Spatio-temporal analysis of Vegetation Dynamics of New Delhi (India) using Satellite Data

Thesis

Submitted to the

G. B. Pant University of Agriculture & Technology
Pantnagar -263145 Uttarakhand, India

By

Navtej Anand
ID. No. 48905

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Master of Technology
In
Information Technology

October, 2022

ACKNOWLEDGEMENT

First of all, I bow my head before 'God' who inspired me to face challenges of uneven times. All my sincere gratitude goes to him for the help he has given to me and his unfailing mercies over my life.

I express my deep sense of reverence and heartfelt gratitude to ***Dr. Subodh Prasad****, Assistant Professor, Department of Information Technology and Chairman of Advisory Committee for his invaluable guidance, constant encouragement, abundant counsel and constructive suggestions throughout the investigation. I am extremely indebted to him and thanks him from the bottom of his heart.*

With profound sense of gratitude, I express my warmest thanks to the members of Advisory Committee, ***Er. Ashok Kumar,*** *Assistant Professor, Department of Information Technology and* ***Er. Govind Verma,*** *Assistant Professor, Department of Information Technology, for their inspiring and constructive suggestions at every stage of this study.*

I tender my sincere thanks to ***Dr. Ajit Kumar,*** *Head, Department of Information Technology,* ***Dr. Alaknanda Ashok,*** *Dean,* ***College of Technology*** *and* ***Dr. Kiran P. Raverkar,*** *Dean,* ***College of Post Graduate Studies****, G. B. Pant University of Agriculture and Technology, Pantnagar for their ardent engrossment and assiduous efforts in providing all the necessary facilities for the completion of the research.*

I am also thankful to AICTE, MHRD, government of India for providing financial assistance and support through GATE.

I owe a very special word of thanks to my father Mr. Harindra Prasad, mother Smt. Shivbachchi Devi for their boundless generosity, everlasting inspiration, blessing abundant love and affection throughout.

This list is obviously incomplete but allow me to submit as the omissions are inadvertent and I once again record my heartfelt gratitude to all those who helped me directly or indirectly in this endeavour

Pantnagar **(Navtej Anand)**

October, 2022 **Author**

CERTIFICATE-I

This is to certify that the thesis entitled **"Spatio-temporal analysis of Vegetation Dynamics of New Delhi (India) using Satellite Data"** submitted in partial fulfilment of the requirements for the degree of **Master of Technology** in **Information Technology** of the College of Post-Graduate Studies, G. B. Pant University of Agriculture and Technology, Pantnagar, is a record of bona fide research carried out by **Mr. Navtej Anand, ID. No. 48905,** under my supervision and no part of the thesis, has been submitted for any other degree or diploma.

The assistance and help received during the course of this investigation have been acknowledged.

Pantnagar
October, 2022

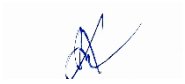
(Subodh Prasad)
Chairperson
Advisory Committee

CERTIFICATE-II

We, the undersigned members of the Advisory Committee of **Mr. Navtej Anand, ID.No. 48905,** a candidate for the degree **Master of Technology** in **Information Technology**, agrees that the thesis entitled **"Spatio-temporal analysis of Vegetation Dynamics of New Delhi (India) using Satellite Data"** may be submitted in partial fulfilment of the requirements for the degree.

(Subodh Prasad)
Chairperson
Advisory committee

(Govind Verma)
Member



(Ashok Kumar)
Member

TABLE OF CONTENTS

(a) List of Tables

(b) List of Figures

(c) List of abbreviations/Symbols

LIST OF TABLES

LIST OF FIGURES

List of Symbols and Abbreviations

Abbreviation	Extended/ Full Form
NDVI	Normalized Difference Vegetation Index
MODIS	Moderate Resolution Imaging Spectroradiometer
DEM	Digital Elevation Model
NIR	Near-infrared
EKG	Electrocardiogram (heart rate monitoring)
EEG	Electroencephalogram
LSTM	Long short-term memory
RNN	Recurrent neural network
IEEE	Institute of Electrical and Electronics Engineers
U.C.L.A	University of California, Los Angeles
LAI	leaf area index
TI NDVI	Time-integrated normalised difference vegetation index
AVHRR	Advanced Very High Resolution Radiometer
CI	Climate indicator
NOAA	National Oceanic and Atmospheric Administration
LANDSAT	Land Remote-Sensing Satellite
SPOT	Satellite Pour l'Observation de la Terre
MRA	Multi-resolution analysis
WT	Wavelet transform
MEDOKADS	Mediterranean Extended Daily One-km AVHRR Data Set
GIMMS	Global Inventory Monitoring and Modeling System
QA	
GEOBIA	Geographic Object-Based Image Analysis
TC	Tree cover

CG	Cropland/grass
GA	Global accuracy
SWIR	Short Wave Infrared
NDWI	Normalized Difference Water Index
DTW	Dynamic Time Warping
GEE	Google Earth Engine
AOI	Area of Interest
GAUL	Global Administrative Unit Layers
EVI	Enhanced Vegetation Index
ADF	Augmented Dickey Fuller Test
KNN	K-Nearest Neighbour
HMM	Hidden Markov Model
ANN	Artificial Neural Network
ReLU	Rectified Linear Unit
MAE	Mean Absolute Error
RMSE	Root Mean Square Error

Chapter 1 INTRODUCTION

1.1 General

Understanding and analyzing the changes in vegetation cover is very important in several aspects including climatic changes, water budget, and ecological balance and specially to undertake necessary conservation measures. The concept of neural network has gained much significance in the analysis of vegetation dynamics using remote sensing satellite data. This work will present an enhanced Change Detection method for the analysis of Satellite image based on Normalized Difference Vegetation Index (NDVI). NDVI employs the Multi-Spectral Remote Sensing data technique to find Vegetation Index, land cover classification, vegetation, water bodies, open area, scrub area, hilly areas, agricultural area, thick forest, thin forest with few band combinations of the remote sensed data. Land Resources are easily interpreted by computing their Normalized Difference Vegetation Index for Land Cover classification. Remote Sensing data from MODIS/Terra satellite image along with NDVI and DEM data layers have been used to perform multi-source classification **(Reddy and Prasad, 2018)**. The Change Detection method used is NDVI differencing. NDVI method is applied according to its characteristic like vegetation at different NDVI threshold values such as 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4 and 0.5. Many studies have shown that the NDVI is highly useful in detecting the surface features of the visible area which are extremely beneficial for policy makers in decision making. The Vegetation analysis can be helpful in predicting the unfortunate natural disasters to provide humanitarian aid, damage assessment and furthermore to device new protection strategies.

1.2 NDVI

The NDVI is a dimensionless index that describes the difference between visible and near-infrared reflectance of vegetation cover and can be used to estimate the density of green on an area of land. To the human eye, a plant is green because the chlorophyll pigment in it reflects green waves and absorbs red waves. Cell structures in plants reflect the near-infrared (NIR) waves. So, when photosynthesis occurs, the plant develops and grows and contains more cell structures.

This means that a healthy plant, one with a lot of chlorophyll and cell structures, actively absorbs red light and reflects NIR. An unhealthy plant will do the exact opposite. This relationship between light and chlorophyll is how we can use NDVI to tell apart a healthy plant from a diseased one. Satellite sensors in space measure wavelengths of light

absorbed and reflected by green plants. They are an excellent source of spectral signature data for NDVI analysis. The NDVI index detects and quantifies the presence of live green vegetation using this reflected light in the visible and near-infrared bands. Put simply, NDVI is an indicator of the vegetation greenness, the density and health of each pixel in a satellite image.

1.3 Time Series

Time series is collecting data points over some time. A point to consider while doing time series analysis is that one should take a fixed interval of periods for effective calculations and should not rely on the random period intervals. One thing which makes time series unique from other forms of analysis is that it can show how variables (independent and dependent) change with time. That's why we say that time plays a pivotal role as a variable because it shows the way the data changes or adjusts with the course of data points as well as with the final result. It also gives us an extra source of information and a defined order of dependencies between the data.

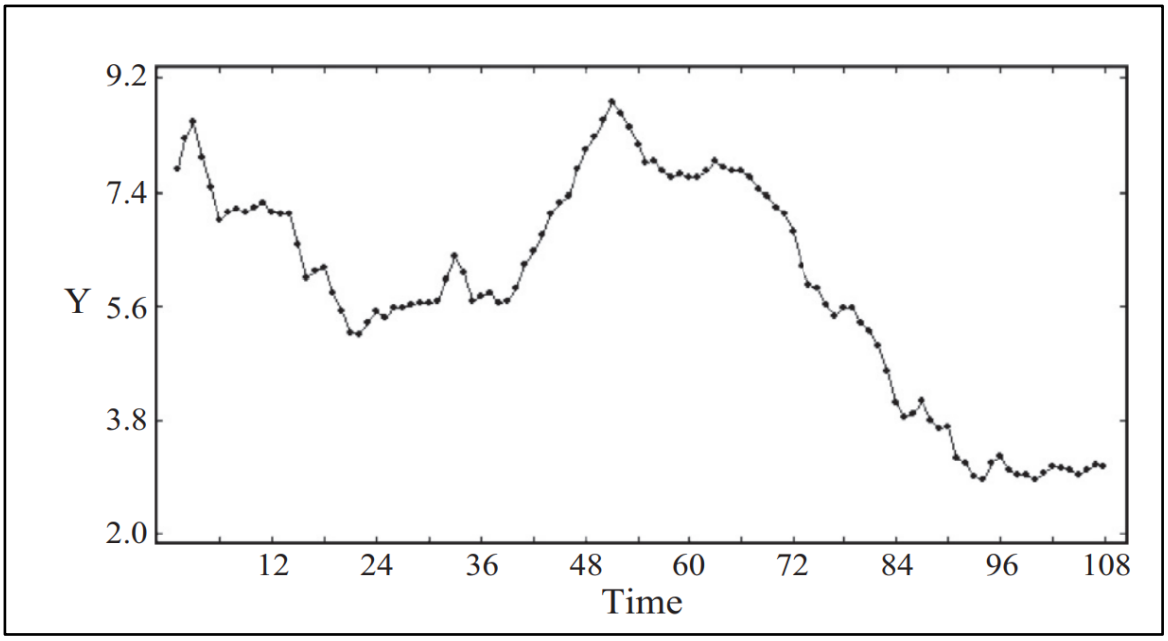


Fig. 1.1 Monthly interest rates of 90-day Treasury bills, 1985–93.
(source: (Tiao, 2015))

The Fig. 1.1 is an example for time series data for monthly interest rates variations recorded for years 1985 to 1983. Graphical representation makes visualization and analysis easier and convenient. Some models of time series analysis include classification, curve fitting, descriptive analysis, explanative analysis, explorative analysis, forecasting, intervention analysis, and segmentation. Time series analysis is used for non-stationary

data that changes with time. Examples of time series analysis include weather data, stock prices, temperature readings, etc.

These methods of analysing a series of data points gathered over a period of time is called a "time series analysis." Instead of just recording the data points intermittently or randomly, time series analysts record the data points at regular intervals over a predetermined period of time. But this kind of analysis involves more than just gathering data over time. Data from time series can be analysed to reveal how variables change over time, which distinguishes them from other types of data. To put it another way, time is a key variable because it both reveals how the data changes over the course of the data points and the outcomes. It offers a second source of information and establishes a specific hierarchy of dependencies between the data.

To ensure consistency and reliability, time series analysis typically needs a lot of data. A large data set guarantees that your analysis can sift through erratic data and that your sample size is representative. Additionally, it guarantees that any trends or patterns are not outliers and can take seasonal variation into account. Time series data can also be used to forecast future data by basing predictions on past data.

Time series analysis is also excellent for predicting weather variations, assisting meteorologists in foreseeing everything from tomorrow's weather report to upcoming years of climate change. Few examples of time series analysis in action are stated below:

- Weather data
- Rainfall measurements
- Temperature readings
- Heart rate monitoring (EKG)
- Brain monitoring (EEG)
- Quarterly sales
- Stock prices
- Automated stock trading
- Industry forecasts
- Interest rates

Time series analysis models include:

- **Classification:** Identifies and assigns categories to the data.
- **Curve fitting:** Plots the data along a curve to study the relationships of variables within the data.
- **Descriptive analysis:** Identifies trends, cycles, or seasonal variation in time series data.
- **Explanative analysis:** Attempts to understand the data and the relationships within it, as well as cause and effect.
- **Exploratory analysis:** Highlights the main characteristics of the time series data, usually in a visual format.
- **Forecasting:** Predicts future data. This type is based on historical trends. It uses the historical data as a model for future data, predicting scenarios that could happen along future plot points.
- **Intervention analysis:** Studies how an event can change the data.
- **Segmentation:** Splits the data into segments to show the underlying properties of the source information.

In this research we are going to develop a machine learning model and analyse the outputs and results of that model with other models. From the above time series models we are applying forecasting model using LSTM network. So we will take a quick look at LSTM.

1.4 Long short time memory (LSTM)

An artificial neural network called long short-term memory (LSTM) is used in deep learning and artificial intelligence. LSTM has feedback connections as opposed to typical feedforward neural networks. Such a recurrent neural network (RNN) can process entire data sequences (such as speech or video) in addition to single data points (such as images). For instance, LSTM can be used for tasks like speech recognition, robot control, machine translation, connected, unsegmented handwriting recognition, video games, and healthcare. The most frequently used neural network of the 20th century is LSTM.

The name of LSTM refers to the analogy that a standard RNN has both "long-term memory" and "short-term memory". The activation patterns in the network change once per time-step, analogous to how physiological changes in synaptic strengths store short-term memories. The connection weights and biases in the network change once per episode of training, analogous to how physiological changes in synaptic strengths store long-term memories. The "long short-term memory" of the LSTM architecture is intended to give

RNN a short-term memory that can last thousands of timesteps. The below Fig. 1.2 shows the basic structure of a LSTM.

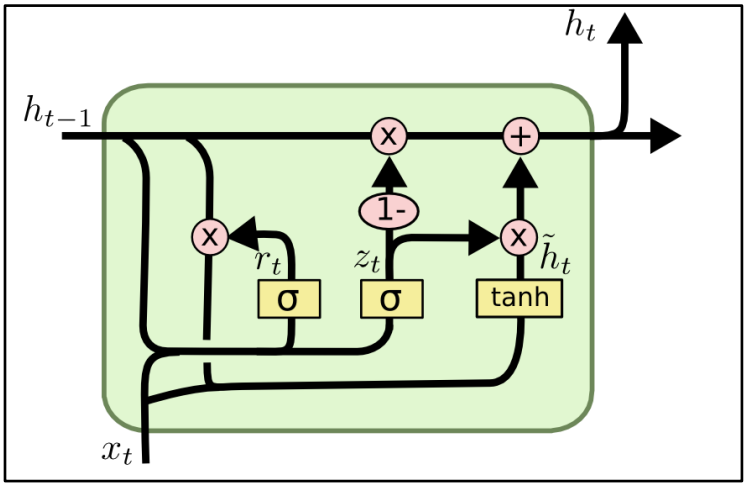


Fig. 1.2 Unit structure of LSTM model
Source: (colah.github.io/posts/2015-08-Understanding-LSTMs/)

1.5 Problem Statement

This thesis work is all about the "Spatio-temporal analysis of vegetation dynamics of New Delhi (India) using satellite data" where an analysis has been carried out over the dataset that has been derived from MODIS satellite to analyse vegetation change and forecast the changes using NDVI time series.

The main focus of the study is to analyse the forecasted output and accuracy assessment of our model. Long short term memory (LSTM), an advanced artificial neural network technique, used in the model has been deployed on the MODIS terra satellite dataset to forecast the future vegetation changes.

The specific objectives of the study are:

I. To obtain NDVI values from satellite dataset.
II. To generate time series from the NDVI values obtained.
III. To deploy Long Short-Term Memory (LSTM) machine learning model to analyse NDVI time series data for the prescribed region.
IV. To forecast future NDVI time series.
V. Accuracy assessment comparison of the outputs obtained using LSTM models.

Thesis Outline

Chapter 1 describes the background information, motivation, problem statement, objectives and significance of the study.

Chapter 2 is about literature review related to vegetation change analysis and prediction using NDVI time series and LSTM.

Chapter 3 explains computational technique, hardware and software tools that have been used in the development of the proposed work.

Chapter 4 mentions to the results and discussion of the proposed work.

Chapter 5 deals with the summary, conclusions and future scope of the work.

Chapter 2 REVIEW OF LITERATURE

A literature review locates and summarises research on a certain topic. In the context of blended procedures, the written material examination is regarded as a trustworthy method for determining the true type of concentration, whether quantitative or subjective. In order to gather pertinent data that helped in addressing the research questions, a thorough literature review is conducted using sources such as IEEE and other journals on vegetation dynamics analysis. References to studies relating to NDVI time analysis using a long short term memory network model have also been made in order to aid in the execution of this proposed work.

Anderson (1971) stated that the land use classification schemes for use with orbital imagery for making thematic maps of land use in the United States ranging from 1:250,000 to 1:2,500,000 in scale is a problem. A perspective on recent attempts to develop classification systems that would be usable with imagery from remote sensors in orbiting spacecraft is given by some background on approaches to land use classification based mainly on aerial photographs. Several criteria are proposed to provide a framework for such a review and classification. Effective classification schemes must be developed for use with orbital imagery, so it is hoped that the review and evaluation of land use maps prepared by N.J. Thrower and others at U.C.L.A. from Gemini to Apollo imagery will direct attention to some of the serious problems. That must be recommendations are shown as directives for additional study. There are two proposed tentative land use classification schemes for future orbital imagery testing.

Anderson *et al.* (1976) presented a framework of a national land use and land cover classification system for use of data from remote sensors, The classification system was created to satisfy the demands of Federal and State agencies for a current overview of land use and land cover throughout the nation on a basis that is uniform in categorization at the more generalised first and second levels and that will be open to data from satellite and aircraft remote sensors. The suggested system makes use of popular, currently in use classification systems that can handle data from remote sensing sources. U.S. Geological Survey Circular 671 presents a revision of the land use classification system that takes into account the findings of in-depth testing and review of the categorization and definitions.

Carlson and Ripley (1997) demonstrated the relationships between the NDVI, leaf area index (LAI), and fractional vegetation cover, we use a straightforward radiative transfer model with vegetation, soil, and atmospheric components. The following points are made: i) LAI and fractional vegetation cover are not completely independent variables. ii) An scaled NDVI taken in clear or hazy conditions between the limits of minimum (bare soil) and maximum fractional vegetation cover is insensitive to atmospheric correction. iii) A simple relation between scaled NDVI and fractional vegetation cover is further confirmed by the simulations. iv) The sensitive dependence of LAI on NDVI when the former is below a value of about 2-4 may be viewed as being due to the variation in the bare soil component.

Yang *et al.* (1998) determined that as a substitute for primary production, the time-integrated normalised difference vegetation index (TI NDVI), which was derived from the multitemporal satellite imagery (1989-1993), was used to look into the effects of climate change on the performance of grasslands in the central and northern Great Plains. The findings point to the most significant controls on grassland performance and productivity as being spatial and temporal variability in growing season precipitation, potential evapotranspiration, and growing degree days. A statistical model revealed a significant positive correlation between the TI NDVI and accumulated spring and summer precipitation when TI NDVI and climate data of all grassland land cover classes were examined collectively, and a significant negative correlation when TI NDVI and spring potential evapotranspiration were examined. The general model's coefficient of determination (R2) was 0.45. In order to track the dynamics of grassland ecosystems on a regional scale, the study recommends an integrated approach involving numerical models, satellite remote sensing, and field observations.

Ichii *et al.* (2002) used the Pathfinder AVHRR Land NDVI data set and observed climate data for the years 1982 to 1990, the relationship between the Normalized Difference Vegetation Index (NDVI) and climatic variables was examined on a global scale. In the northern mid- to high-latitude regions between spring and autumn, a significant correlation between interannual NDVI and temperature variation was found. In both the northern and southern semi-arid regions, a significant correlation between the NDVI, temperature, and precipitation was also found. A comparison of global NDVI trends reveals that temperature increases are related to NDVI increases in the northern mid- and high latitudinal zones, while precipitation decreases during the survey period are related to

NDVI decreases in the southern semiarid regions. The combined effects of forest regrowth, deforestation, and fertilisation may have an impact on the NDVI trend, even though the exact cause of NDVI increases in the equatorial regions is still unknown.

Roerink *et al.* (2003) discovered that the dynamics of the vegetation are significantly impacted by climatic variability. A study using Normalized Difference Vegetation Index (NDVI) satellite images and meteorological data is conducted over a portion of Sahelian Africa and Europe over a number of years to quantify this impact. Through a time series analysis of NDVI satellite images using the Harmonic Analysis of Time Series algorithm, the vegetation dynamics are quantified as the total amount of vegetation (mean NDVI) and the seasonal difference (annual NDVI amplitude). Meteorological information is used to create a climate indicator (CI) (precipitation over net radiation). Both spatial and temporal factors determine the relationships between the dynamics of the vegetation and the CI. The driest regions turn out to be the most vulnerable to climate change. While the seasonal difference partially alters the NDVI's spatial and temporal patterns, they remain the same overall. With these sparse datasets, it is impossible to provide a satisfactory answer to the question of whether the impact of climate on vegetation dynamics is the same everywhere on Earth in the time and space domain.

Anyamba and Tucker (2005) created a 23-year time series by NOAA-remote AVHRR's sensing measurements and expressed as the normalised difference vegetation index (NDVI) is suitable for long-term studies of the Sahel region. The ability to use NDVI data as a proxy for the land surface response to precipitation variability results from the close relationship between Sahelian rainfall and vegetation growth. A closer look at this time series reveals two distinct periods: (a) 1982–1993, marked by persistent drought and below average NDVI, with a signature large-scale drought occurring during the 1982–1985 timeframe; and (b) 1994–2003, marked by a trend towards "wetter" conditions and region-wide above normal NDVI conditions with maxima in 1994 and 1999. These patterns are consistent with recent regional rainfall trends in the Sahel. These conditions are still significantly below the more humid conditions that prevailed in the area from 1930 to 1965, even when viewed in the context of long-term Sahelian climate history. Therefore, it is only possible to interpret these trend patterns as a slow recovery from the severe drought conditions that peaked between 1983 and 1985. A detailed spatial quantification and

description of the recovery patterns at the local scale will be provided by systematic studies of changes on the landscape using high spatial resolution satellite data sets, such as those from LANDSAT, SPOT, and MODIS.

Jiang *et al.* (2006) used linear spectral mixing models to examine the spatial scale dependencies of NDVI and the relationship between NDVI and fractional vegetation cover at various resolutions. The results indicate that NDVI values at various resolutions may not be comparable because of the strong spatial scale dependencies of NDVI over heterogeneous surfaces. With darker soil backgrounds and the presence of shadow, the NDVI over partially vegetated surfaces exhibits nonlinear behaviour that is more pronounced. Due to its nonlinearity and scale effects, the NDVI may not be appropriate to infer the vegetation fraction.

Martínez and Gilabert (2009) implemented multi-resolution analysis (MRA) based on the wavelet transform (WT) to study NDVI time series. These series, which are non-stationary and present short-term, seasonal and long-term variations, can be decomposed using this MRA as a sum of series associated with different temporal scales. The main focus of the paper is to check the potential of this MRA to capture and describe both intra- and inter-annual changes in the data, i.e., to discuss the ability of the proposed procedure to monitor vegetation dynamics at regional scale. Our approach concentrates on what wavelet analysis can tell us about a NDVI time series. On the one hand, the intra-annual series, linked to the seasonality, has been used to estimate different key features related to the vegetation phenology, which depend on the vegetation cover type. On the other hand, the inter-annual series has been used to identify the trend, which is related to land-cover changes, and a Mann-Kendall test has been applied to confirm the significance of the observed trends. NDVI images from the MEDOKADS (Mediterranean Extended Daily One-km AVHRR Data Set) imagery series over Spain are processed according to a per-pixel strategy for this study. Results show that the wavelet analysis provides relevant information about vegetation dynamics at regional scale, such as the mean and minimum NDVI value, the amplitude of the phenological cycle, the timing of the maximum NDVI and the magnitude of the land-cover change. The latter, in combination with precipitation data, has been used to interpret the observed land-cover changes and identify those subtle changes associated to land degradation.

Zhang *et al.* (2009) used GIMMS NDVI database and geo-statistics to depict the spatial distribution and temporal stability of NDVI on the Mongolian Plateau. The results demonstrated that: (1) Regions of interest with high NDVI indices were distributed primarily in forested mountainous regions of the east and the north, areas with low NDVI indices were primarily distributed in the Gobi Desert regions of the west and the southwest, and areas with moderate NDVI values were mainly distributed in a middle steppe strap from northwest to southeast. (2) The maximum NDVI values maintained for the past 22 years showed little variation. The average NDVI variance coefficient for the 22-year period was 15.2%. (3) NDVI distribution and vegetation cover showed spatial autocorrelations on a global scale. NDVI patterns from the vegetation cover also demonstrated anisotropy; a higher positive spatial correlation was indicated in a NW-SE direction, which suggested that vegetation cover in a NW-SE direction maintained increased integrity, and vegetation assemblage was mainly distributed in the same specific direction. (4) The NDVI spatial distribution was mainly controlled by structural factors, 88.7% of the total spatial variation was influenced by structural and 11.3% by random factors.

de Jong *et al.* (2011) carried out a study and noticed that in particular, the global coverage of time-series normalised difference vegetation index (NDVI) data, which are available from 1981, are widely used to identify greening and browning trends. We use harmonic analyses and non-parametric trend tests to the GIMMS NDVI dataset as an alternative to reducing the temporal resolution, which has previously been used to address seasonality and serial auto-correlation in the data (1981-2006). A seasonal non-parametric model and a linear model with seasonality removed were used to analyse the greening and browning trends using the entire dataset. In a third method, vegetation development stages rather than the calendar days were used to analyse the time-series in order to account for phenological shift and variations in growing season length. Despite having the same input data, the models produced very different results. Multiple other studies' identification of prominent regional greening trends was confirmed, but the models were inconsistent in regions with weak trends. Similar trend slopes to those described in earlier work using linear models on yearly mean values were revealed by the linear model using data that had been adjusted for seasonality. The non-parametric models showed the significant impact of phenological variation; accounting for this variation should lead to more reliable trend analyses and improved comprehension of vegetation trends.

Bhandari *et al.* (2012) presented an improved method for the analysis of satellite image based on Normalized Difference Vegetation Index (NDVI). For land cover classification, some band combinations of the remote sensed data are exploited and the spatial distribution such as road, urban area, agriculture land and water resources are easily interpreted by computing their normalized difference vegetation index. Different values of threshold of NDVI are used for generating the false colour composite of the classified objects. The simulation results show that the NDVI is highly useful in detecting the surface features of the visible area which are extremely beneficial for municipal planning and management. The vegetation analysis can be used for the situation of unfortunate natural disasters to provide humanitarian aid, damage assessment and furthermore to device new protection strategies.

Fensholt and Proud (2012) found that a new and updated version of the AVHRR (Advanced Very High-Resolution Radiometer) based GIMMS (Global Inventory Modelling and Mapping Studies) NDVI (Normalized Difference Vegetation Index) dataset is now available covering 1981 to 2010 (GIMMS3g). Earlier versions of this global coverage 15-day composite dataset have been used for numerous local to global scale vegetation time series studies during recent years. However, several aspects of the AVHRR sensor design and data processing potentially introduce substantial noise into the NDVI dataset if not corrected for. The more recent NDVI dataset from Terra MODIS (Moderate Resolution Imaging Spectroradiometer) is considered an improvement over AVHRR data and with the release of GIMMS3g an overlapping period of 11 years now provides a possibility to perform a robust evaluation of the accuracy of GIMMS3g data and derived trends. In this study the accuracy is evaluated by comparison with the global Terra MODIS NDVI (MOD13C2 Collection 5) data using linear regression trend analysis. The trends of GIMMS NDVI were found to be in overall acceptable agreement with MODIS NDVI data. A significant trend in NDVI ($\alpha = 0.05$) was found for 11.8% of the MODIS NDVI pixels on a global scale (5.4% characterised by positive trends and 6.3 with negative trends) whereas GIMMS NDVI analysis produced a total of 10.5% significant pixels (4.9% positive, 5.6% negative). However, larger differences were found for the Southern Hemisphere land masses (South America and Australia) and the high northern latitude Arctic regions. From a linear regression analysis the correlation coefficient between the two datasets was found to be highly significant for areas with a distinct phenological cycle. Discrepancies between the GIMMS and MODIS datasets were found in equatorial areas

(broadleaved, evergreen forest), Arctic areas (sparse herbaceous or sparse shrub cover) and arid areas (herbaceous cover, closed-open). Linear regression of QA filtered Terra and Aqua MODIS NDVI (2003-2010) revealed similar inconsistencies for Arctic and equatorial areas suggesting that robust long-term NDVI trend estimates in these areas are difficult to obtain from both GIMMS and MODIS data. Additionally, GIMMS based NDVI trend analysis in arid areas of limited photosynthetic activity should be interpreted with caution. The regression coefficient (slope value) ($p < 0.01$) was found to be close to 1 for most land cover types on a global scale (global land cover class average slope = 1.00) suggesting overall compatibility between MODIS and GIMMS NDVI, but with land cover class specific variations (within class and between classes).

Wang *et al.* (2012) concluded that time series of satellite data offer unmatched insight into how vegetation reacts to climate change. Consistent satellite-based measurements are required to identify minute changes in vegetation over time. Using Collection 5 data from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors on the Terra and Aqua platforms, the effect of sensor degradation on trend detection was assessed in this study. Under a variety of simulated aerosol conditions and surface types, the Normalized Difference Vegetation Index (NDVI) for Terra MODIS decreased by 0.001-0.004yr-1 at near-nadir view angles due to the impact of blue band (Band 3, 470nm) degradation on simulated surface reflectance. The negative NDVI trends derived from Terra (17.4%) and Aqua (6.7%) MODIS sensors between 2002 and 2010 were nearly three times different from observed trends in MODIS NDVI over North America. This unfavourable bias in the detection of NDVI trends will be largely eliminated by Terra MODIS calibration adjustments planned for Collection 6 data reprocessing.

Nyamekye *et al.* (2014) carried out the study by applying principal component analysis and classification of average NDVI values. It showed that the dominant land cover change process was conversion of natural vegetation to savannah and shrub thicket, which occurred at an annual rate of 4% and 6.5% respectively. Most of the land cover change process occurred in the first period (1982-1992). The overall annual rate of change in land cover (1982-1992) was highest for agriculture (3.5%) and lowest for water (1.03%). The results suggest that, yearly phonological behaviour as revealed by NDVI data can be used to map general patterns in the spatial distribution of Ghana's main vegetation formation.

Aburas *et al.* (2015) studied two Landsat TM images from 1990 to 2010 to extract NDVI values. The Natural Breaks (Jenks) method is initially used to compute the NDVI values in order to categorise the NDVI map. After that, a Difference NDVI map between 1990 and 2010 is created to evaluate the values of changing land cover. Results showed that the area devoid of vegetation, including built-up areas, barren lands, and bodies of water, increased from 3.55 % to 7.25 % in 2010. The city must develop new policies for the protection of vegetation areas during the urban and economic development in Seremban as the dense vegetation area decreased from 78.57 % to 65.44 %.

Gandhi *et al.* (2015) presented an enhanced Change Detection method for the analysis of Satellite image based on Normalized Difference Vegetation Index (NDVI). Remote Sensing data from Landsat TM image along with NDVI and DEM data layers have been used to perform multi-source classification. The Change Detection method used was NDVI differencing. NDVI method is applied according to its characteristic like vegetation at different NDVI threshold values such as 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4 and 0.5. The Simulation results show that the NDVI is highly useful in detecting the surface features of the visible area which are extremely beneficial for policy makers in decision making. The Vegetation analysis can be helpful in predicting the unfortunate natural disasters to provide humanitarian aid, damage assessment and furthermore to device new protection strategies. From the empirical study, the forest or shrub land and Barren land cover types have decreased by about 6% and 23% from 2001 to 2006 respectively, while agricultural land, built-up and water areas have increased by about 19%, 4 % and 7% respectively.

Reddy and Prasad (2018) made an attempt to predict the vegetation dynamics using MODIS NDVI time series data sets and long short term memory network, an advanced technique adapted from the artificial neural network. The dataset of 861 NDVI images from January 2000 to June 2016 is used for making the time series. These include different terrains in the Great Nicobar Islands, one region along the coast where vegetation has severe ecological damage due to 2004 Indian Ocean tsunami and the other, an interior region which remained imperturbable during the tsunami. Long short term memory network, an advanced neural network is trained with these NDVI values for both the regions separately to predict the future vegetation dynamics. To measure the accuracy of the LSTM network, root mean square error is calculated. The resulting plots from both the experiments indicate that the long short-term memory neural network follows the series in addition to coinciding with the required time series. Also, an unanticipated change in the

trend of the NDVI series were well adapted by the network and was able to predict the future NDVI values with good accuracy maintaining RMSE less than 0.03 without providing any supplementary data.

Dutra *et al.* (2018) characterized the land cover and use over the 2000–2010 time periods for the Brazilian Caatinga seasonal biome using a temporal Normalized Difference Vegetation Index (NDVI) series and Geographic Object-Based Image Analysis (GEOBIA). For each of the target years NDVI images were derived from a Moderate Resolution Imaging Spectroradiometer (MODIS, MOD13Q1, at a 250 m spatial and 16-day temporal scale) sensor during the dry season to predict wood cover in the municipality of Buriti dos Montes, in the state of Piaui in the north-east region of Brazil (H13V09 tile). The images were automatically pre-processed and the GEOBIA approach was performed for image segmentation, spatial and spectral attribute extraction and labelling according to the following legend, tree cover (TC) and cropland/grass (CG), to obtain a classification using the decision tree supervised algorithm. Our results showed that the approach using GEOBIA presented a Kappa index of 0.58 and global accuracy (GA) of 0.81% and showed better accuracy for the tree cover.

Li *et al.* (2019) saw that recent years have seen notable advancements in time series prediction using deep learning techniques, particularly Long Short-Term Memory Neural Network (LSTM). While LSTM can aid in capturing long-term dependencies, it is insufficient for paying different amounts of attention to sub-window features over the course of multiple time steps. A solution for this problem is a multivariate time series prediction method using evolutionary attention-based LSTM training and competitive random search. LSTM is given an evolutionary attention learning approach by transferring shared parameters. Thus, during temporal relationship mining, the pattern for importance-based attention sampling can be confirmed, just like that for biological evolution. A competitive random search method inspired by evolutionary computation is proposed, which can effectively configure the parameters in the attention layer. This method avoids being caught in partial optimization like traditional gradient-based methods. According to experimental findings, the proposed model can produce competitive prediction performance when compared to other industry standards.

Taufik *et al.* (2019) created the Normalized Different Vegetation Index (NDVI) image using the Landsat 8 visible and Near Infrared (NIR) bands. The NIR and Short Wave Infrared (SWIR) bands are used to create the Normalized Difference Water Index (NDWI), which is a satellite-derived index. The classification experiment of three unsupervised methods, ISODATA, K-means, and fuzzy c-means were then conducted with the aid of ground truth data from the study area to analyse the classes of vegetation, non-vegetation, and water features. The outcome demonstrates that unsupervised methods classifications are better able than the other two methods to classify the data from the Landsat 8 satellite with a high degree of accuracy.

Jiang *et al.* (2021) used long time series NDVI data (1998–2018) to evaluate the ecological environment change and spatiotemporal change that has taken place in China over the last 20 years, at different scales (regional, provincial, county, and grid scales). Combined with typical areas, the causes of ecological environment change are revealed. The results show that in about one-third of China, the ecological environment improved, and in 10%, the ecological environment deteriorated. Ningxia, Shaanxi, Shanxi, Guizhou, and Guangxi are the provinces in which the improvement of the ecological environment is most pronounced. However, the ecological environments with vegetation as the core in Shanghai and Jiangsu Province have deteriorated significantly. The areas with the fastest improvement in their ecological environment are those with precipitation of 200–600 mm. The Yellow River Basin is an area where the ecological environment has significantly improved. Forestry projects and the development of oasis agriculture effectively promote the improvement of the ecological environment.

Zhao *et al.* (2021) proposed that the NDVI is a crucial agricultural indicator that has a strong relationship with crop development. In this study, a novel method for predicting the short and medium-term winter wheat NDVI was developed by combining the dynamic time warping (DTW) model and the long short-term memory (LSTM) deep recurrent neural network model. Although LSTM is effective at modelling long-term dependencies, this approach may be prone to overfitting. In contrast, DTW is more resistant to overfitting and has good predictive ability. Therefore, by combining these two models, the overfitting-related prediction error is decreased, increasing the accuracy of the final prediction. Using historical MODIS time series data with an 8-day time resolution from 2015 to 2020, the combined method is presented here. The DTW model is used for prediction in the first section, which reflects the inter-annual and seasonal variation

characteristics of winter wheat NDVI. The LSTM model is used for prediction in the second section, which reflects the winter wheat, NDVI's short-term change characteristics. The output of both models is then combined to create the final prediction. The DTW-LSTM model proposed here outperforms the LSTM model based on multiple evaluation indicators, according to a case study in Hebei Province that forecasts the NDVI of winter wheat at five prediction horizons in the future.

Chapter 3 MATERIALS AND METHODS

This chapter describes the Study Area, Collection of data, Methods of analyzing the data, and detailed comparison between the methods and parameters to measure the difference in accuracy.

3.1 Study Area

The geographical region chosen here for study is Delhi. Delhi is the capital of India. The state is spread over an area of 1483 square kilometer. According to the Indian geography the state is located at the center of the Indian subcontinent, amidst the ranges of Himalaya and the Aravalli. Delhi geography encompasses the location, climatic conditions, vegetation and so on. The latitudinal and longitudinal locations of Delhi are 23.38 degree north and 77.13 degree east. Delhi geography divides the state into three parts- the Delhi ridge, the Yamuna flood plain and the plains. The Yamuna river plains are very fertile as they are flooded by the river and are rich in alluvial soil. The Delhi ridge is the most important characteristic of the state and is a part of the Aravalli range that passes through Delhi.

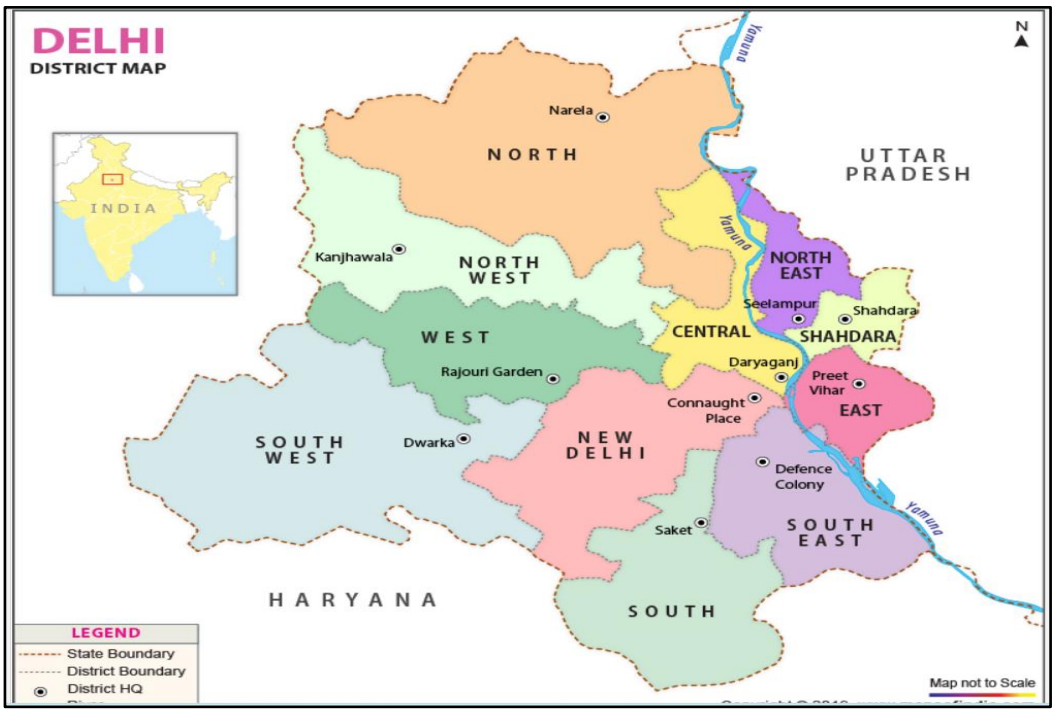


Fig. 3.1 District map of Delhi
Source: (mapsofindia.com/delhi/districts/)

To study the nature of vegetation that grows in the different parts of the city, it is first important to know about the topographical variations of Delhi. The entire topography of Delhi is divided into a ridge, the Yamuna Flood Plain, the Plain. It is interesting to note here that each of these regions is marked by distinct type of vegetation. The ridge area of the city offers the right factors that favour the growth of acacias and other cacti. However, during the monsoon, herbaceous plants grow in abundance in the ridge. As far as the plain region of Delhi is concerned, it is characterized by Shisham trees. And finally, riverine type of vegetation grows along the plain of Yamuna.

Vegetation of Delhi mainly comprise of medium size trees and herbs. However, Delhi is known for its varied flowering plants. Weeds and grass grow on the banks of the Yamuna River. Delhi Weather varies with the different climatic conditions that are faced by this city. Delhi is a city characterized by weather extremes. The geographical location of this city influences the weather conditions of Delhi. Delhi experiences tropical steppe type of climate and hence its seasons are marked with extreme temperatures.

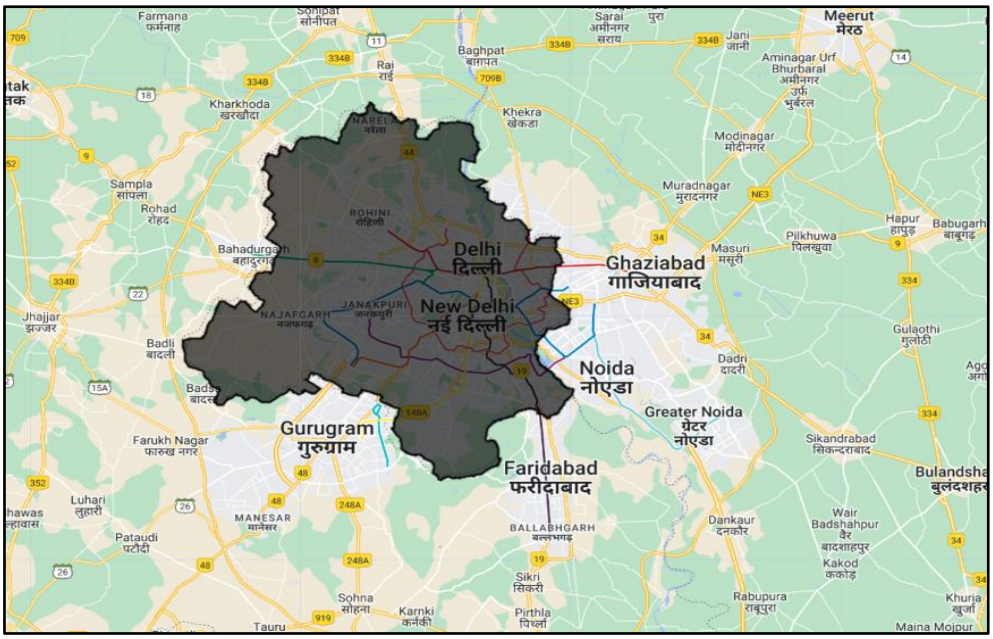


Fig. 3.2 Shape file of Delhi
Source: (GEE)

The area of interest (AOI) has been marked on the Google earth map using Global Administrative Unit Layers (GAUL) shape file. The Global Administrative Unit Layers (GAUL) contributes to the standardisation of the spatial dataset that represents administrative units by gathering and disseminating the best information on administrative units for every nation in the world. The GAUL always maintains global layers with a

unified coding system at country, first (e.g. departments), and second administrative levels (e.g. districts). When data is available, it offers layers that are broken down into third, fourth, and lower levels on a nation basis. The overall methodology entails:

a) Gathering the best data available from the most trustworthy sources,

b) Establishing validation periods of the geographic features (when possible),

c) Adding selected data to the global layer based on the most recent UN Cartographic Unit (UNCS). Map of country boundaries,

d) Creating codes using the GAUL Coding System, and

e) Providing users with data.

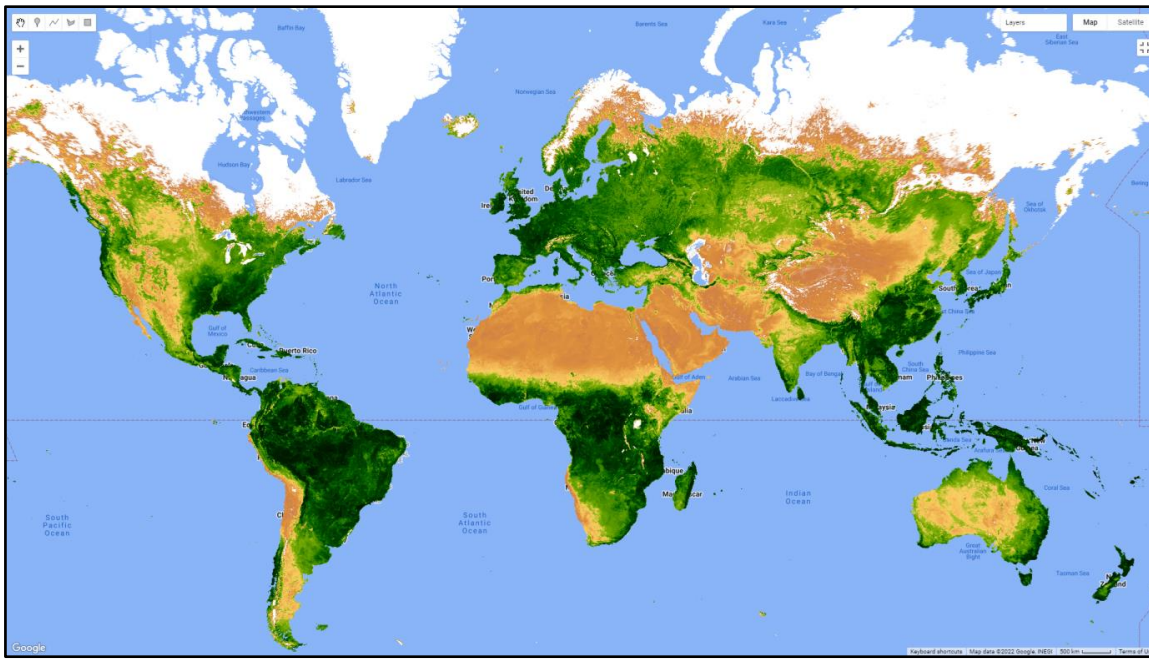


Fig. 3.3 NDVI map of the world
Source: (GEE)

In this study we are going to analyse the vegetation changes based on these kinds of Ndvi images and later based on the time series data obtained from satellite imagery we are going to forecast the future pattern of the change. Analysis has been divided into two parts. First part analyses visual changes, charts, graphs and tables. While second part has been termed as the time series analysis and prediction which has been done based on LSTM (Long Short Term Memory) model.

3.2 Methodology

The following Fig. 3.4 shows the methodology has been adopted for the vegetation change detection and forecasting the future NDVI time series. Two different LSTM models have been developed. Both models have been trained and tested on same dataset to assess their accuracy and to find out best fit model.

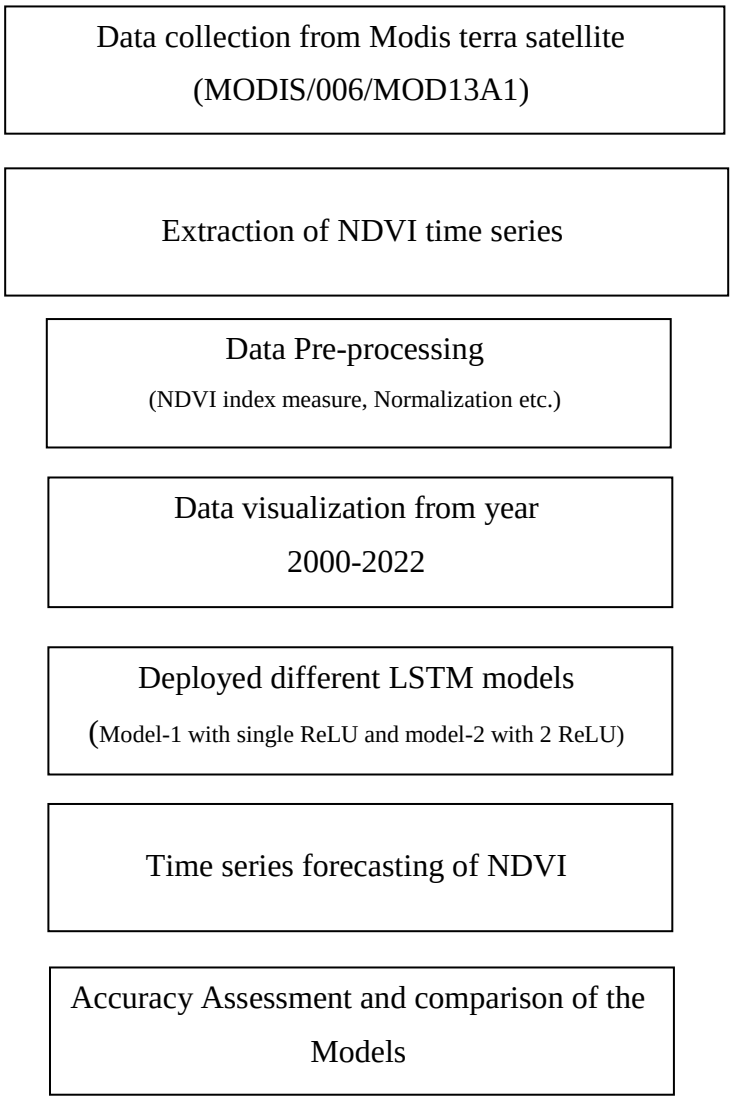


Fig. 3.4 Flowchart of the methodology

3.2.1 Data Collection

The dataset used for the study has been derived from MODIS terra satellite using Google Earth Engine. The dataset is present by the name "MOD13A1.006 Terra Vegetation Indices 16-Day Global 500m [deprecated]". The MOD13A1 V6 product offers a Vegetation Index (VI) value per pixel. There are two main layers of vegetation. The first is the Normalized Difference Vegetation Index (NDVI), also known as the continuity index to the current NDVI derived from the NOAA-AVHRR, i.e. National Oceanic and Atmospheric Administration-Advanced Very High Resolution Radiometer. The Enhanced Vegetation Index (EVI), which maintains sensitivity over conditions of dense vegetation, is the second layer of vegetation. It reduces canopy background variations. The blue band is also used by the EVI to eliminate any remaining smoke and sub-pixel-thin cloud contamination in the atmosphere. The atmospherically corrected bi-directional surface reflectances used to calculate the MODIS NDVI and EVI products have been masked for water, clouds, heavy aerosols, and cloud shadows (**Didan, K., 2015)**.

Dataset Availability from MOD13A1.006:

2000-02-18 to Present

Dataset Provider:

NASA LP DAAC at the USGS EROS Center

Earth Engine Snippet:

ee.ImageCollection("MODIS/006/MOD13A1")

Resolution:

500 meters

Bands:

Bands from the satellite are shown in
Table **3.1** and Table 3.2

Table 3.1: Bands of MOD13A1.006 dataset

Name	Units	Min	Max	Scale	Wavelength	Description
NDVI		-2000	10000	0.0001		Normalized Difference Vegetation Index
EVI		-2000	10000	0.0001		Enhanced Vegetation Index

Table 3.2 Bands in MOD13A1.006 dataset

Name	Units	Min	Max	Scale	Wavelength	Description
sur_refl_b01		0	10000	0.0001	645nm	Red surface reflectance
sur_refl_b02		0	10000	0.0001	858nm	NIR surface reflectance
sur_refl_b03		0	10000	0.0001	469nm	Blue surface reflectance
sur_refl_b07		0	10000	0.0001	2130nm/2105 - 2155nm	MIR surface reflectance
ViewZenith	Degrees	0	18000	0.01		View zenith angle
SolarZenith	Degrees	0	18000	0.01		Solar zenith angle
RelativeAzimuth	Degrees	-18000	18000	0.01		Relative azimuth angle
DayOfYear		1	366			Julian day of year
SummaryQA						Quality reliability of VI pixel

Bitmask for SummaryQA:

- Bits 0-1: VI quality (MODLAND QA Bits)
    - 0: Good data, use with confidence
    - 1: Marginal data, useful but look at detailed QA for more information
    - 2: Pixel covered with snow/ice
    - 3: Pixel is cloudy

3.2.2 NDVI Extraction

For this study the data is used only from NDVI layer of the MODIS dataset. NDVI is highly useful in detecting the surface features of the visible area. The MODIS satellite provides direct NDVI values as it has a dedicated band for NDVI. While nearly all other

satellite Vegetation Indices employ the difference formula to quantify the density of plant growth on the Earth- near-infrared radiation minus visible radiation divided by near-infrared radiation plus visible radiation. This formula's output is known as the Normalized Difference Vegetation Index (NDVI).

As shown below, Normalized Difference Vegetation Index (NDVI) uses the NIR and Red channels in its formula.

$$\mathrm{NDVI} = \frac{(\mathrm{NIR}-\mathrm{RED})}{(\mathrm{NIR}+\mathrm{RED})} \quad (1)$$

Healthy vegetation (chlorophyll) reflects more near-infrared (NIR) and green light compared to other wavelengths. But it absorbs more red and blue light. This is why our eyes see vegetation as the colour green. If you could see near-infrared, then it would be strong for vegetation too.

We receive radiation from the sun in a variety of wavelengths and intensities. Extraterrestrial radiation is the term used to describe the electro-magnetic solar radiation that strikes the upper edge of the atmosphere. 1,367 W/m2 is the mean integral for the entire spectrum (the Solar Constant). Infrared radiation is typically measured in micrometres (10^{-6} m), while the wavelength of solar and atmospheric radiation is typically measured in nanometers (nm, 10^{-9} m). The Table 3.3 below displays the range.

Table 3.3 Wavelengths of solar and atmospheric radiation

Wavelengths of Solar and Atmospheric Radiation for Meteorological Applications			
Short- wave	UV-C	100-280nm	Emitted from the sun, totally absorbed by the earth's atmosphere before reaching the ground
	UV-B	280- 315 nm	Emitted from the sun, 90% absorbed by the earth's atmosphere but biologically very active, causes sunburn
	UV-A	315- 400 nm	Emitted from the sun. most reaches the ground but not

			biologically very active
	Visible	400- 780 nm	Visible light from violet to red (colours of the rainbow)
Long wave (infrared)	NIR	780 nm - 3 µm	Heat radiation from the sun
	FIR	3 µm- 50 µm	Heat radiation from the atmosphere, clouds, earth and surroundings

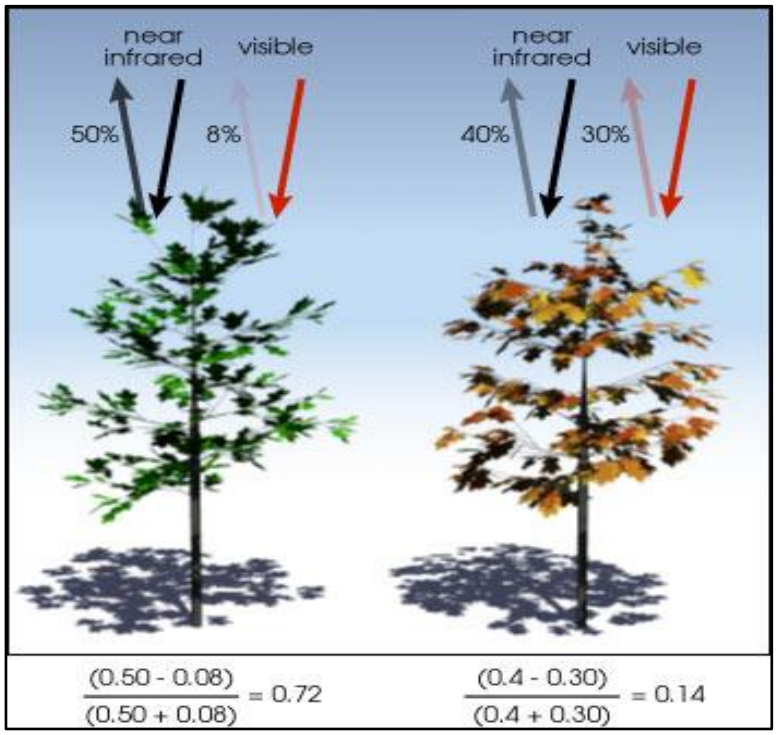


Fig. 3.5 Example of NDVI extraction
Source: (up42.com/blog/tech/5-things-to-know-about-ndvi)

The Fig. 3.5 shows an example of NDVI extraction procedure. The result of an NDVI calculation for a given pixel is always a number between minus one (-1) and plus one (+1); however, the absence of green leaves yields a value that is very close to zero. A value of zero denotes the absence of vegetation; while a value near +1 (0.8–0.9) denotes the greatest possible density of green leaves. Higher value of NDVI refers to healthy and dense vegetation. Lower NDVI values show sparse vegetation.

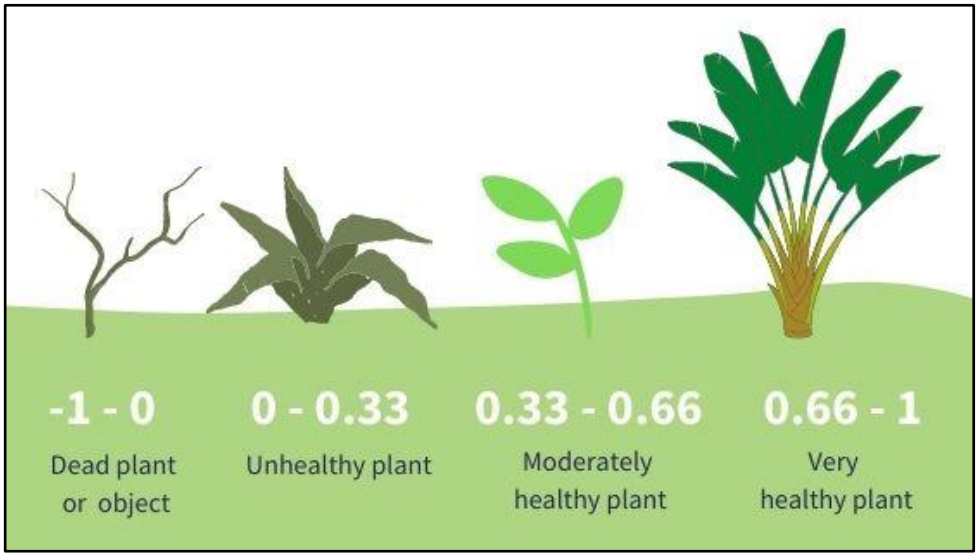


Fig. 3.6 NDVI indices indicating different types of vegetation
Source: (up42.com/blog/tech/5-things-to-know-about-ndvi)

3.2.3 Data Preprocessing

Preparing raw data to be suitable for a machine learning model is known as data pre-processing. In order to build a machine learning model, it is the first and most important step. It is not always the case that we come across the clean and formatted data when developing a machine learning project. Additionally, any time you work with data, you must clean it up and format it. Therefore, we use a data preprocessing task for this.

Real-world data typically includes noise, missing values, and may be in an unusable format, making it impossible to use machine learning models on it directly. Data preprocessing is necessary to clean the data and prepare it for a machine learning model, which also improves the model's accuracy and effectiveness.

3.2.4 Adfuller test

The Augmented Dickey Fuller Test (ADF) is a stationarity unit root test. Your time series analysis may yield unexpected results if you use unit roots. With serial correlation, the Augmented Dickey-Fuller test can be applied. Compared to the Dickey-Fuller test, the ADF test is more potent and capable of handling more complicated models. However, because it has a relatively high Type I error rate, like the majority of unit root tests, it should be used with caution **(Harris, 1992).**

The stationary component of a time series is crucial. Because a model cannot forecast on non-stationary time series data, the first step in ARIMA (autoregressive integrated moving average.) time series forecasting is to ascertain how many differencing are necessary to

make the series stationary. A stationary series is one whose statistical characteristics such as mean, variance, and covariance, do not change over time or are not affected by time. In other words, series without a Trend or Seasonal component are said to be Stationary in Time Series. Now the question comes that why a time series should be stationary? So it should be stationary because statistical models can more easily and accurately predict stationary series.

Types of Stationary

1. Series Strict Stationary – The mean, variance, and covariance are not time-dependent, and the process satisfies the mathematical definition of a stationary process.
2. Seasonal Stationary – Series shows seasonality.
3. Trend stationary – Series shows trend.

The following presumptions are used when conducting the ADF test.

1. Series is non-stationary or series has a unit root, according to the null hypothesis.
2. A second possibility is that the series is stationary or that it lacks a unit root.

This test may show that the series is not stationary if the null hypothesis cannot be proved.

If

Test statistic < Critical Value and p-value < 0.05

Reject Null Hypothesis i.e**.,** time series does not have a unit root, meaning it is stationary. It does not have a time-dependent structure. The snapshot of code used for Adfuller test is referred below.

```
[ ]  from pandas import read_csv
     from statsmodels.tsa.stattools import adfuller
     X = df['NDVI'].values
     result = adfuller(X)
     print('ADF Statistic: %f' % result[0])
     print('p-value: %f' % result[1])
     print('Critical Values:')
     for key, value in result[4].items():
       print('\t%s: %.3f' % (key, value))

     ADF Statistic: -3.175837
     p-value: 0.021427
     Critical Values:
             1%: -3.444
             5%: -2.867
             10%: -2.570
```

Fig. 3.7 Code for Adfuller test and outputs

As the obtained p-value and critical values are less than 0.05 i.e. 0.02 and -2.570, so it clearly denotes that time series used is stationary.

3.2.5 Data splitting

```
[4]  train_size = int(len(df) * 0.8)
     test_size = len(df) - train_size
     train, test = df.iloc[0:train_size], df.iloc[train_size:len(df)]
     print(len(train), len(test))

     407 102
```

Fig. 3.8 Train and test data split

The dataset has been split into training set and testing set. This split is 8:2 i.e. 80% of the data set was fed to the network for training and learning the model. Rest 20% is used to test the model's accuracy. As we have a total of 509 data values so total training values are 407 and 102 data values are for testing.

3.2.6 Reshaping of data

We have taken 'timesteps = 5', it means that the inputs to the model are the data for the first 4 timesteps, and the output will be the data for the 5th timestep.

3.3 Machine learning

Machine learning, a developing technology, allows computers to automatically learn from historical data. It is a subset of artificial intelligence. It uses a variety of algorithms to create mathematical models and make predictions based on previous information or data. Currently, it is used for a wide range of tasks, including recommender systems, email filtering, image and speech recognition, and many others.

Machine learning algorithms create a mathematical model with the aid of historical sample data, or "training data," that aids in making predictions or decisions without being explicitly programmed. Computer science and statistics are combined with machine learning to create predictive models. Algorithms that learn from historical data are created by machine learning or used in it. The level of performance will increase as we provide more information.

Imagine that we have a complex problem that requires some predictions. Instead of writing code for it, we can simply feed the data to generic algorithms, and the machine will build the logic according to the data and predict the output. Our perspective on the issue

has changed as a result of machine learning. The machine learning algorithm's operation is explained in the block diagram below:

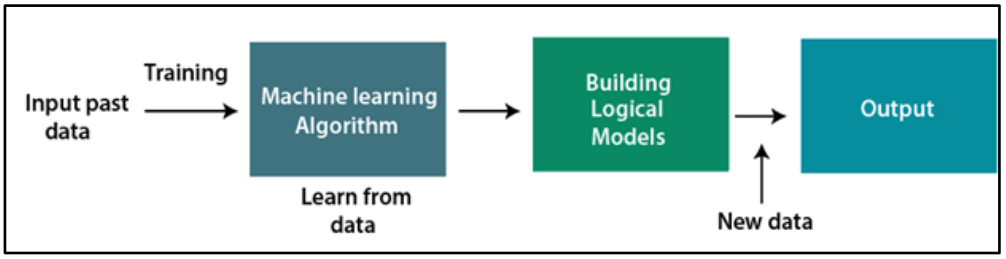


Fig. 3.9 Basic structure of machine learning

3.3.1 Supervised Machine Learning

A feature of pattern recognition called supervised learning makes use of a set of labelled instances that are used as training data and produce the desired results. In the training phase, a predictive model is created using the labelled examples to categorise new datasets. The labelled instances are fed into a specific machine learning algorithm to accomplish this. In supervised learning, mapping input and output data is the main objective. The foundation of supervised learning is supervision, just like when a student is learning under a teacher's supervision. Some of the supervised machine learning approaches is Naive Bayes, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Hidden Markov Model (HMM), and Decision Trees like the C4.5 and ID3 algorithms are a few examples of supervised machine learning techniques.

3.3.2 Unsupervised Machine Learning

Unsupervised learning works by discovering patterns in an unlabelled dataset used as the training data in order to make the rightful classification decisions in a set of new instances. This usually involves the use of clusters to identify the classes to which instances belong.

3.3.3 Semi-Supervised Machine Learning

Semi-supervised learning combines the power of both supervised and unsupervised learning approaches in the process of building a model for classifying new instances of a dataset. With this approach, good results are produced in terms of a high detection rate and low false positive rate.

3.3.4 Reinforcement Machine Learning

One subfield of machine learning is reinforcement learning. It involves acting appropriately to maximise reward in a specific circumstance. It is used by a variety of programmes and machines to determine the best course of action to take in a given circumstance. There is no correct answer in reinforcement learning, but the reinforcement agent decides what to do to complete the task. This is different from supervised learning, where the training data includes the answer key and the model is trained with that answer. It is obligated to gain knowledge from its experience in the absence of a training dataset. The best solution is decided based on the maximum reward.

3.4 Long short-term memory (LSTM)

Long Short-Term Memory Network is an advanced RNN, a sequential network that allows information to persist. It is capable of handling the vanishing gradient problem faced by RNN **(Yu *et al.* 2019)**. A recurrent neural network also known as RNN is used for persistent memory.

An advanced method adapted from ANN called the LSTM network was taken into consideration for predicting the dynamics of the vegetation using NDVI. The operation of neurons in the human brain served as the model for the development of ANN. Neurons and links that connect the nodes make up an ANN. Nodes process the information and output the information received by them to other neurons through the links connecting the nodes. Each link has a set of weights attached to it. By changing the weights attached to the links, ANNs can learn all of the functional relationships in the data. Because of this characteristic, ANNs are well suited for situations in which there is a surplus of data but insufficient understanding of the data. Additionally, they are very helpful in predicting time series and perform fairly well even with noisy data because they learn the hidden dependencies in the data based on previous values. The RNN has been modified to form the LSTM network. RNN is a feed-forward network with an internal memory and feedback loop. Similar to humans, they retain information in a sequential order. In contrast to the traditional feed forward network, they keep this sequence in their internal memory for a long time, which aids in making decisions about the following step accurately. In contrast to RNN, the recurrent structure in LSTM is a gated structure. LSTMs are better suited for time series prediction than RNN because they have a longer-term memory. Furthermore, the introduction of the memory cell concept in LSTM networks solves the vanishing gradient descent problem in a traditional RNN. Gates in a memory cell determine what data should be read from, written to, or stored in the network. They hold onto the back-propagated error

for a very long time. They are ideal for time series prediction because they can learn from over 1000-time steps. The four main parts of a memory cell are a candidate hidden state, an input gate, an output gate, and a forget gate. A sigmoid layer makes up these gates. If the output of the sigmoid layer is not zero, they permit the information to pass through them. No information is passed through the gate if the output is zero; however, if it is one, then all information is passed through the network.

3.4.1 LSTM Architecture

LSTM functions on a high level very similarly to an RNN cell. The LSTM network's internal operation is shown here. As seen in the image below, the LSTM is composed of three parts, each of which has a distinct function.

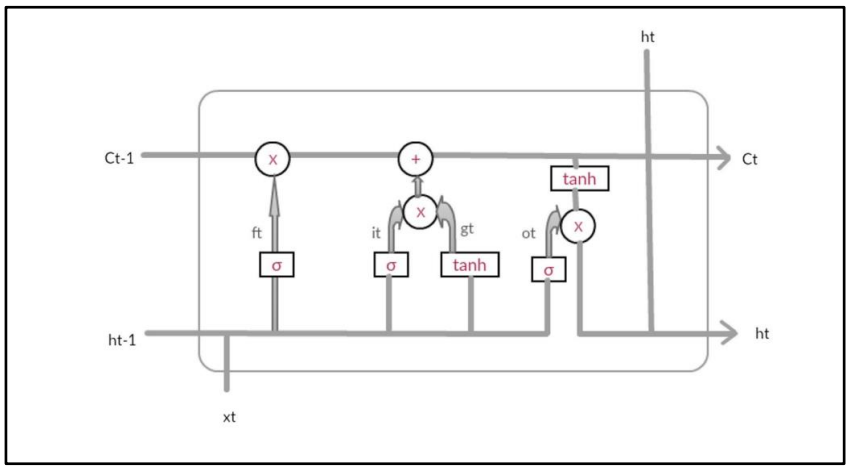


Fig. 3.10 Internal gated structure of memory cell in LSTM
Source: (Reddy and Prasad, 2018)

A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The three gates control the flow of information into and out of the cell, and the cell remembers values over arbitrary time intervals.

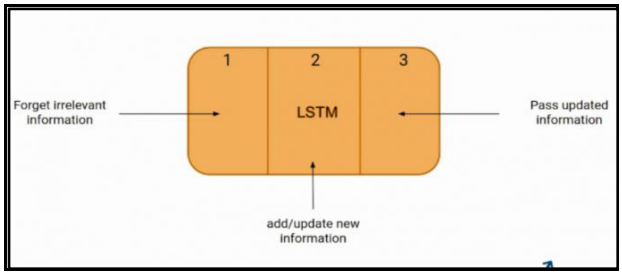


Fig. 3.11 LSTM primary structure
Source: (cdn.analyticsvidhya.com/wp-content/uploads/2021/03/Screenshot-from-2021-03-16-13-26-39-768x331.png)

The first section determines whether the information from the previous timestamp needs to be remembered or can be ignored. The cell attempts to learn new information from the input to this cell in the second section. The cell finally transmits the updated data from the current timestamp to the following timestamp in the third section.

The three components of an LSTM cell are referred to as gates. The Forget gate is the first component, the Input gate is the second, and the Output gate is the final component.

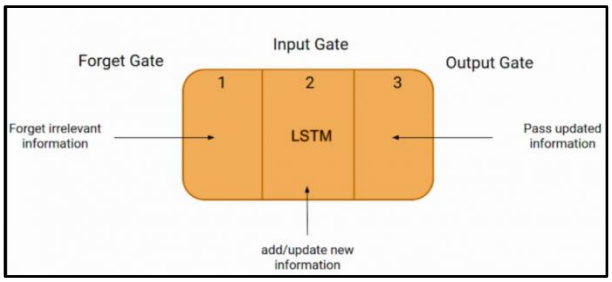


Fig. 3.12 LSTM unit showing different gates
Source: (cdn.analyticsvidhya.com/wp-content/uploads/2021/03/Screenshot-from-2021-03-16-13-41-03-768x342.png)

An LSTM has a hidden state, just like a straightforward RNN, with H(t-1) standing for the hidden state of the prior timestamp and Ht for the hidden state of the present timestamp. Additionally, LSTMs have a cell state that is denoted by the timestamps C(t-1) and C(t), which stand for the previous and current timestamps, respectively. In this case, the cell state is referred to as the long-term memory and the hidden state as the short-term memory. It is interesting to note that the cell state carries the information along with all the timestamps. The gates shown in Fig. 3.12 have their own meaning and work. They are defined below.

3.4.1.1 Forget Gate

In a cell of the LSTM network, the first step is to decide whether we should keep the information from the previous timestamp or forget it. So, it is done by forget gate of the LSTM network.

3.4.1.2 Input Gate

Input gate is used to quantify the importance of the new information carried by the input. We want the model to remember only the relevant data. This is the main task of the input gate.

3.4.1.3 Output Gate

Based on the current expectation the model has to give a relevant word to fill in the blank or to guess the future element. That word is our output and this is the function of our Output gate.

Based on this architecture of the LSTM network, this is how the models are trained. The relevant information is gathered by the gates else forget gate discards those values. The input gate lets the model add the upcoming relevant information to the previous data. These values are later used by the output gate to predict the future values based on the learned model.

Advantages of LSTM

- The constant error back propagation within memory cells results in LSTM's ability to bridge very long-time lags.
- For long time lag problems, LSTM can handle noise, distributed representations, and continuous values.
- LSTM quickly learns to distinguish between two or more widely separated occurrences of a particular element in an input sequence, without depending on appropriate short-term lag training exemplars.

3.4.2 LSTM Models

To study and assess the accuracy of the LSTM model, we have taken two simultaneous models. Both are LSTM models but with different orientations. We are going to assess that which one works best with our dataset. In the end of the study we are going to compare both models' outputs and name the better of the two. Model-1 has been built with just single ReLU. While model-2 have two ReLU functions.

```
[11] model = keras.Sequential()
     model.add(keras.layers.LSTM(128,return_sequences=True,input_shape=(X_train.shape[1], X_train.shape[2])))
     model.add(keras.layers.LSTM(128,activation='relu',return_sequences=True))
     model.add(keras.layers.LSTM(128))
     model.add(keras.layers.Dense(1))
     model.compile(loss='mean_squared_error', optimizer=keras.optimizers.Adam(0.001))
```

Fig. 3.13 Model-1

```
[ ]  model = keras.Sequential()
     model.add(keras.layers.LSTM(128,return_sequences=True,input_shape=(X_train.shape[1], X_train.shape[2])))
     model.add(keras.layers.LSTM(128,activation='relu',return_sequences=True))
     model.add(LSTM(units =128, activation='relu', return_sequences=False))
     model.add(Dropout(0.2))
     model.add(keras.layers.Dense(1))
     model.compile(loss='mean_squared_error', optimizer=keras.optimizers.Adam(0.001))
```

Fig. 3.14 Model-2

3.4.2.1 ReLU Activation function

Without an activation function, a neural network is essentially just a linear regression model. Thus, we apply a nonlinear transformation to the neuron's inputs, and an activation function introduces this nonlinearity into the network. Although the neural network is made simpler by linear transformations, it is less powerful and is unable to learn complex patterns from the data.

When we think of an activation function, the first thing that comes to mind is a threshold-based classifier, which determines whether or not to activate the neuron based on the result of the linear transformation. In other words, a neuron is activated if the input to the activation function exceeds a threshold; otherwise, it is deactivated, meaning that its output is not taken into account for the next hidden layer.

A non-linear activation function that has grown in popularity in the deep learning field is the ReLU function. Rectified Linear Unit is referred to as ReLU. The ReLU function's primary advantage over other activation functions is that it does not simultaneously activate all of the neurons.

In the case of classifiers, sigmoid functions and their combinations typically perform better. Due to the issue with the vanishing gradient, sigmoids and tanh functions are occasionally avoided. When compared to models trained on the Sigmoid and tanh function, the model trained with ReLU converged quickly and therefore takes much less time. In the model developed using ReLU, overfitting is clearly visible. This is as a result of the rapid convergence. When using ReLU for model training, the model's performance is noticeably improved.

Today, the majority of cases use the ReLU function, which is a general activation function. The leaky ReLU function is the best option if we come across a case of dead neurons in our networks. ReLU function should only be used in hidden layers, keep in mind. As a general rule, you should start by using the ReLU function and then switch to other activation functions if ReLU doesn't yield the best results.

3.4.2.2 Adam Optimizer

When a neural network is being trained, its weights are initially initialised randomly and then they are updated in each epoch in a way that increases the network's overall accuracy. The output of the training data is compared to the actual data in each epoch with the aid of the loss function to determine the error, and the weight is then updated appropriately. But how can we update the weight in a way that improves accuracy? This is known as optimization problem. And the objective of this problem is to optimise the loss function and find the ideal weights. The process of optimization is referred to as Optimizer. The stochastic gradient descent method is used by the Adam algorithm, which is used by the Adam optimizer, to carry out the optimization process. It uses very little memory and is effective in use. When a large amount of data and parameters are available for use, it is appropriate.

The most well-known and frequently employed optimizer for neural network training is Keras Adam.

Upon applying the proposed model, the training and validation loss graphs has been plotted to observe the fit of the model. Depending upon the validation loss and training loss different kinds of fits have been defined. Following are the detail explanation about the conditions of particular fit.

3.4.3 Underfitting

Sometimes the validation loss exceeds the training loss. This may indicate that the model is underfitting. When a model is unable to accurately model the training set of data, underfitting occurs and large errors are produced. In order to lessen the loss experienced during training, more training is required. Additionally, we can increase the training data by either acquiring more samples or enhancing the existing data.

3.4.4 Overfitting

The validation loss in overfitting is greater than the training loss. This typically means the model is overfitting and unable to generalise to fresh data. The model performs particularly well on training data but poorly on the fresh data in the validation set. The validation loss initially decreases but then gradually increases again.

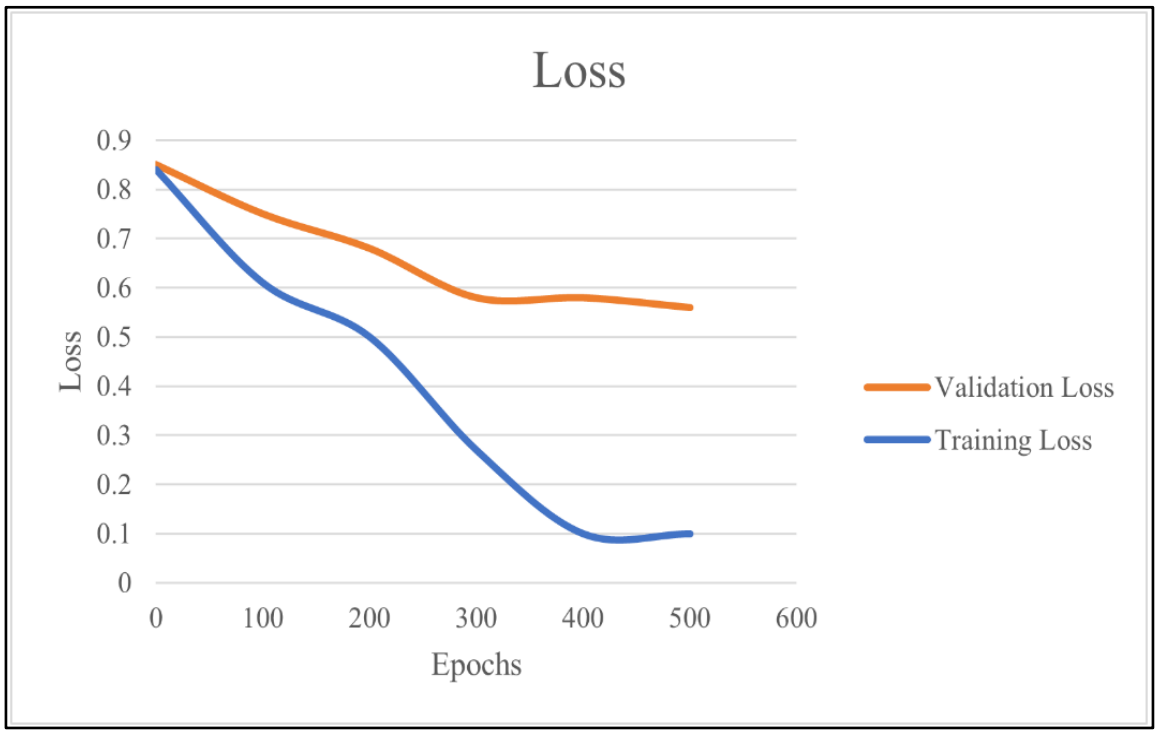


Fig. 3.15 Underfitting graph

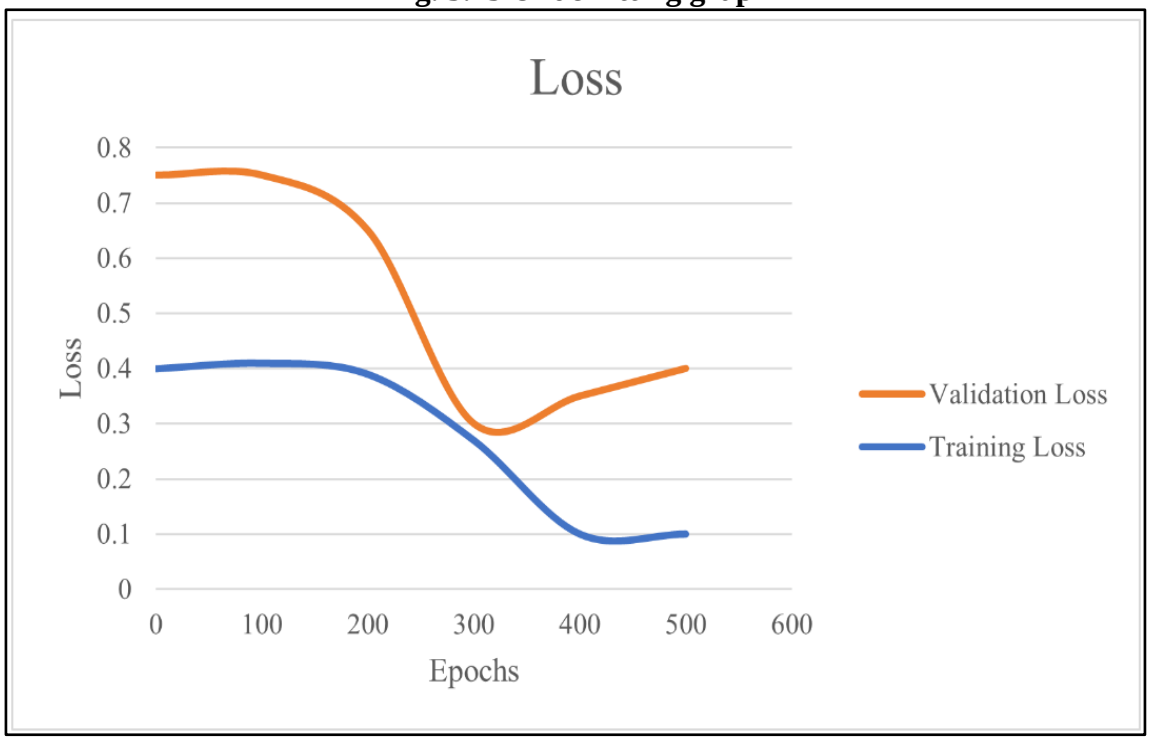


Fig. 3.16 Overfitting graph

The model may be too complex for the data, or it may have been trained for a long time, as a notable cause of this occurrence. When the loss is small and stable, training can be stopped—a technique known as early stopping—in this situation. One of the many strategies used to avoid overfitting is early stopping.

3.4.5 Good fit

At some point, both the training loss and the validation loss start to decline and stabilise. This is the condition of good fit for a model.

Because the learning algorithm did not have enough data to draw conclusions from, accuracy on training and test data might be subpar. Performance could be enhanced by the following:

- The number of training data examples should be increased.
- Increase the number of times the current training data is passed.

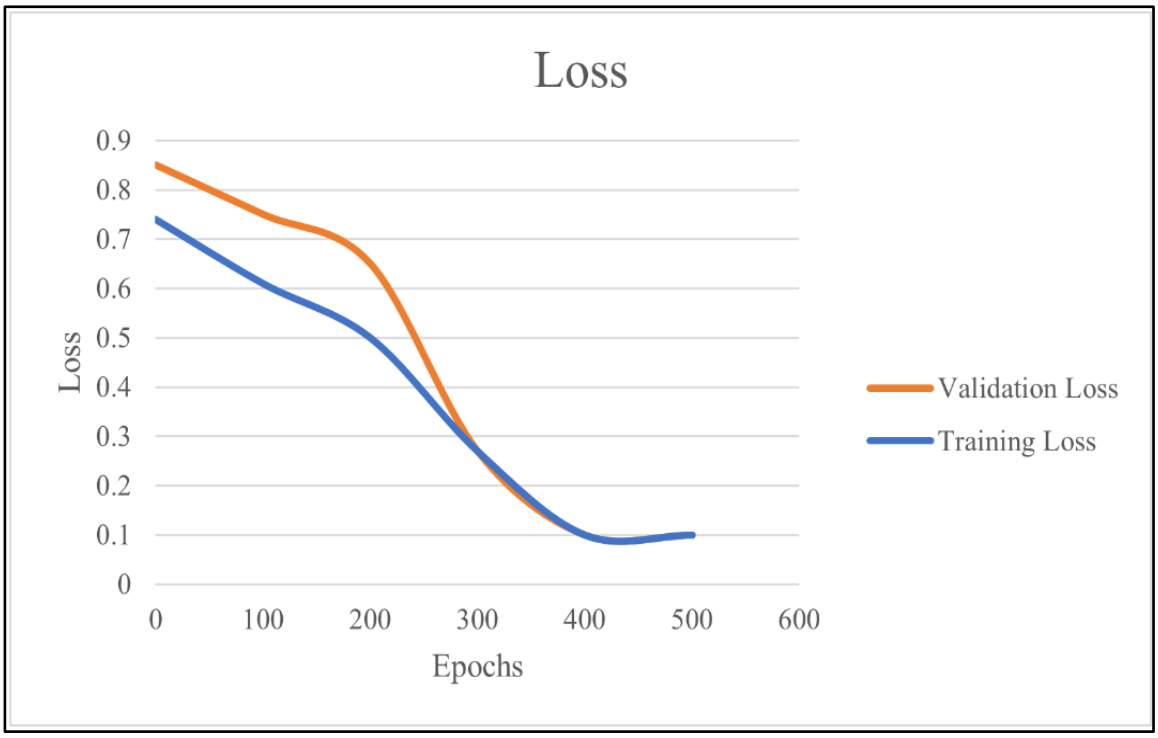


Fig. 3.17 Good fit graph

3.4.6 Mean Absolute Error (MAE)

A model evaluation metric used with regression models is mean absolute error. The average of the absolute values of each prediction error across all test set instances is the mean absolute error of a model with respect to the test set. The difference between the instance's true value and the predicted value represents each prediction error.

$$MAE = \frac{\sum_{i=1}^{n} abs(yi - \lambda(\mathrm{xi}))}{n}$$

Where n is the number of test instances, y_i is the actual target value for test instance x_i and $\lambda(x_i)$ is the predicted target value for test instance x_i.

3.4.7 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). The distance between the data points and the regression line is measured by residuals, and the spread of these residuals is measured by RMSE. In other words, it provides information on how tightly the data is clustered around the line of best fit. In climatology, forecasting, and regression analysis, root mean square error is frequently used to validate experimental results.

The formula is:

$$RMSE = \sqrt{\overline{(f-o)^2}}$$

Where:

f = forecasts (expected values or unknown results),

o = observed values (known results).

The bar above the squared differences is the mean (similar to $\bar{x}$). The same formula can be written with the following, slightly different, notation.

$$RMSE_{fo} = \left[\sum_{i=1}^{N}(z_{fi} - z_{oi})^2/N\right]^{1/2}$$

Where:

Σ = summation ("add up")

$(z_{fi} - Z_{oi})^2$ = differences squared

N = sample size.

RMSE of test > RMSE of train => OVER FITTING of the data.

RMSE of test < RMSE of train => UNDER FITTING of the data.

3.4.8 R^2(r square)

Basically, R^2 value is the percentage of the dependent variable variation that a linear model explains.

$$R^2 = \frac{variance\ explained\ by\ the\ model}{total\ variance}$$

R-squared is always between 0 and 100%:

0% represents a model that does not explain any of the variation in the response variable around its mean. The mean of the dependent variable predicts the dependent variable as

well as the regression model.100% represents a model that explains all of the variation in the response variable around its mean. Usually, the larger the R^2, the better the regression model fits your observations.

The plots of fitted values against observed values can be seen visually through a graphical representation. Fig. 3.18 demonstrates how the scatter around the regression line is represented by R-squared values.

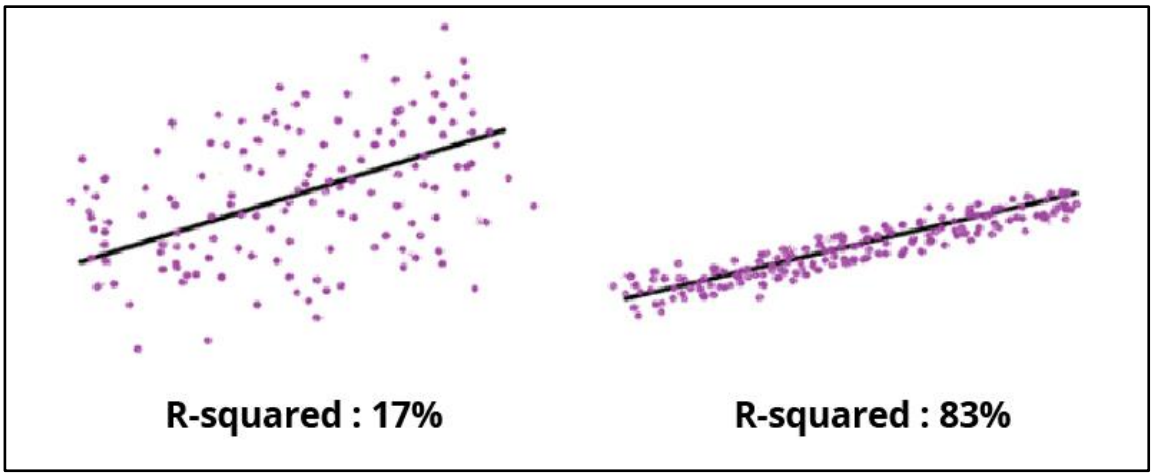


Fig. 3.18 Example of R-square

As seen in the images above, the regression model on the left has an R-squared value of 17%, while the model on the right has an R-squared value of 83%. The data points in a regression model typically fall closer to the fitted regression line when the variance is high.

However, an ideal case that is actually not attainable is a regression model with an R^2 of 100%. This results in all the data points falling precisely on the regression line because the predicted values are equal to the observed values.

3.5 Hardware used

In order to accomplish the proposed work a high-performance computer is required which has the following hardware specifications:

CPU Cores: 6 Threads: 24

GPU Cores: 10, 240

Boost Clock: 1,665MHz

Base Clock: 3.2GHz P-core, 2.4GHz E-core

Boost Clock: 5.2GHz P-core, 3.9GHz E-core

L3 Cache (smart): 30MB

L2 Cache: 14MB

Memory: 8GB GDDR6X,

Memory speed: 19Gbps

Memory Bandwidth: 912GB/s

Primary Memory: Capacity: 16GB, Speed: 6400MHz

Or we can use a comparable cloud computing service. For our project, we opted to use Google Earth Engine.

3.6 Software used

The data extraction and visualization of the proposed work was done in Google Earth Engine Code Editor. The programming language that was used here to process the Spatial Dataset files was JavaScript. Below is the snapshot of Google earth engine code editor. Extraction of NDVI values and generation of time series is done using Google earth engine only. The NDVI changes have also been demonstrated using this software. The NDVI values are extracted in excel using .csv files. The plot of time series was exported in .PNG format.

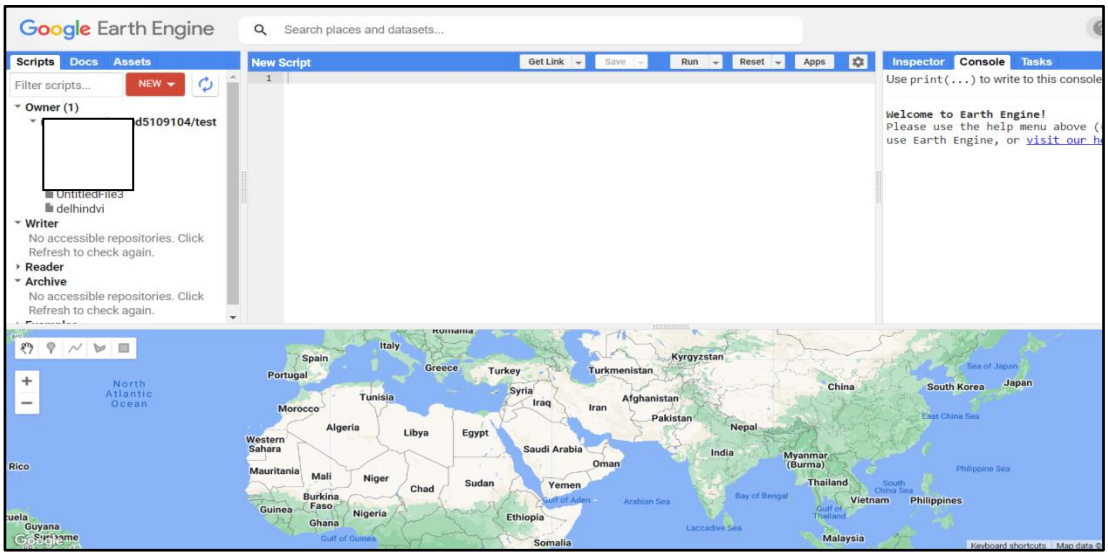


Fig. 3.19 Google earth engine code editor
Source: (GEE)

After processing the output from the Google earth engine code editor, it was imported to Google colab software for further analysis and visualization. The data files obtained from the earth engine have been processed and modified in colab using python programming language. The LSTM model has been applied on those data files and further analysis has been done. Following figure shows the snapshot of colab notebook.

Fig. 3.20 Colab code editor
Source: (Google colab)

Chapter 4 RESULTS AND DISCUSSION

To access the environmental variables of urban areas, especially mega cities like Delhi, can be assessed by understanding various components of the city and its spatial pattern such as land use, vegetation, built-up areas, temperature variations, impervious surface **(Prafull *et al.*, 2022)**. The dynamics of land use and land cover over time determine a city's capacity to successfully balance its expansion with its natural ecosystem. Changes in the amount of open space, agricultural land, and urban areas are good indicators of urban growth, but they also have negative effects on the viability of the urban area.

The result analysis has been divided into two phases. During the first phase the comparative study has been carried out across different time period. It was carried out based on the maps, charts and graphs generated from Google earth engine. The complete dataset from the MODIS terra satellite has been fed to Google earth engine. Based on the data, the NDVI images and NDVI time series of the Delhi are generated. The NDVI images that have been generated for different time period has been used for the change analysis.

NDVI time series has been used in the second phase of the analysis. Here the time series was imported to the 'Colab Notebook'. Python language was used for predicting the future values of the NDVI for Delhi region. Two different Long short term memory (LSTM) network models has been used to learn the series and based on the same the future values have been predicted. Later at the end of the study we have compared and the results and accuracy of both the models. Based on the performance one has been claimed better than the other model.

The findings clearly demonstrate that the LSTM network is capable of accurately learning all seasonal and annual changes in vegetation, changes brought on by deforestation or natural disasters, etc. in order to predict future NDVI values. Both the models predicted the future NDVI with acceptable accuracy with slight differences. As a result, it is advised to use an LSTM network to predict vegetation dynamics with reliability and accuracy. The accuracy is much higher when the LSTM network is trained using large amounts of data. Because of this, LSTM is a much more effective and distinctive method than other methods for predicting NDVI dynamics.

4.1 NDVI visualization and change analysis

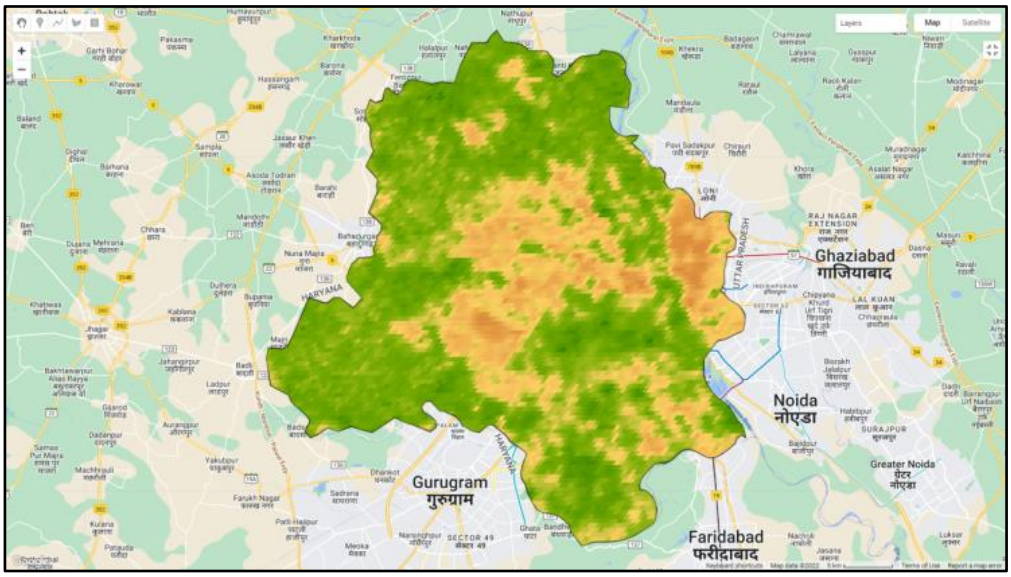


Fig. 4.1 NDVI map for 2001

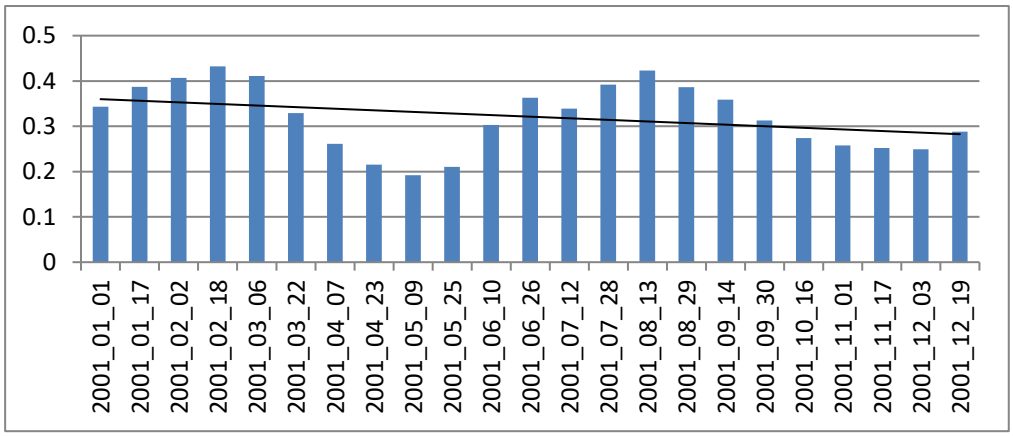


Fig. 4.2 16-day interval NDVI graph for year 2001

Table 4.1 NDVI table for year 2001

NDVI for Year 2001	
Minimum NDVI	0.192
Maximum NDVI	0.432

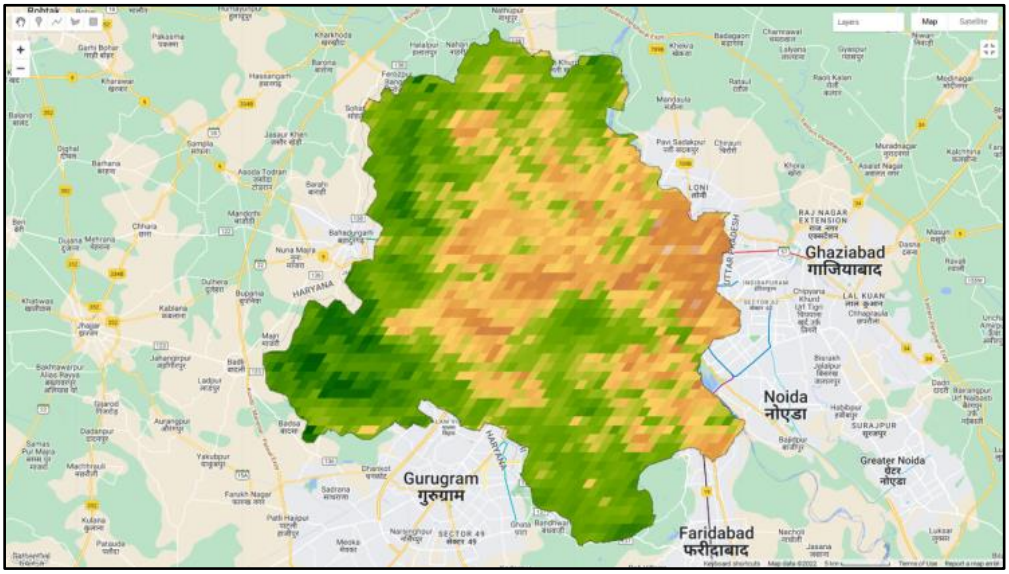


Fig. 4.3 NDVI map for 2010

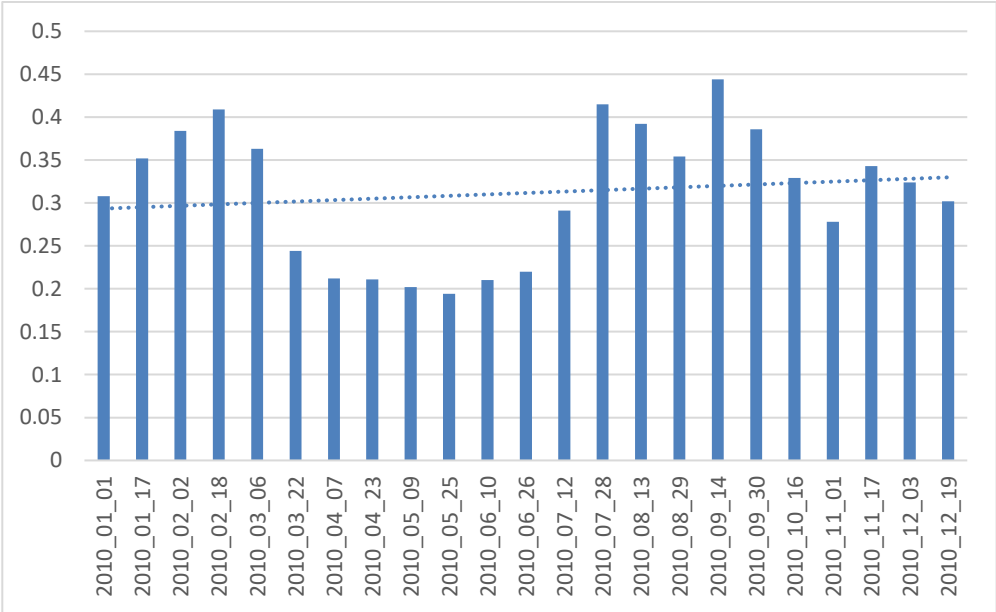


Fig. 4.4 16-day interval NDVI graph for year 2010

Table 4.2 NDVI table for year 2010

NDVI for Year 2010	
Minimum NDVI	0.194
Maximum NDVI	0.444

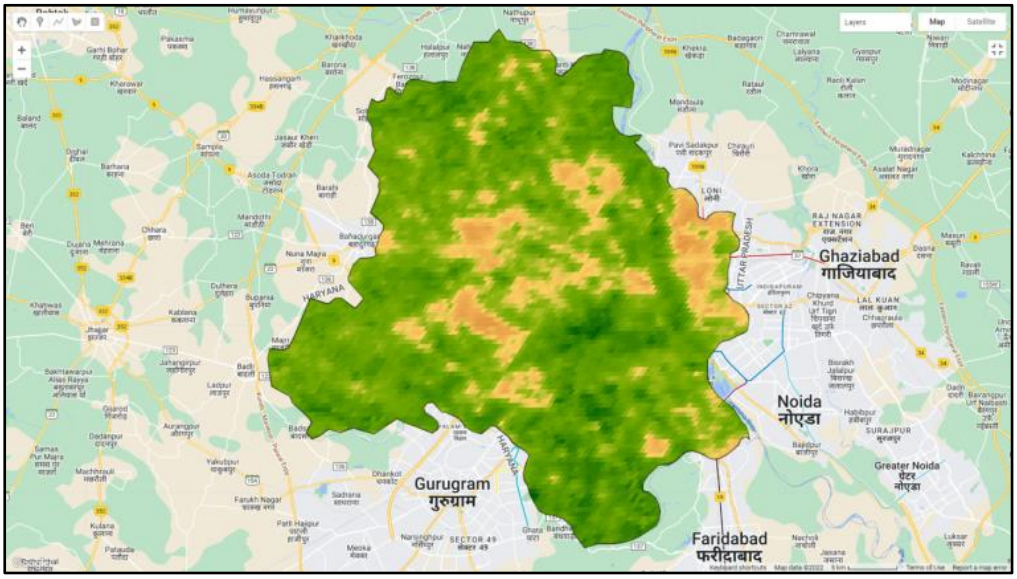


Fig. 4.5 NDVI map for 2021

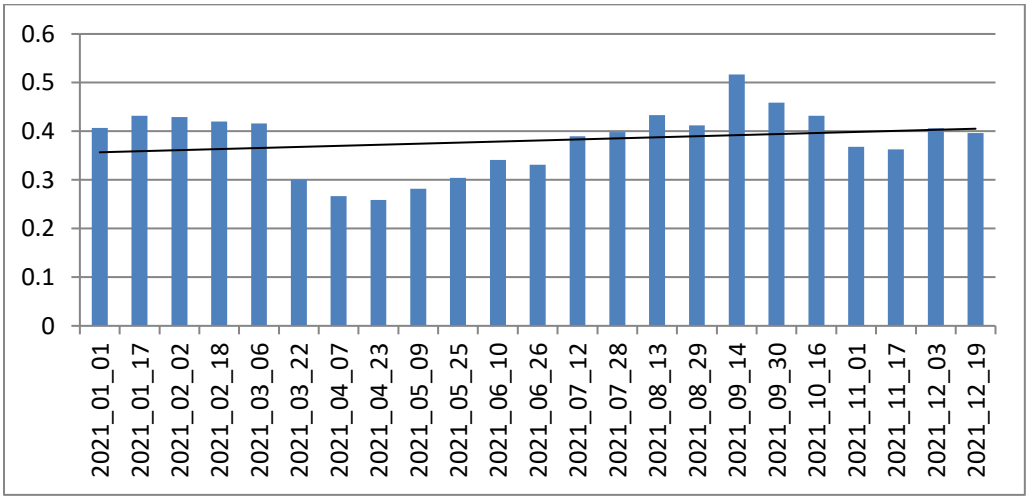


Fig. 4.6 16-day interval NDVI graph for year 2021

Table 4.3 NDVI table for year 2021

NDVI for Year 2021	
Minimum NDVI	0.259
Maximum NDVI	0.517

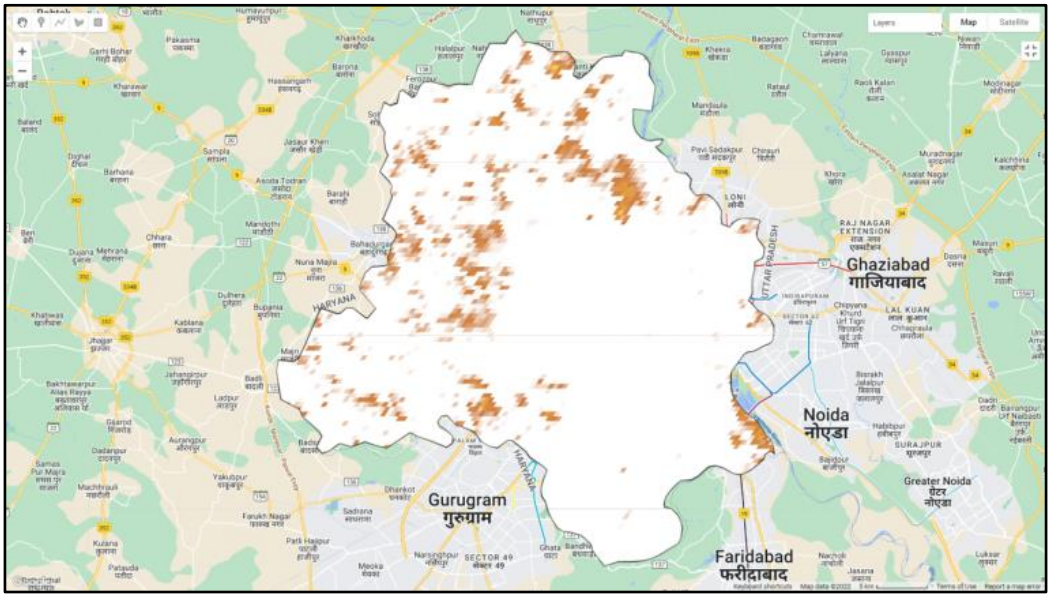


Fig. 4.7 Change in NDVI between year 2001 and 2021

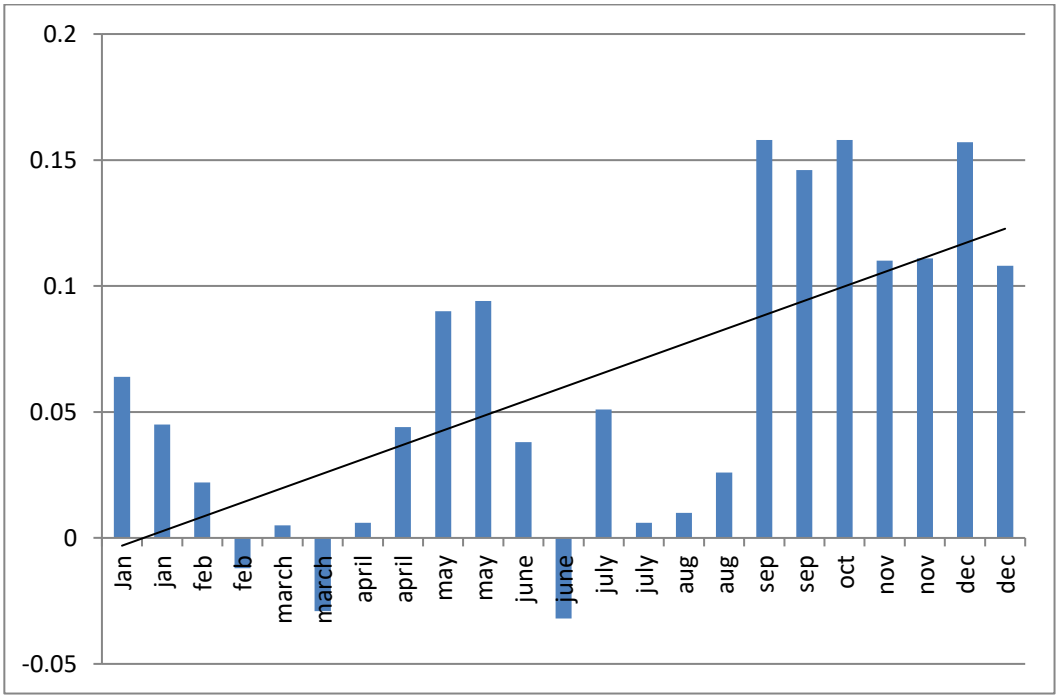


Fig. 4.8 NDVI difference graph between year 2001 and 2021

The above figures show NDVI v/s time graphs. The Fig. 4.2 shows the NDVI values across different months of the year 2001. Similarly, Fig. 4.6 shows the graph for year 2021. It can be clearly seen that the NDVI for year 2021 are greater for every month except for February, April and July. All the bar graphs plotted above are added with linear trend lines also. As the dataset taken here is at interval of 16 days, we are seeing two NDVI bars for each month. The highest average NDVI value for year 2001 was 0.432 for the month of

February. While for the year 2021, the highest NDVI value is 0.517. Similarly, for lowest values we can see that for 2001 is 0.192 while for 2021 it is 0.259. These stats clearly state that in recent year there has been significant positive change in vegetation of Delhi. Here also we can clearly see that every month have seen positive change except for Feb, April and July.

To verify the above statements, we have plotted the graph for average monthly NDVI difference (for year 2001 and 2021) with trend line. Total percentage change has been analysed using the average yearly NDVI values for year 2001 and 2021. It is calculated using the formula below:

$$\frac{total\ NDVI\ difference\ between\ 2001\ and\ 2021}{total\ average\ NDVI\ for\ year\ 2001} * 100\%$$

This equation gives the %bias which has been calculated in the table below. As the difference calculated here is by subtracting NDVI of 2001 from 2021. And the total average difference obtained is positive. This means that total vegetation of Delhi has improved over the years. From the table below sum of NDVI for year 2001 is 7.3860 and for year 2021 is 8.7620. The NDVI difference calculated here is 1.3760. Therefore, total percentage change seen here is 18.63%.

Table 4.4 Table showing total NDVI change from year 2001 to 2021

Sno.	NDVI for 2001	NDVI for 2021	Difference (2021-2001)	%BIAS
1	0.3430	0.4070	0.06	
2	0.3870	0.4320	0.05	
3	0.4070	0.4290	0.02	
4	0.4320	0.4200	-0.01	
5	0.4110	0.4160	0.01	
6	0.3290	0.3000	-0.03	
7	0.2610	0.2670	0.01	
8	0.2150	0.2590	0.04	
9	0.1920	0.2820	0.09	
10	0.2100	0.3040	0.09	
11	0.3030	0.3410	0.04	**18.6298**
12	0.3630	0.3310	-0.03	
13	0.3390	0.3900	0.05	
14	0.3920	0.3980	0.01	
15	0.4230	0.4330	0.01	
16	0.3860	0.4120	0.03	
17	0.3590	0.5170	0.16	
18	0.3130	0.4590	0.15	
19	0.2740	0.4320	0.16	
20	0.2580	0.3680	0.11	
21	0.2520	0.3630	0.11	
22	0.2490	0.4060	0.16	
23	0.2880	0.3960	0.11	
SUM	**7.3860**	**8.7620**	**1.3760**	

4.2 Time series forecasting and LSTM model's accuracy analysis

The time series derived from Google earth engine from the MODIS dataset has been used as the primary dataset for the time series analysis. The MODIS dataset consists of 509 NDVI values of different months of the years ranging from February 2000 to March 2022. The MOD13A1 Version 6 product provides Vegetation Index (VI) values at a per pixel basis at 500 meter (m) spatial resolution. The VI values taken here are NDVI at an interval of 16 days. This NDVI time series plot has been stated in the chart below:

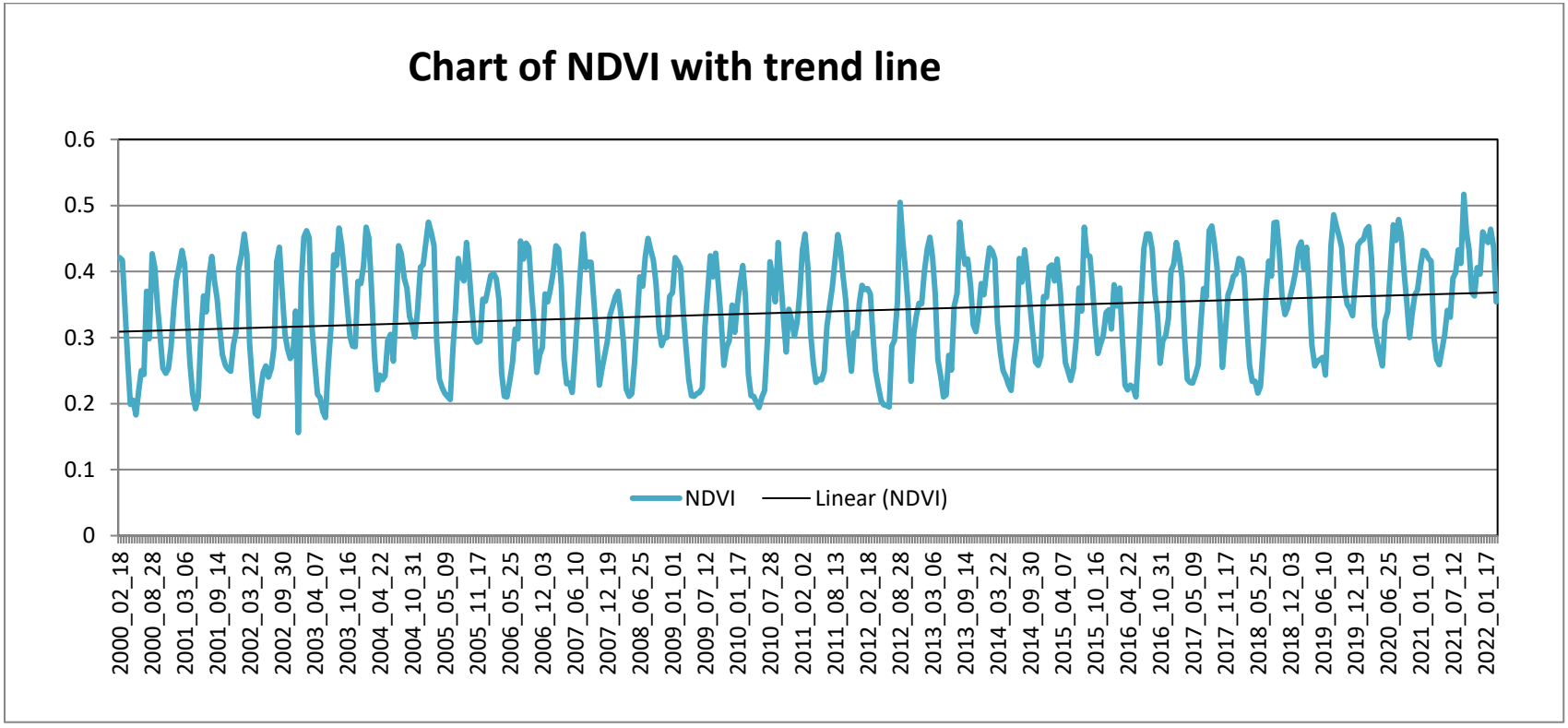


Fig. 4.9 NDVI time series for the time period 2000 to 2022

The Fig. 4.9 shows the NDVI data plot over the time period of 18th February 2000 to March 2022 for Delhi state. The time series clearly shows some seasonal effect as there are particular months showing high values of NDVI every year and it is same for lowest NDVI values also. The blue line is actual NDVI value plot over the years which have been referred here as the time series. The black line is linear trend line. It shows upward trend from year 2000 to year 2022. This indicates that NDVI of Delhi has increased over the years. Since increased NDVI means more green vegetation. So we can say that now the greenness of Delhi have increased compared to year 2000. This can also be seen in the visual maps of year 2001 and 2021 in Fig. 4.1 and Fig. 4.5 respectively.

The .csv file of the above time series is then used by the proposed LSTM models to forecast the future values. The LSTM models used here are model-1 and model-2 indicated in Fig. 3.13 and Fig. 3.14 respectively.

4.3 Training of dataset on the proposed LSTM model

The dataset of 509 NDVI values have been used for predicting the model's accuracy. The dataset has been divided into 8:2 ratios for training and testing purpose. 407 entries i.e. 80% of the total data was used for training and 102 i.e. 20% was used for testing the model. We have tried both the models with different number of epochs and same has been defined in the below table.Table 4.5 Metrics table corresponding different epochs for Model-1Table 4.5 shows the error metrics and r^2 values corresponding different epochs for model-1 whereas Table 4.6 shows for model-2. Model-1 shows best results for 70 epochs with RMSE 0.034 and R^2 of 0.7796 while model-2 gives better results for 65 epochs with RMSE of 0.0365 and R^2 of 0.7463.

Table 4.5 Metrics table corresponding different epochs for Model-1

No. of Epochs	MAE	RMSE	R^2
75	0.026501	0.034276	0.776634
80	0.026369	0.034332	0.775910
85	0.026710	0.034644	0.771837
70	**0.026239**	**0.034043**	**0.779662**
65	0.026478	0.034239	0.777118
60	0.026672	0.034465	0.774166

Table 4.6 Metrics table corresponding different epochs for Model-2

No. of Epochs	MAE	RMSE	R^2
75	0.03034	0.03900	0.710799
70	0.029278	0.03787	0.72732
65	**0.02855**	**0.036528**	**0.74631**
60	0.032216	0.040803	0.683477
80	0.02879	0.037706	0.72969
85	0.030602	0.0393264	0.70597

The above metrics show that model-1 has better accuracy than model-2 as it has greater R^2 and lesser RMSE value. Even MAE is lesser for model-1 than model-2.

For validation we have taken a small part of testing dataset. Following graphs Fig. 4.10 and Fig. 4.11 plot indicates how validation loss decreased significantly with each epoch. For model-1 we trained data for 70 epochs to train our model while for model-2 the 65 epochs have been taken. Batch size taken was of 10 and validation split was just 0.07. As the validation loss almost remained constant after 70 epochs so we stopped training our model 75 epochs.

Blue lines indicate the training loss with respect to number of epochs. We can see that training loss curve falls sharply for the initial epochs. After that it decreases slowly and then attains a constant value after certain epochs. The orange line indicates validation loss on test data. It also shows similar pattern and becomes constant after particular number of epochs.

Both training as well as validation loss stabilizes after a point, this means that the proposed model is neither under fit nor over fit. It is good fit for the dataset taken.

Model-1 attains a stable value of validation as well as training loss sooner than model-2. Even the loss for the starting epochs is greater in model-2 graph plot. This means that model-1 learns the data variations quickly than model-2. But both the models produce good fit lines for loss. So, we can say that LSTM model is good for predicting this type of NDVI time series. The accuracy of the models are very close that means there is no much difference if we are using a LSTM model.

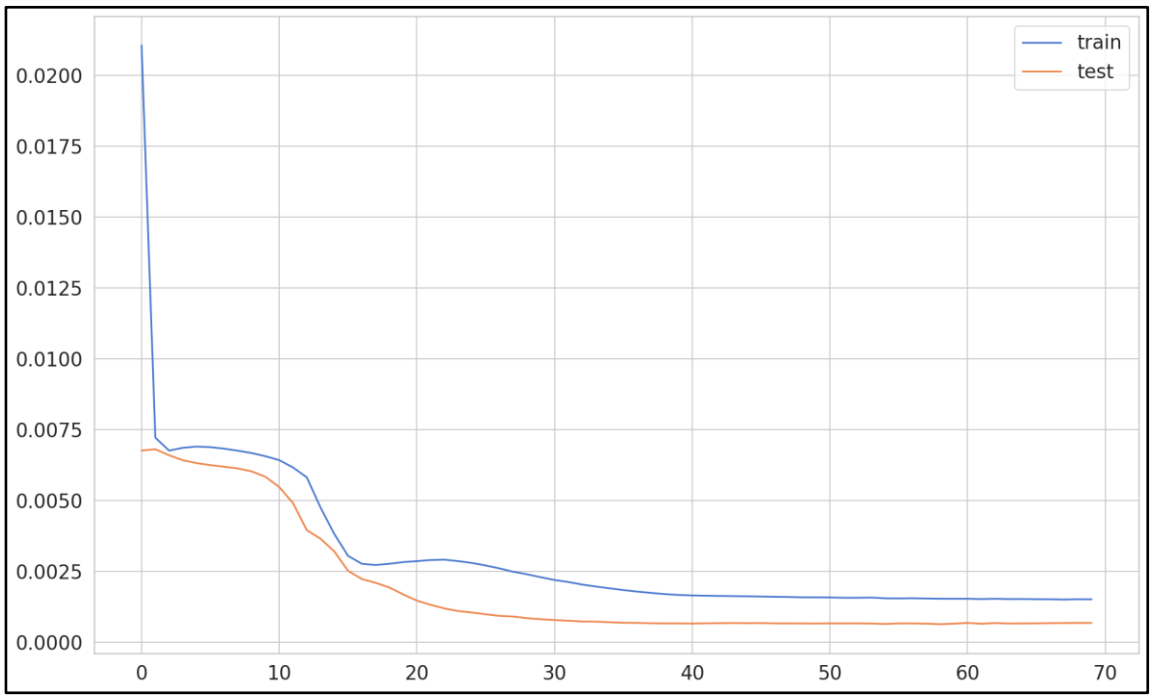


Fig. 4.10 Validation and training loss graph for model-1

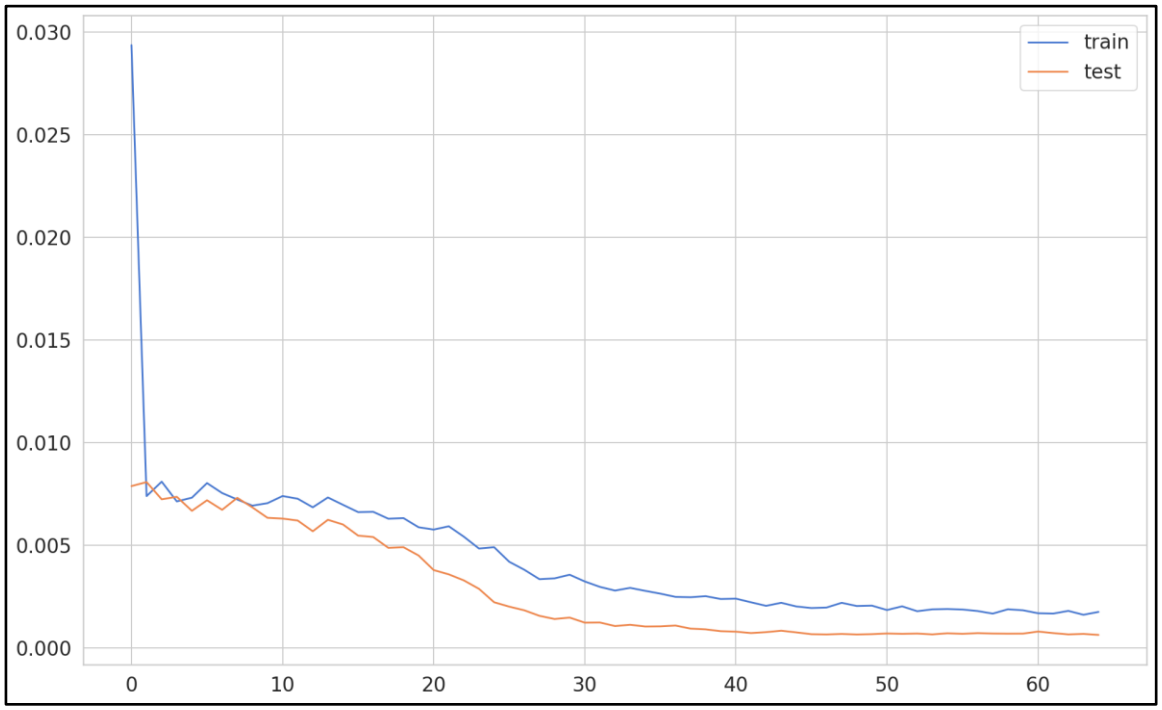


Fig. 4.11 Validation and training loss graph for model-2

4.4 Prediction

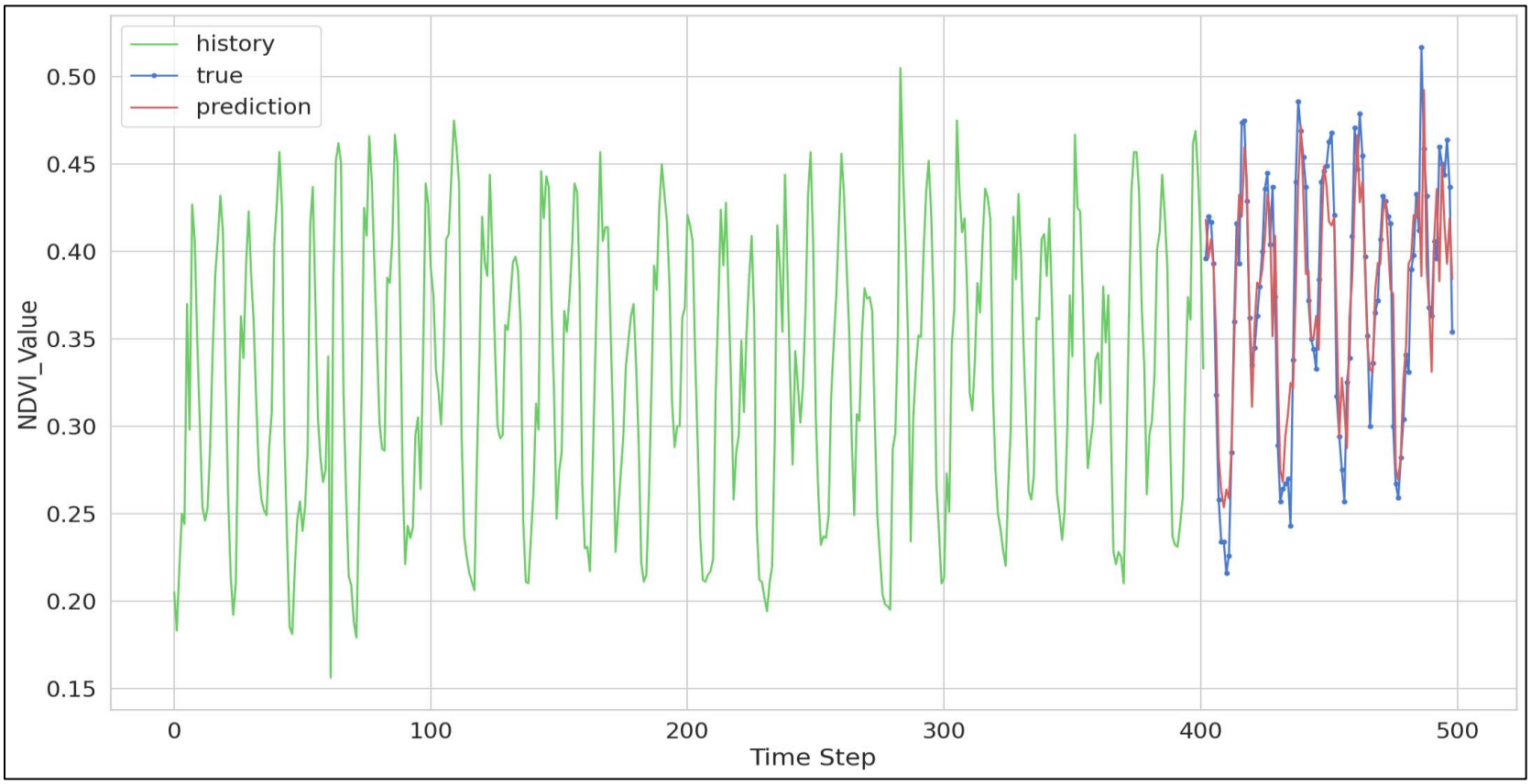


Fig. 4.12 Forecasted Time Series Plot for model-1

The above Fig. 4.12 shows the plot of observed NDVI values over the years plus the forecasted values by the model. The green part is the training data that has been ploted by the model and the blue lines are the test data values. Train and test data are the true values. The red line is the prediction made by the model after learning from the train data.

4.5 True VS Prediction

In the Fig. 4.13 and Fig. 4.14 below we have taken forecasted vs true time series by model-1 and model-2 respectively. The true line here means the real observed values we have fed to the model for the testing. While the red line is the forecasted time series by the models. We can see that the prediction line is very close to the true line. The peak and lowest points of both the lines are almost at the same place. This adds more confidence to the accuracy of the model.

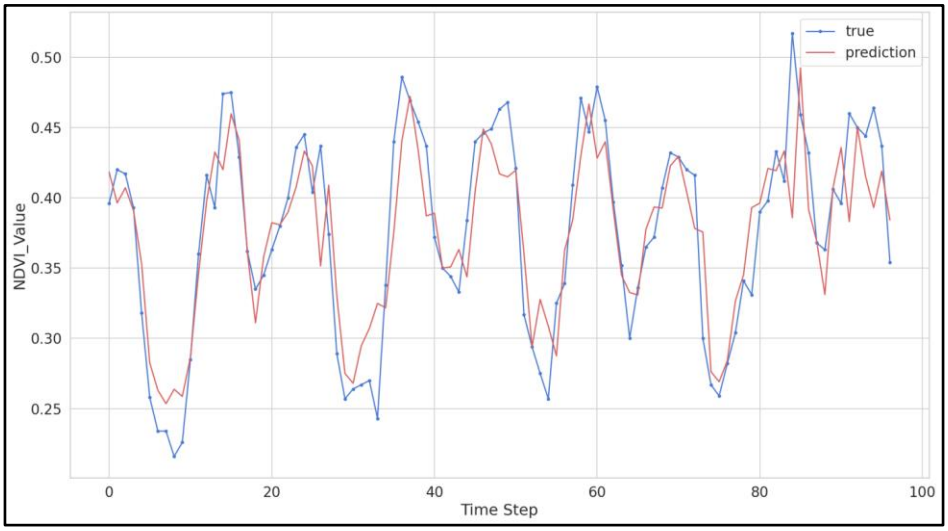


Fig. 4.13 True values Vs Predicted values graph for model-1

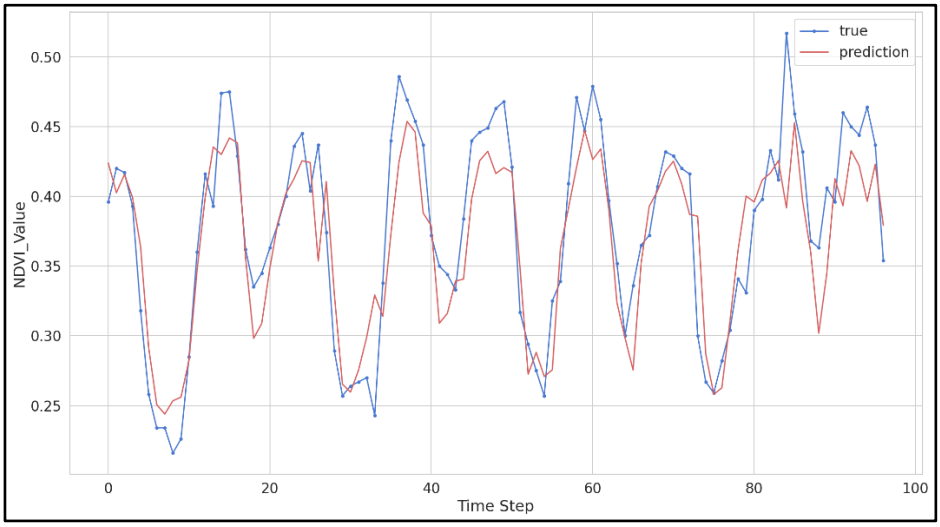


Fig. 4.14 True values Vs Predicted values graph for model-2

If we closely observe the above figures, we can see that model-2 misses some of the peak values or the lowest points. The prediction looks accurate for rest of the time except the peaks and lowest points. For model-1 these misses are little lesser than model-2.

4.6 Mean and Variance

Mean-1 = 0.326539,

Mean-2 = 0.350537

Variance-1= 0.006320,

Variance-2 = 0.005907

Here mean-1 is the mean for the first half of the NDVI data and mean-2 for the second half. Similarly, variance-1 and variance-2 are the variances for first half and second half respectively.

Mean-2 is greater than mean-1 which means that in the second half of the data NDVI values have increased. In case of variance it is the opposite case. Variance-2 is smaller than variance-1 which indicates that data is more clustered around the mean in the second half than first half of data.

4.7 MAE

The observed MAE for the model-1 is 0.026239.

The observed MAE for the model-2 is 0.02855.

Model-1 has better accuracy than model-2. To justify these values, we will consider more parameters like RMSE and R^2.

4.8 R^2(r-square)

The r^2 calculated between the true values and predicted values for model-1 is 0.779662 while for model-2 it is 0.74631. The r^2 value close to one indicates very good model. This means that the predicted values do not vary much from the original value. MAE close to zero and R-Square approaching 1 are indicative of high accuracy between observed and predicted values. The value of R^2 also indicates that model-1 has better accuracy than model-2.

These results evidently show that the model-1 LSTM network is more effective in predicting the future NDVI values by learning all the seasonal and annual changes in vegetation, changes due to deforestation or natural calamities, etc. with an appreciable accuracy. Thus, it is recommended to use LSTM network for reliable outcomes.

Table 4.7 Calculated Results

S. No.	Parameters	Results
1.	Total % change from 2001 to 2021	18.6298%
2.	mean-1	0.326539
3.	mean-2	0.350537
4.	variance-1	0.006320
5.	variance-2	0.005907

Table 4.8 Error metrics for models

Parameters	Model-1	Model-2
MAE	0.026239	0.02855
RMSE	0.034043	0.03653
R^2	0.779662	0.74631

Chapter 5 SUMMARY AND CONCLUSIONS

5.1 Summary

Using a temporal MODIS Terra dataset, the current study examined how urbanization, different climate conditions, long-term changes in vegetation cover, and their spatial relationships with NDVI changed over the state of Delhi between 2000 and 2022. The quantitative values found during the assessment demonstrate how specific land use practises have an impact on the loss of vegetation cover, its density, and the development of densely populated areas as well as other open and barren land.

This research has observed that Delhi has seen a slight decrease in the vegetation from year 2001 to 2010. But after that from 2010 onwards there has been significant change and we have observed increase in healthy vegetation. The analysis has observed a positive change of 18.63% from year 2001 to 2021. But the visual analysis shows that still there are some regions that haven't seen any change and we need to take necessary measures.

The current research has also developed two different machine learning models to learn the changes and forecast the upcoming values. The outcomes indicate that the LSTM model developed for the change prediction is capable enough to learn and forecast with acceptable accurate results. The mean absolute error of the predicted values with respect to test values for model-1 is 0.026239 and the R^2 for the model-1 on our dataset is 0.779662. Similarly, for model-2 RMSE is 0.03653 and R^2 is 0.74631. So, we can conclude from these results that LSTM model-1 is better than model-2. There is not much of difference in both model's accuracy so we can say that for vegetation change prediction using NDVI time series LSTM model is good choice.

The current study endorsed the conclusion that when urbanisation is being done, the vegetation cover and natural environment must be taken into account. To prevent the further deterioration of urban health and greenery, it is also suggested that the area needs further evaluation by taking high resolution data sets into account to estimate the hot spot points in the city.

5.2 Future scope

Since the vegetation change analysis has been carried out using the mean yearly NDVI of Delhi. The visual change analysis has been done between the years 2001 and 2021. For more precise analysis this time range can be reduced. The change analysis could be done for between two consecutive years or for every five years. This clearly show us that how the changes are taking place.

The LSTM model can accurately predict the time series up to 50-time intervals in the future, which will help with efforts to take care for vegetation well in advance to prevent any disturbance that could result in the loss of vegetation. The proposed model predicted the future values with R^2 of 0.7796, which a good value but this can be increased future with certain changes in the model or by using dataset with more standardization. There is always room for improvement. The study is an experiment to examine the application of LSTM for predicting vegetation dynamics. By experimenting with LSTM, interpretation can be made even better. Furthermore, incorporating climatic variables may result in a highly accurate prediction of vegetation dynamics.

Literature cited

Aburas, M. M., Abdullah, S. H., Ramli, M. F. and Ash'aari, Z. H. 2015. Measuring land cover change in seremban, malaysia using NDVI index. *Procedia Environ. Sci.*, 30(1): 238–243.

Anderson, J. R. 1971. Land-use classification schemes. *Photogramm. Eng. Remote Sens.*, 37(4): 379–387.

Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E. 1976. Land use and land cover classification system for use with remote sensor data. *In U S Geol Surv.*, 64(5): 154-168.

A**nyamba, A. and Tucker, C. J. 2005.** Analysis of Sahelian vegetation dynamics using NOAA-AVHRR NDVI data from 1981-2003. *J. Arid Environ.* 63(3): 596–614.

Bhandari, A. K., Kumar, A. and Singh, G. K. 2012. Feature extraction using normalized difference vegetation index (ndvi): a case study of Jabalpur city. *Procedia Tech.*, 6(2): 612–621.

Carlson, T. N. and Ripley, D. A. 1997. On the relation between NDVI, fractional vegetation cover and leaf area index. *Remote Sens. Environ.*, 62(3): 241–252.

de Jong, R., de Bruin, S., de Wit, A., Schaepman, M. E. and Dent, D. L. 2011. Analysis of monotonic greening and browning trends from global NDVI time-series. *Remote Sens. Environ.*, 115(2): 692–702.

Dutra, A. C., Shimabukuro, Y. E. and Escada, M. I. S. 2018. Data mining using ndvi time series applied to change detection. *J. Arid Environ.*, 58(6): 356-378.

Fensholt, R. and Proud, S. R. 2012. Evaluation of earth observation based global long term vegetation trends - comparing GIMMS and MODIS global NDVI time series. *Remote Sens. Environ.*, 119(5): 131–147.

Gandhi, G. M., Parthiban, S., Thummalu, N. and Christy, A. 2015. Ndvi: vegetation change detection using remote sensing and gis - a case study of Vellore district. *Procedia Comput. Sci.*, 57(2): 1199–1210.

Harris, R. I. 1992. Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test. *Economics letters*, 38(4), 381-386.

Ichii, K., Kawabata, A. and Yamaguchi, Y. 2002. Global correlation analysis for NDVI and climatic variables and NDVI trends: 1982-1990. *Int. J. Remote Sens.*, *23*(18): 3873–3878.

Jiang, L., Liu, Y., Wu, S. and Yang, C. 2021. Analyzing ecological environment change and associated driving factors in China based on NDVI time series data. *Ecol. Indic.*, 129(3): 354-367.

Jiang, Z., Huete, A. R., Chen, J., Chen, Y., Li, J., Yan, G. and Zhang, X. 2006. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sens. Environ.*, 101(3): 366–378.

Li, Y., Zhu, Z., Kong, D., Han, H. and Zhao, Y. 2019. EA-LSTM: Evolutionary attention-based LSTM for time series prediction. *Knowl Based Syst.*, 181(1): 104-115.

Martínez, B. and Gilabert, M. A. 2009. Vegetation dynamics from NDVI time series analysis using the wavelet transform. *Remote Sens. Environ.*, 113(9): 1823–1842.

Nyamekye, C., Osei Jnr, E. and Oseitutu, A. 2014. Classification of time series NDVI for the assessment of land cover change in Ghana using NOAA/AVHRR data. *Int. j. geomat. geosci.*, 8(2): 34–39.

Prafull, S., Anindita, S. C., Pradipika, V., Kumar, S. V. and Raj, M. S. 2022. Earth observation data sets in monitoring of urbanization and urban heat earth observation data sets in monitoring of urbanization and urban heat island of Delhi, India. *Geomatics, Nat. Hazards Risk.*, 13(1): 1762–1780.

Reddy, D. S. and Prasad, P. R. C. 2018. Prediction of vegetation dynamics using NDVI time series data and LSTM. *Model. Earth Syst. Environ.*, 4(1): 409–419.

Roerink, G. J., Menenti, M., Soepboer, W. and Su, Z. 2003. Assessment of climate impact on vegetation dynamics by using remote sensing. *Phys Chem Earth.*, *28*(3): 103–109.

Taufik, A., Syed Ahmad, S. S. and Azmi, E. F. 2019. Classification of landsat 8 satellite data using unsupervised methods. *Lect. Notes Netw.*, 67(3): 275–284.

Tiao, G. C. 2015. Time Series: ARIMA Methods. In *Int. Encyclopedia of the Social & Behav. Sci.*, 23(2): 114-132.

Wang, D., Morton, D., Masek, J., Wu, A., Nagol, J., Xiong, X., Levy, R., Vermote, E. and Wolfe, R. 2012. Impact of sensor degradation on the MODIS NDVI time series. *Remote Sens. Environ.*, 119(1): 55–61.

Yang, L., Wylie, B. K., Tieszen, L. L. and Reed, B. C. 1998. An analysis of relationships among climate forcing and time-integrated NDVI of grasslands over the U.S. northern and central Great Plains. *Remote Sens. Environ.*,65(1): 25–37.

Yu, Y., Si, X., Hu, C. and Zhang, J. 2019. A review of recurrent neural networks: Lstm cells and network architectures. *Neural Comput.*, 31(7): 1235–1270.

Zhang, X., Hu, Y., Zhuang, D., Qi, Y. and Ma, X. 2009. NDVI spatial pattern and its differentiation on the Mongolian Plateau. *J. Geogr. Sci.*, 19(4): 403–415.

Zhao, F., Yang, G., Yang, H., Zhu, Y., Meng, Y., Han, S. and Bu, X. 2021. Short and medium-term prediction of winter wheat ndvi based on the dtw–lstm combination method and modis time series data. *Remote Sens Environ.*, 13(2): 46-55.

Appendices

Dataset:

Date	NDVI
2000_02_18	0.421
2000_03_05	0.417
2000_03_21	0.344
2000_04_06	0.266
2000_04_22	0.199
2000_05_08	0.205
2000_05_24	0.183
2000_06_09	0.218
2000_06_25	0.25
2000_07_11	0.244
2000_07_27	0.37
2000_08_12	0.298
2000_08_28	0.427
2000_09_13	0.406
2000_09_29	0.348
2000_10_15	0.302
2000_10_31	0.254
2000_11_16	0.246
2000_12_02	0.253
2000_12_18	0.287
2001_01_01	0.343
2001_01_17	0.387
2001_02_02	0.407
2001_02_18	0.432
2001_03_06	0.411
2001_03_22	0.329
2001_04_07	0.261
2001_04_23	0.215
2001_05_09	0.192
2001_05_25	0.21
2001_06_10	0.303
2001_06_26	0.363
2001_07_12	0.339
2001_07_28	0.392
2001_08_13	0.423
2001_08_29	0.386
2001_09_14	0.359
2001_09_30	0.313
2001_10_16	0.274
2001_11_01	0.258

2001_11_17	0.252
2001_12_03	0.249
2001_12_19	0.288
2002_01_01	0.308
2002_01_17	0.404
2002_02_02	0.427
2002_02_18	0.457
2002_03_06	0.425
2002_03_22	0.291
2002_04_07	0.235
2002_04_23	0.185
2002_05_09	0.181
2002_05_25	0.219
2002_06_10	0.247
2002_06_26	0.257
2002_07_12	0.24
2002_07_28	0.254
2002_08_13	0.286
2002_08_29	0.415
2002_09_14	0.437
2002_09_30	0.372
2002_10_16	0.305
2002_11_01	0.282
2002_11_17	0.268
2002_12_03	0.275
2002_12_19	0.34
2003_01_01	0.156
2003_01_17	0.382
2003_02_02	0.452
2003_02_18	0.462
2003_03_06	0.451
2003_03_22	0.318
2003_04_07	0.26
2003_04_23	0.214
2003_05_09	0.209
2003_05_25	0.188
2003_06_10	0.179
2003_06_26	0.254
2003_07_12	0.316
2003_07_28	0.425
2003_08_13	0.409
2003_08_29	0.466
2003_09_14	0.44
2003_09_30	0.394
2003_10_16	0.348

2003_11_01	0.302
2003_11_17	0.287
2003_12_03	0.286
2003_12_19	0.385
2004_01_01	0.382
2004_01_17	0.408
2004_02_02	0.467
2004_02_18	0.451
2004_03_05	0.374
2004_03_21	0.273
2004_04_06	0.221
2004_04_22	0.243
2004_05_08	0.236
2004_05_24	0.242
2004_06_09	0.295
2004_06_25	0.305
2004_07_11	0.264
2004_07_27	0.341
2004_08_12	0.439
2004_08_28	0.427
2004_09_13	0.39
2004_09_29	0.375
2004_10_15	0.332
2004_10_31	0.32
2004_11_16	0.301
2004_12_02	0.341
2004_12_18	0.407
2005_01_01	0.41
2005_01_17	0.443
2005_02_02	0.475
2005_02_18	0.46
2005_03_06	0.439
2005_03_22	0.295
2005_04_07	0.237
2005_04_23	0.225
2005_05_09	0.216
2005_05_25	0.211
2005_06_10	0.206
2005_06_26	0.283
2005_07_12	0.343
2005_07_28	0.42
2005_08_13	0.394
2005_08_29	0.386
2005_09_14	0.444
2005_09_30	0.399

2005_10_16	0.346
2005_11_01	0.3
2005_11_17	0.293
2005_12_03	0.295
2005_12_19	0.358
2006_01_01	0.355
2006_01_17	0.374
2006_02_02	0.394
2006_02_18	0.397
2006_03_06	0.389
2006_03_22	0.357
2006_04_07	0.246
2006_04_23	0.211
2006_05_09	0.21
2006_05_25	0.236
2006_06_10	0.263
2006_06_26	0.313
2006_07_12	0.298
2006_07_28	0.446
2006_08_13	0.419
2006_08_29	0.443
2006_09_14	0.437
2006_09_30	0.371
2006_10_16	0.32
2006_11_01	0.247
2006_11_17	0.274
2006_12_03	0.285
2006_12_19	0.366
2007_01_01	0.354
2007_01_17	0.373
2007_02_02	0.4
2007_02_18	0.439
2007_03_06	0.434
2007_03_22	0.38
2007_04_07	0.268
2007_04_23	0.23
2007_05_09	0.231
2007_05_25	0.217
2007_06_10	0.272
2007_06_26	0.334
2007_07_12	0.397
2007_07_28	0.457
2007_08_13	0.406
2007_08_29	0.414
2007_09_14	0.414

2007_09_30	0.358
2007_10_16	0.304
2007_11_01	0.228
2007_11_17	0.252
2007_12_03	0.273
2007_12_19	0.294
2008_01_01	0.334
2008_01_17	0.349
2008_02_02	0.363
2008_02_18	0.37
2008_03_05	0.335
2008_03_21	0.291
2008_04_06	0.222
2008_04_22	0.211
2008_05_08	0.215
2008_05_24	0.261
2008_06_09	0.323
2008_06_25	0.392
2008_07_11	0.378
2008_07_27	0.424
2008_08_12	0.45
2008_08_28	0.433
2008_09_13	0.417
2008_09_29	0.382
2008_10_15	0.314
2008_10_31	0.288
2008_11_16	0.3
2008_12_02	0.3
2008_12_18	0.362
2009_01_01	0.368
2009_01_17	0.421
2009_02_02	0.415
2009_02_18	0.406
2009_03_06	0.341
2009_03_22	0.289
2009_04_07	0.236
2009_04_23	0.212
2009_05_09	0.211
2009_05_25	0.215
2009_06_10	0.217
2009_06_26	0.224
2009_07_12	0.318
2009_07_28	0.368
2009_08_13	0.424
2009_08_29	0.392

2009_09_14	0.428
2009_09_30	0.376
2009_10_16	0.324
2009_11_01	0.258
2009_11_17	0.285
2009_12_03	0.295
2009_12_19	0.349
2010_01_01	0.308
2010_01_17	0.352
2010_02_02	0.384
2010_02_18	0.409
2010_03_06	0.363
2010_03_22	0.244
2010_04_07	0.212
2010_04_23	0.211
2010_05_09	0.202
2010_05_25	0.194
2010_06_10	0.21
2010_06_26	0.22
2010_07_12	0.291
2010_07_28	0.415
2010_08_13	0.392
2010_08_29	0.354
2010_09_14	0.444
2010_09_30	0.386
2010_10_16	0.329
2010_11_01	0.278
2010_11_17	0.343
2010_12_03	0.324
2010_12_19	0.302
2011_01_01	0.323
2011_01_17	0.366
2011_02_02	0.433
2011_02_18	0.457
2011_03_06	0.404
2011_03_22	0.306
2011_04_07	0.261
2011_04_23	0.232
2011_05_09	0.237
2011_05_25	0.236
2011_06_10	0.249
2011_06_26	0.317
2011_07_12	0.347
2011_07_28	0.375
2011_08_13	0.415

2011_08_29	0.456
2011_09_14	0.433
2011_09_30	0.392
2011_10_16	0.356
2011_11_01	0.293
2011_11_17	0.249
2011_12_03	0.307
2011_12_19	0.303
2012_01_01	0.353
2012_01_17	0.379
2012_02_02	0.373
2012_02_18	0.374
2012_03_05	0.366
2012_03_21	0.303
2012_04_06	0.249
2012_04_22	0.226
2012_05_08	0.204
2012_05_24	0.198
2012_06_09	0.197
2012_06_25	0.195
2012_07_11	0.287
2012_07_27	0.296
2012_08_12	0.353
2012_08_28	0.505
2012_09_13	0.446
2012_09_29	0.403
2012_10_15	0.348
2012_10_31	0.234
2012_11_16	0.305
2012_12_02	0.333
2012_12_18	0.352
2013_01_01	0.351
2013_01_17	0.399
2013_02_02	0.434
2013_02_18	0.452
2013_03_06	0.421
2013_03_22	0.366
2013_04_07	0.267
2013_04_23	0.24
2013_05_09	0.21
2013_05_25	0.213
2013_06_10	0.273
2013_06_26	0.251
2013_07_12	0.347
2013_07_28	0.367

2013_08_13	0.475
2013_08_29	0.434
2013_09_14	0.411
2013_09_30	0.419
2013_10_16	0.378
2013_11_01	0.319
2013_11_17	0.309
2013_12_03	0.337
2013_12_19	0.382
2014_01_01	0.365
2014_01_17	0.406
2014_02_02	0.436
2014_02_18	0.431
2014_03_06	0.419
2014_03_22	0.321
2014_04_07	0.276
2014_04_23	0.25
2014_05_09	0.241
2014_05_25	0.229
2014_06_10	0.22
2014_06_26	0.265
2014_07_12	0.299
2014_07_28	0.42
2014_08_13	0.384
2014_08_29	0.433
2014_09_14	0.397
2014_09_30	0.345
2014_10_16	0.303
2014_11_01	0.263
2014_11_17	0.258
2014_12_03	0.272
2014_12_19	0.362
2015_01_01	0.361
2015_01_17	0.407
2015_02_02	0.41
2015_02_18	0.386
2015_03_06	0.419
2015_03_22	0.374
2015_04_07	0.314
2015_04_23	0.262
2015_05_09	0.249
2015_05_25	0.235
2015_06_10	0.252
2015_06_26	0.298
2015_07_12	0.375

2015_07_28	0.34
2015_08_13	0.467
2015_08_29	0.425
2015_09_14	0.423
2015_09_30	0.376
2015_10_16	0.319
2015_11_01	0.276
2015_11_17	0.29
2015_12_03	0.302
2015_12_19	0.338
2016_01_01	0.342
2016_01_17	0.313
2016_02_02	0.38
2016_02_18	0.348
2016_03_05	0.375
2016_03_21	0.299
2016_04_06	0.228
2016_04_22	0.221
2016_05_08	0.228
2016_05_24	0.225
2016_06_09	0.21
2016_06_25	0.281
2016_07_11	0.359
2016_07_27	0.435
2016_08_12	0.457
2016_08_28	0.457
2016_09_13	0.434
2016_09_29	0.369
2016_10_15	0.335
2016_10_31	0.261
2016_11_16	0.295
2016_12_02	0.303
2016_12_18	0.329
2017_01_01	0.401
2017_01_17	0.411
2017_02_02	0.444
2017_02_18	0.421
2017_03_06	0.391
2017_03_22	0.299
2017_04_07	0.237
2017_04_23	0.232
2017_05_09	0.231
2017_05_25	0.244
2017_06_10	0.259
2017_06_26	0.323

2017_07_12	0.374
2017_07_28	0.361
2017_08_13	0.462
2017_08_29	0.469
2017_09_14	0.442
2017_09_30	0.405
2017_10_16	0.333
2017_11_01	0.255
2017_11_17	0.31
2017_12_03	0.365
2017_12_19	0.376
2018_01_01	0.393
2018_01_17	0.396
2018_02_02	0.42
2018_02_18	0.417
2018_03_06	0.393
2018_03_22	0.318
2018_04_07	0.258
2018_04_23	0.234
2018_05_09	0.234
2018_05_25	0.216
2018_06_10	0.226
2018_06_26	0.285
2018_07_12	0.36
2018_07_28	0.416
2018_08_13	0.393
2018_08_29	0.474
2018_09_14	0.475
2018_09_30	0.429
2018_10_16	0.362
2018_11_01	0.335
2018_11_17	0.345
2018_12_03	0.363
2018_12_19	0.38
2019_01_01	0.4
2019_01_17	0.436
2019_02_02	0.445
2019_02_18	0.404
2019_03_06	0.437
2019_03_22	0.374
2019_04_07	0.289
2019_04_23	0.257
2019_05_09	0.264
2019_05_25	0.267
2019_06_10	0.27

2019_06_26	0.243
2019_07_12	0.338
2019_07_28	0.44
2019_08_13	0.486
2019_08_29	0.469
2019_09_14	0.454
2019_09_30	0.437
2019_10_16	0.372
2019_11_01	0.35
2019_11_17	0.344
2019_12_03	0.333
2019_12_19	0.384
2020_01_01	0.44
2020_01_17	0.446
2020_02_02	0.449
2020_02_18	0.463
2020_03_05	0.468
2020_03_21	0.421
2020_04_06	0.317
2020_04_22	0.294
2020_05_08	0.275
2020_05_24	0.257
2020_06_09	0.325
2020_06_25	0.339
2020_07_11	0.409
2020_07_27	0.471
2020_08_12	0.447
2020_08_28	0.479
2020_09_13	0.455
2020_09_29	0.397
2020_10_15	0.352
2020_10_31	0.3
2020_11_16	0.336
2020_12_02	0.365
2020_12_18	0.372
2021_01_01	0.407
2021_01_17	0.432
2021_02_02	0.429
2021_02_18	0.42
2021_03_06	0.416
2021_03_22	0.3
2021_04_07	0.267
2021_04_23	0.259
2021_05_09	0.282
2021_05_25	0.304

2021_06_10	0.341
2021_06_26	0.331
2021_07_12	0.39
2021_07_28	0.398
2021_08_13	0.433
2021_08_29	0.412
2021_09_14	0.517
2021_09_30	0.459
2021_10_16	0.432
2021_11_01	0.368
2021_11_17	0.363
2021_12_03	0.406
2021_12_19	0.396
2022_01_01	0.46
2022_01_17	0.45
2022_02_02	0.444
2022_02_18	0.464
2022_03_06	0.437
2022_03_22	0.354

CURRICULUM VITAE

Name	**:**	Navtej Anand	**Phone Number**	**:**	+917251054701
Mailing Address	**:**	70 - Vani Vihar, Raipur Road, Dehradun, Uttarakhand, India (248001)	**Permanent Address**	**:**	Same as mailing address
E-mail	**:**	navtejnt1@gmail.com			
Career objectives	**:**	To secure a good position in a competitive and creative engineering environment that will benefit from my creative ideas and experience.			

Educational Qualification

Sr.No.	Examination Passed	Institution	Year	Percentage/CGPA
1	M.Tech (Information Technology)	G.B.P.U.A & T, Pantnagar	2022	8.052
2	B. Tech (Computer Engineering)	G.B.P.U.A & T, Pantnagar	2019	6.562
3	Intermediate	Touch Wood School, Dehradun	2014	69.40%
4	High school	Touch Wood School, Dehradun	2012	78.00%

Specialization: **Major:** Information Technology **Minor:** Nil

Thesis Title: "Spatio-temporal analysis of Vegetation Dynamics of New Delhi (India) using Satellite Data"

Publication: Nil

Conference/Seminars/Workshops/Training Attended: Nil

List of papers presented in conference/seminar during degree programme: Nil

Software Skills: Python, Java, Adobe Photoshop, MS Office.

Professional Skills: Communication and Motivational skill.

Professional Affiliations (Membership, etc.,): Nil

Awards/Honours/Achievements: Nil

Place: Pantnagar
Date: 01-10-2022

(Navtej Anand)

Name	:	Navtej Anand	**Id. No.**	:	48905
Sem. and year of admission	:	1st Sem., 2020-21	**Degree**	:	M. Tech (I.T.)
			Department	:	Information Technology
Major	:	Information Technology			
Thesis title	:	**"Spatio-temporal analysis of Vegetation Dynamics of New Delhi (India) using Satellite Data"**			
Page No.	:	1 - 57	**Advisor**	:	Dr. Subodh Prasad

ABSTRACT

Understanding and analysing vegetation cover changes is crucial for a number of reasons, especially when it comes to taking the necessary conservation measures. This study asses the vegetation changes in the New Delhi (India) over the years from 2000 to march 2022 based on NDVI (Normalized Difference Vegetation Index). The NDVI values have been collected from MODIS terra satellite imagery. Using this NDVI data the study finds that the vegetation greenness of Delhi has increased by 18.63% from year 2001 to 2021. A dataset of 509 NDVI values have been used for making the time series. An attempt has been carried out to predict the vegetation change using this MODIS NDVI time series data and LSTM (Long Short Term Memory) network. The prediction has been carried out on two different LSTM models side by side on the same data and comparative study has been done. The LSTM networks has been trained with 80% of the data and rest 20% are used for testing the model's accuracy. The results show that both the LSTM model are capable of predicting the future NDVI values with appreciable accuracy but model-1 predicts with better accuracy and lesser errors. Model-1 predicts the future NDVI values with RMSE less than 0.034 and R^2 of more than 0.77. Model-2 is not far behind, it predicts with RMSE of around 0.036 and R^2 of around 0.74. So, this study concludes that using LSTM networks it is possible to accurately predict vegetation changes well in advance and take appropriate proactive measures to protect and enhance the vegetation in any area.

(Dr. Subodh Prasad)
Advisor

(Navtej Anand)
Author

नाम	:	नवतेज आनंद	**परिचयांक**	:	४८९०५
सत्र एवं प्रवेश का वर्ष	:	प्रथम सेमेस्टर, २०२०-२१	**उपाधि**	:	एम.टेक
मुख्य विषय	:	सूचान प्रौद्योगिकी	**विभाग**	:	सूचान प्रौद्योगिकी
शोध का शीर्षक	:	"सैटेलाइट डेटा का उपयोग करके नई दिल्ली (भारत) के वनस्पति परिवर्तन का स्थानिक-अस्थायी विश्लेषण"			
पृष्ठ संख्या	:	१-५७	**सलाहकार**	:	डॉ सुबोध प्रसाद

सारांश

वनस्पति आवरण परिवर्तनों को समझना और उनका विश्लेषण करना कई कारणों से महत्वपूर्ण है, खासकर जब आवश्यक संरक्षण उपाय करने की बात आती है। यह अध्ययन एनडीवीआई (सामान्यीकृत अंतर वनस्पति सूचकांक) के आधार पर २००० से मार्च २०२२ तक नई दिल्ली (भारत) में वनस्पति परिवर्तन का आकलन करता है। एनडीवीआई मान मोडिस टेरा सैटेलाइट इमेजरी से एकत्र किए गए हैं। इस एनडीवीआई डेटा का उपयोग करते हुए अध्ययन में पाया गया है कि दिल्ली की वनस्पति हरियाली में वर्ष २००१ से २०२१ तक १८.६३% की वृद्धि हुई है। समय श्रृंखला बनाने के लिए ५०९ एनडीवीआई मानों के एक डेटासेट का उपयोग किया गया है। इस मोडिस एनडीवीआई समय श्रृंखला डेटा और एलएसटीएम (लॉन्ग शॉर्ट टर्म मेमोरी) नेटवर्क का उपयोग करके वनस्पति परिवर्तन की भविष्यवाणी करने का प्रयास किया गया है। भविष्यवाणी एक ही डेटा पर दो अलग-अलग एलएसटीएम मॉडल पर एक साथ की गई है और तुलनात्मक अध्ययन किया गया है। एलएसटीएम नेटवर्क को ८०% डेटा के साथ प्रशिक्षित किया गया है और शेष २०% का उपयोग मॉडल की सटीकता के परीक्षण के लिए किया जाता है। परिणाम बताते हैं कि दोनों एलएसटीएम मॉडल प्रशंसनीय सटीकता के साथ भविष्य के एनडीवीआई मूल्यों की भविष्यवाणी करने में सक्षम हैं लेकिन मॉडल -१ बेहतर सटीकता और कम त्रुटियों के साथ भविष्यवाणी करता है। मॉडल -१ भविष्य के एनडीवीआई मूल्यों की भविष्यवाणी करता है जिसमें आरएमएसई ०.०३४ से कम और आर^2 ०.७७ से अधिक है। मॉडल -२ भी पीछे नहीं है, यह लगभग ०.०३६ के आरएमएसई और ०.७४ के आर^2 के साथ भविष्यवाणी करता है। इसलिए इस अध्ययन का निष्कर्ष है कि एलएसटीएम नेटवर्क का उपयोग करके पहले से ही वनस्पति परिवर्तनों की सटीक भविष्यवाणी करना संभव है और किसी भी क्षेत्र में वनस्पति की रक्षा और वृद्धि के लिए उचित सक्रिय उपाय करना संभव है।

(डॉ सुबोध प्रसाद)
सलाहकार

(नवतेज आनंद)
लेखक