ENVIRONMENTAL ASSESSMENT USING SATELLITE IMAGES: A STATISTICAL ANALYSIS

Project report submitted in partial fulfillment of the requirement for

POST-GRADUATE DIPLOMA IN
STATISTICAL METHODS AND ANALYTICS

Submitted

by

Anik Saha
(Roll No: DST-23/24-022),
Sourav Chakraborty
(Roll No: DST-23/24-005)
And
Shalini Roy
(Roll No: DST-23/24-020)

June, 2024

INDIAN STATISTICAL INSTITUTE
[North East Centre, Tezpur, Punioni PO Solmara, Assam 784501]

Date: 11/06/2024

CERTIFICATE

This is to certify Mr. Anik Saha, Mr. Sourav Chakraborty and Ms. Shalini Roy has done the project under my supervision and guidance (from February 2024 to May 2024). This is an original project report based on work carried out by them in partial fulfillment of the requirement for the Post-Graduate Diploma in Statistical Methods and Analytics programme of the Indian Statistical Institute, North-East Centre, Tezpur, Assam.

Dr. Kushal Banik Chowdhury
(Assistant Professor)

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to Dr. Kushal Banik Chowdhury for serving as our guide throughout this project and providing us with valuable suggestions and guidance. We are truly grateful for his continuous support.

We would also like to extend our thanks to Dr. Sanjit Maitra and Mr. Rituraj Gogoi, for sharing their valuable knowledge and expertise with us during the course of our project. Their insights and feedback have been invaluable.

Finally, we would like to acknowledge the authorities of ISI-NE for their unwavering support and guidance throughout the project. Their assistance has been instrumental in shaping our research.

Once again, we express our sincere appreciation to all those who have contributed to the completion of this project.

Anik Saha

Sourav Chakraborty

Shalini Roy

CONTENTS

Introduction :

Motivation :

The demand for new housing, schools, hospitals, transportation, park and many other basic needs of people are increasing day by day in developing countries like INDIA. Increase in the urban hubs consumes cultivated land and thus reduces the agricultural output. As such the Govt. and the policymakers should be interested in keeping an eye on the fastly growing regions of India such as Mumbai, Kolkata, etc. so that they can take decisions on various policies and reforms such that in the regions under consideration urban hubs are increased as well as the vegetation is also increased.

So, we would like to analyze the NDVI and the NDBI values over time and how they are dependent on each other over time.

1. Landsat-8 :

Landsat 8 is an American Earth observation satellite launched on 11 February 2013. It is the eighth satellite in the Landsat program; the seventh to reach orbit successfully. Originally called the Landsat Data Continuity Mission (LDCM), it is a collaboration between NASA and the United States Geological Survey (USGS).

With Landsat 5 retiring in early 2013, leaving Landsat 7 as the only on-orbit Landsat program satellite, Landsat 8 ensures the continued acquisition and availability of Landsat data utilizing a two-sensor payload, the Operational Land Imager (OLI) and the Thermal InfraRed Sensor (TIRS). Respectively, these two instruments collect image data for nine shortwave bands and two longwave thermal bands. The satellite was developed with a 5 years mission design life but was launched with enough fuel on board to provide for upwards of ten years of operations.

Providing moderate-resolution imagery, from 15 metres to 100 metres, of Earth's land surface and polar regions, Landsat 8 operates in the visible, near-infrared, short wave infrared, and thermal infrared spectrums. Landsat 8 captures more than 700 scenes a day, an increase from the 250 scenes a day on Landsat 7.

Landsat 8's Operational Land Imager (OLI) improves on past Landsat sensors and was built, under contract to NASA, by Ball Aerospace & Technologies. The nine bands of OLI are as follows:

Spectral Band	Description	Wavelength	Resolution
Band 1	Coastal Aerosol	0.43 - 0.45 µm	30 m
Band 2	Blue	0.450 - 0.51 µm	30 m
Band 3	Green	0.53 - 0.59 µm	30 m

Band 4	Red	0.64 - 0.67 µm	30 m
Band 5	Near-Infrared	0.85 - 0.88 µm	30 m
Band 6	SWIR 1	1.57 - 1.65 µm	30 m
Band 7	SWIR 2	2.11 - 2.29 µm	30 m
Band 8	Panchromatic (PAN)	0.50 - 0.68 µm	15 m
Band 9	Cirrus	1.36 - 1.38 µm	30 m

2. Vector image :

A vector image is an image created using mathematical formulas to represent the image, rather than using a grid of pixels.Vector images are made up of lines and shapes that are mathematically defined.This means that they can be scaled up or down without losing any quality. They are often used for logos and illustrations.

3. Raster image :

Raster images, on the other hand, are made up of pixels, which are small squares of colour. Raster images can't be scaled up without losing quality, but they can be scaled down without any problems. They are often used for photos and other complex images.

4. Virtual Raster:

A virtual raster (VRT) is a GDAL method of combining multiple raster tiles into one file. It enables visualisation of whole raster datasets and faster navigation around your dataset in GIS software.

5. Shape file :

A shapefile is a vector data file format commonly used for geospatial analysis. Shapefiles store the location, geometry, and attribution of point, line, and polygon features.

6. NDVI :

To track changes in vegetation across regions, Normalized Difference Vegetation Index (NDVI) is widely used with satellite data. High NDVI values indicate healthy vegetation, while low values represent other features. Factors like precipitation and methods affect vegetation patterns. However, in urban areas, high population density, industrial growth, and climate change significantly impact the green spaces on Earth's surface. The equation of the NDVI or Normalized Diference Vegetation Index is:

$$NDVI = (NIR - R)/(NIR + R)$$

The NDVI scale ranges from -1 to +1. Healthy vegetation on Earth's surface typically has values between 0 and +1. NIR refers to the near-infrared satellite band used in satellite imagery, and R stands for the Red band found in Landsat satellite images.

7. <u>NDBI</u> :

As urbanization intensifies, the increase in population and infrastructure development has led to a increase of built-up areas in urban environments. These areas are characterized by high population densities and can be monitored using the Normalized Difference Built-up Index (NDBI). Satellite imagery from Landsat TM and OLI/TIRS datasets is utilized to assess built-up conditions. The NDBI has proven effective in mapping built-up areas in both urban and rural settings. Values of the NDBI range from -1 to +1, with values between 0 and +1 indicating the presence of dense built-up structures on the Earth's surface. The NDBI value is calculated by the following equation:

$$NDBI = (SWIR - NIR)/(SWIR + NIR)$$

Two Landsat image bands, SWIR (shortwave infrared) and NIR (near-infrared), are used to track NDBI (Normalized Difference Built-Up Index) values.

8. <u>QGiS</u> :

QGIS is a geographic information system (GIS) software that is free and open-source.[2] QGIS supports Windows, macOS, and Linux.[3] It supports viewing, editing, printing, and analysis of geospatial data in a range of data formats. QGIS was previously also known as Quantum GIS.

QGIS functions as geographic information system (GIS) software, allowing users to analyze and edit spatial information, in addition to composing and exporting graphical maps.[2] QGIS supports raster, vector, mesh, and point cloud layers.[4] Vector data is stored as either point, line, or polygon features. Multiple formats of raster images are supported, and the software can georeference images.

9. <u>Study Area</u> : Mumbai and neighbouring regions

Our study area(**Latitude:** 19.0760° N , **Longitude**:72.8777° E) is located on the west coast of India along the Arabian sea. It is one of the most populous cities in world, with a population exceeding 20 million. Percentage of vegetation and build up areas of study area are approximately 27% and 50% respectively.

10. Literature Review :

There are extensive literatures available on a similar study. They are as follows:

1. Hai Tao, Bassim Mohammed Hashim,Salim Heddam, Leonardo Goliatt, Mou Leong Tan, Zulfaqar Sa'adi, Iman Ahmadi Nafar, Mayadah W. Falah, Bijay Halder, Zaher Mundher Yaseen. Megacities' environmental assessment for Iraq region using satellite image and geo- spatial tools. This research studies the remote sensing-based on Landsat images that are used for investigating the vegetation circumstances of megacities of Iraq.

2. Shoukat Ali Shaha, Madeeha Kiranb, Aleena Nazirc, Shaharyar Hassan Ashrafani. EXPLORING NDVI AND NDBI RELATIONSHIP USING LANDSAT 8 OLI/TIRS IN KHANGARH TALUKA, GHOTKI. This study investigates the relationship between Normalized Difference Vegetation Index (NDVI) and Normalized Difference Built-up Index (NDBI) in Khangarh taluka.

11. Objective :

The present work of us proposed entitled “Environmental assessment using satellite images: A Statistical Analysis” to study the NDVI and NDBI of Mumbai and it’s surrounding regions during January 2014 to January 2024 considering only the first quarter. The specific objectives of the present study are:

1. To find how the NDVI and NDBI changes over time for each tile.
2. To find how NDVI and NDBI are correlated.

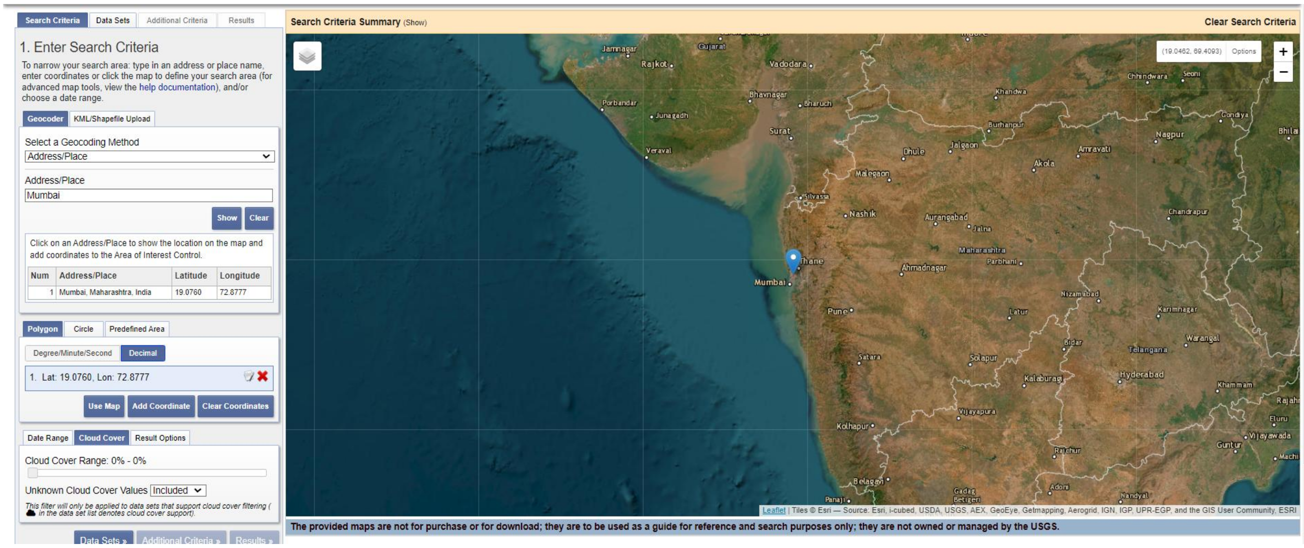
METHODOLOGY

Data Collection :

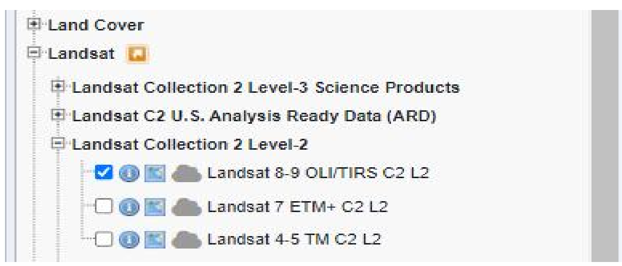
The Landsat TM and OLI/TIRS data were used for this study. The satellite datasets are acquired from the USGS web portal (https://earthexplorer.usgs.gov/).

Range: 2014-2024

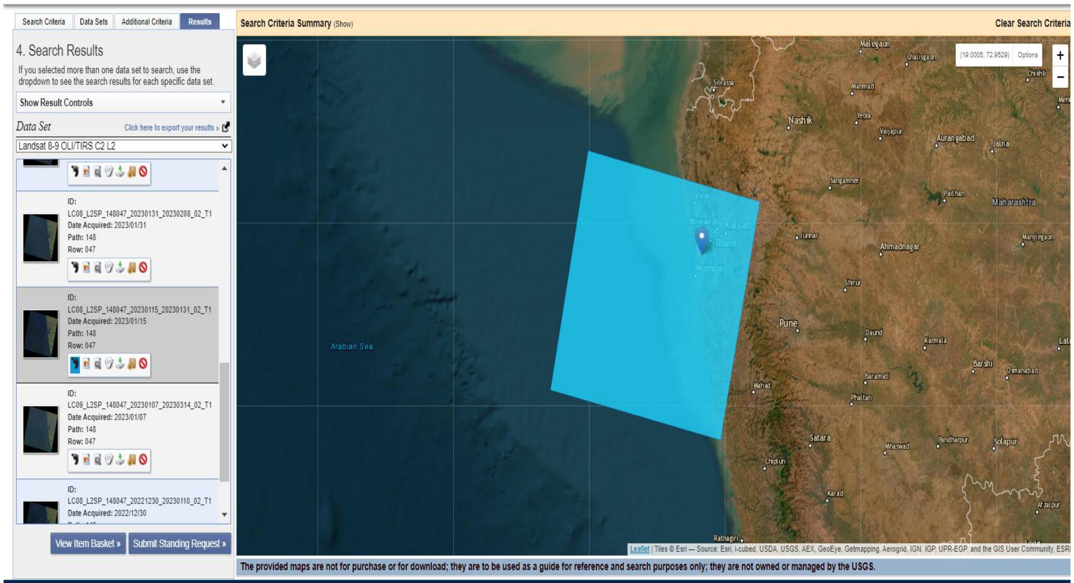
We collect the satellite images of the region under study in the following ways:



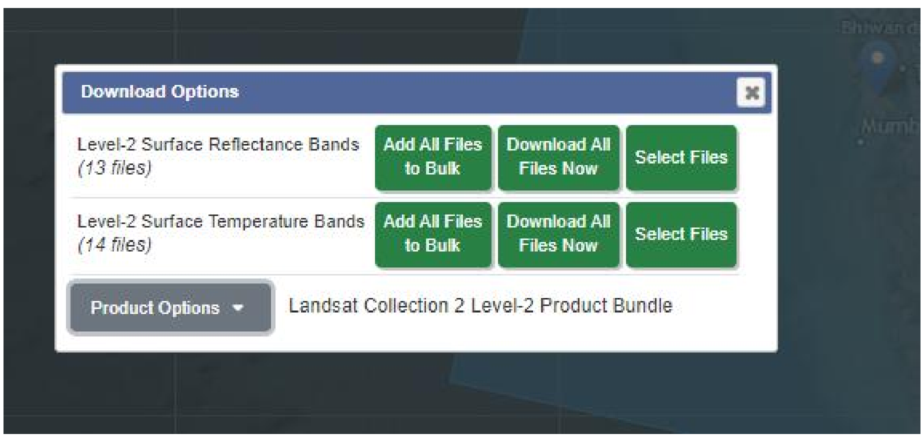
Step 1: After visiting the website and logging in, select the geocoding method as Address/Place and set the Address/Place as Mumbai. Setting cloud cover to zero, tap on datasets.



Step 2: Choose Landsat Collection 2 Level 2 and then select landsat 8-9 OLI/TIRS C2 L2. Tap on results.



Step 3: We get the following on clicking on the footprint icon. In the image, the satellite image for 15th Jan 23 is selected.



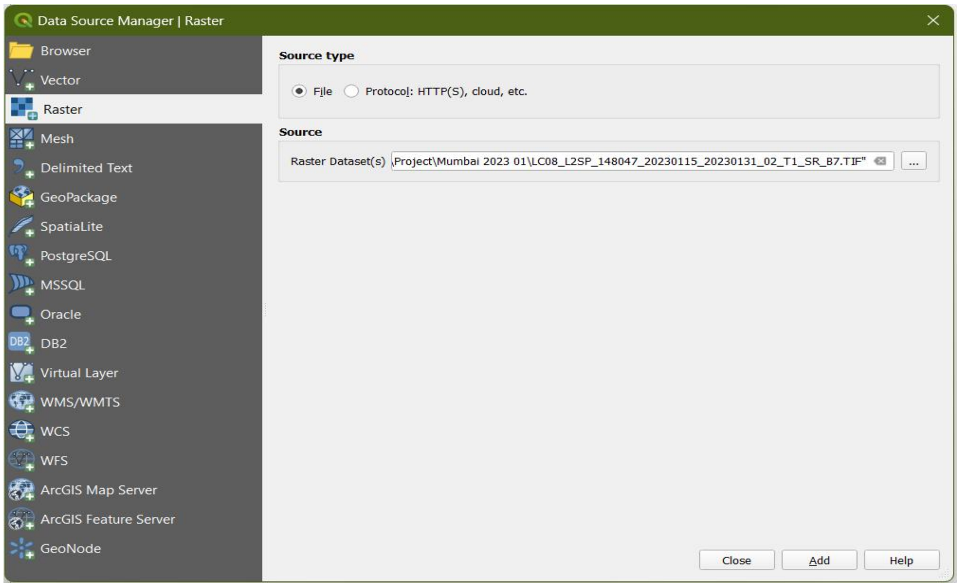
Step 4: On clicking the save icon, the following dialog box appears. We opt for the Surface Reflectance Bands and download it.

Thus, 7 bands of the satellite images are collected for first quarter of each year from 2014 to 2024 using the above steps. The first quarter is chosen to remove seasonality from the dataset.

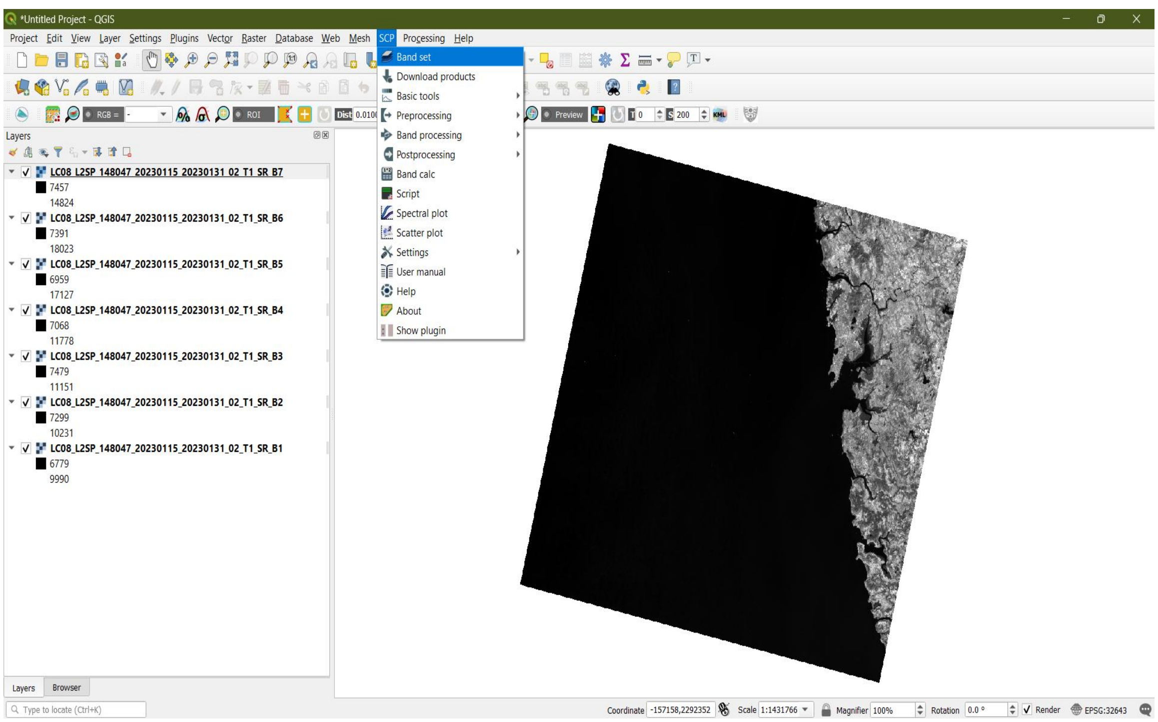
Data Preprocessing :

QGiS : Install the QGIS software using the link: [Download QGIS]

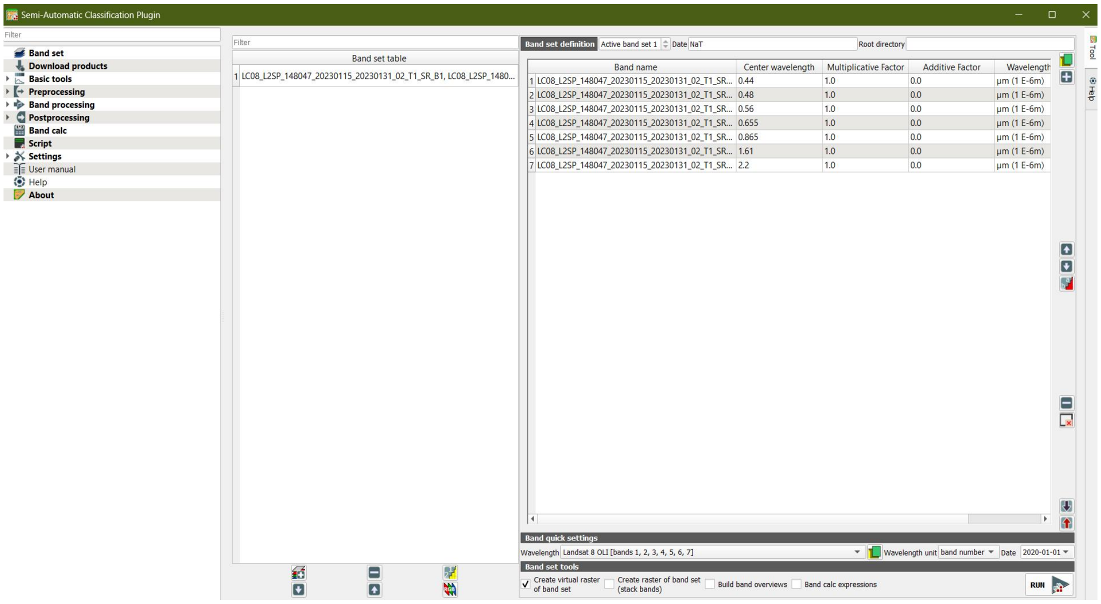
Step 1: Tap on data source manager icon, the dialogue box appears. Click on Raster and upload the bands. Then, tap on add.



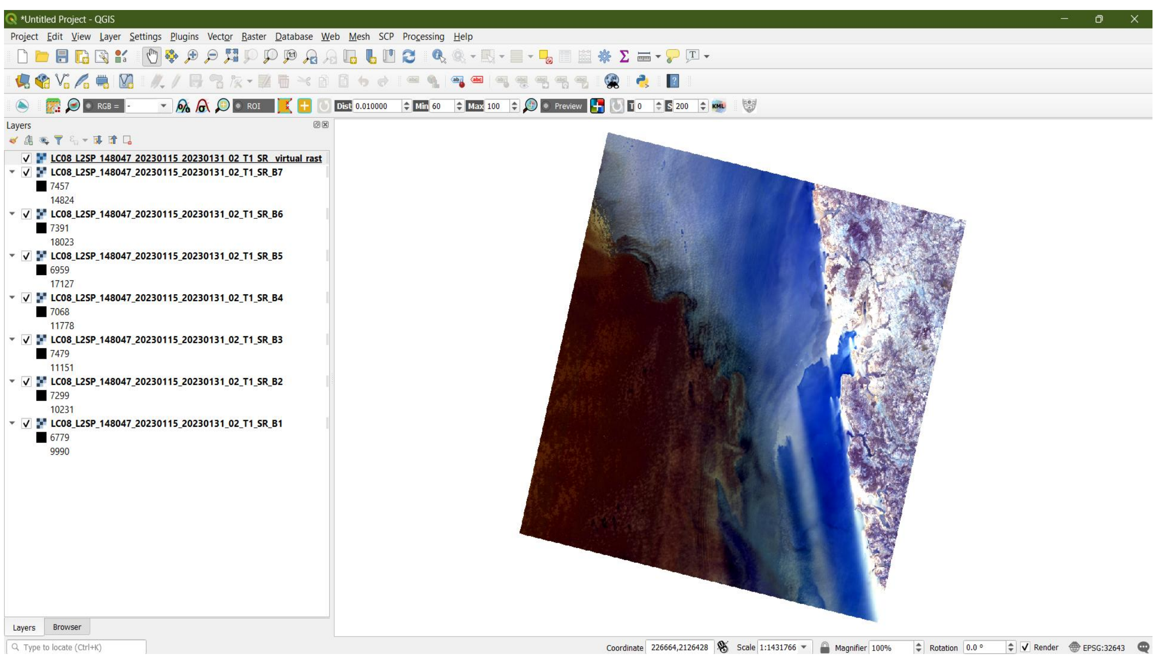
Step 2 : Click on SCP and then Band set (if there is no SCP option download it from the plugins).



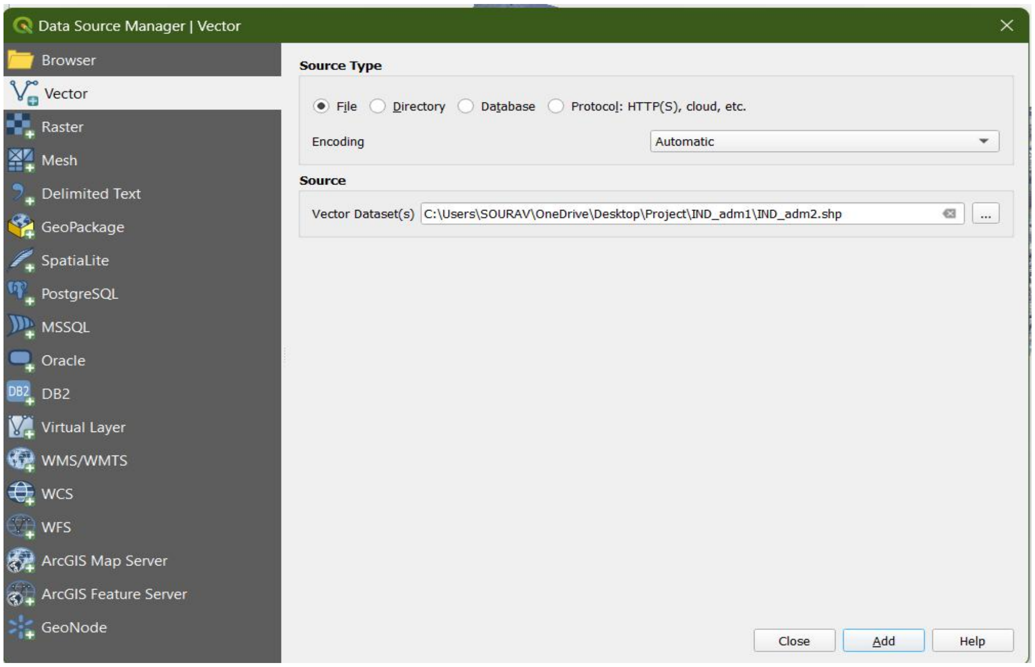
Step 3 : The Semi Automatic Classification plugin dialogue box appears. Select the bands and then choose wavelength and wavelength unit and select Create virtual raster of band set.



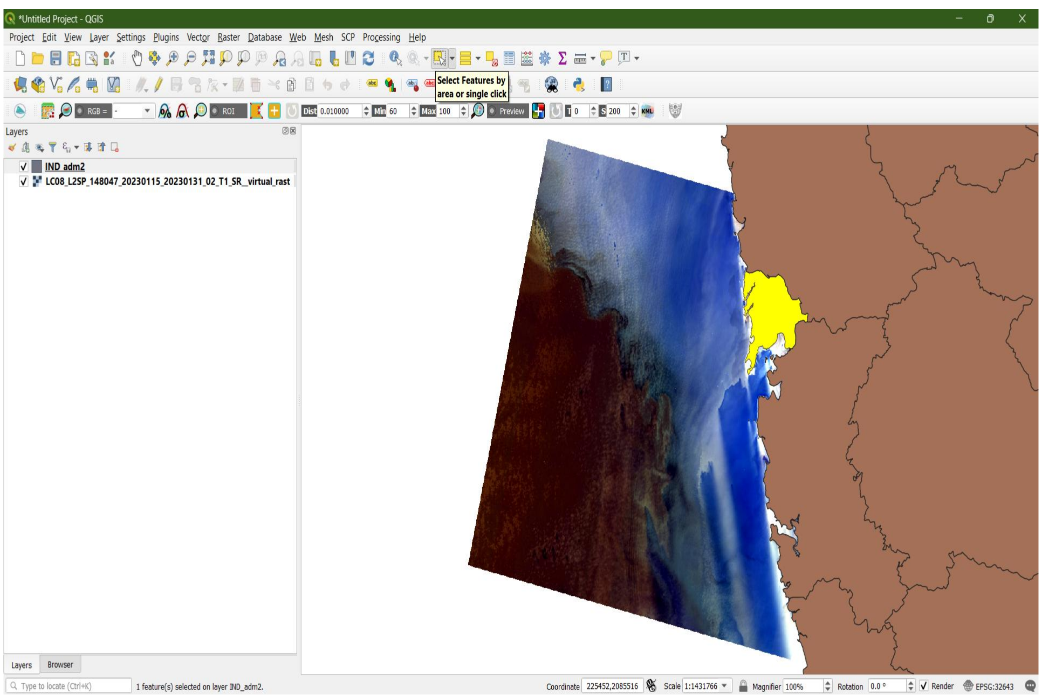
Step 4 : The virtual raster is created. Open the data source manager and choose the raster file.



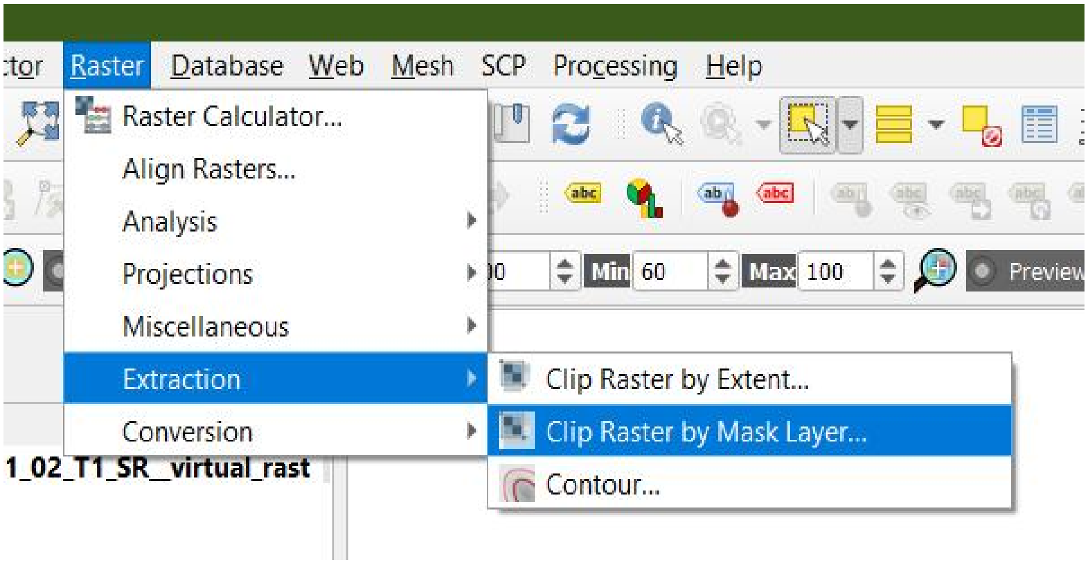
Step 5 : In the Data Source manager, go to Vector and choose the shape file to clip out the desired region.



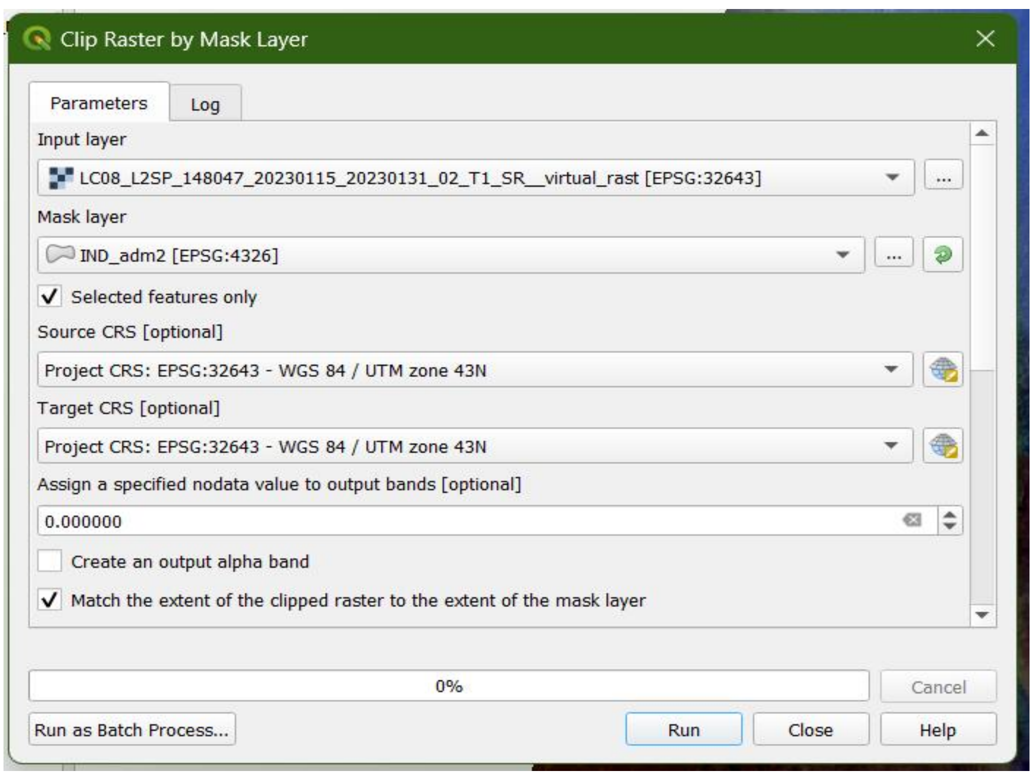
Step 6 : Tap on the select features icon and tap on the region required.



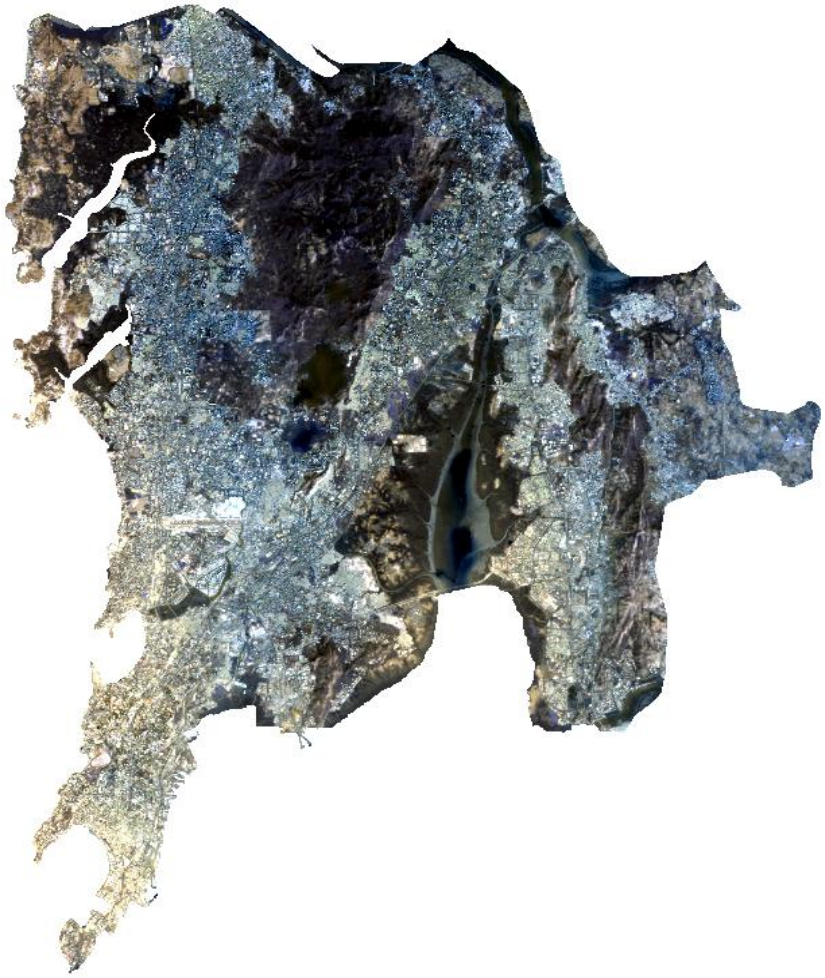
Step 7 : Select Raster→ Extraction→ Clip Raster by Mask Layer.



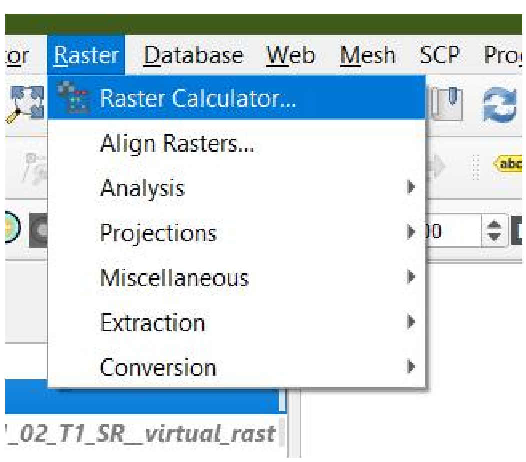
Step 8 : Fill the credentials and tap run.



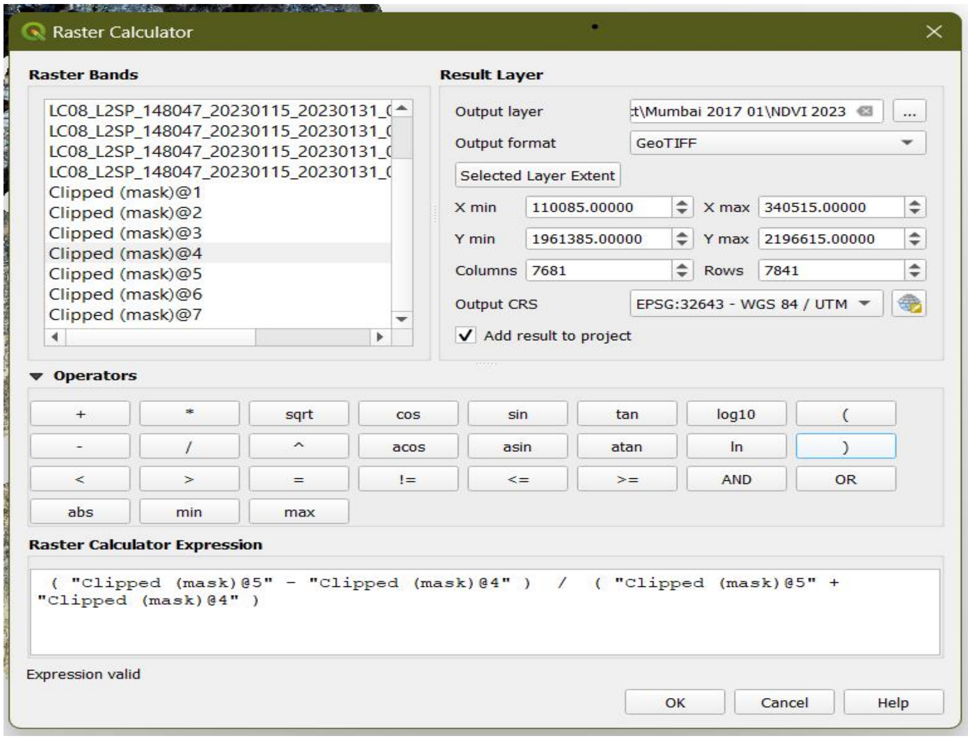
Step 9 : We get the clipped image which is our study area.



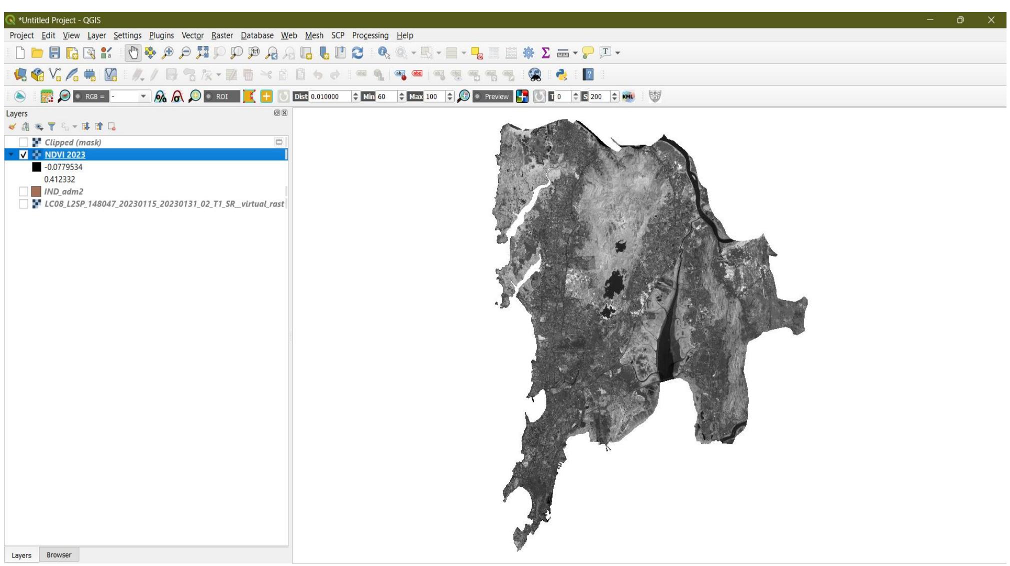
Step 10 : Select raster and then choose Raster Calculator to get the NDVI and NDBI images.



Step 11 : We write the NDVI and NDBI formula on the clipped image and select the folder where it is to be saved [Output Layer].



Step 12 : We obtain the NDVI or NDBI image as given.



NDVI IMAGE

R data extraction :

After creating the NDVI and NDBI image from QGIS, we convert it into a matrix where each element defines a NDVI or NDBI index of each pixel.

Tiling (3*3) :

The NDVI and NDBI images were further divided into 3×3 tiles and mean value of each is calculated for further analysis using R programming.

Data description :

The dataset is as follows:

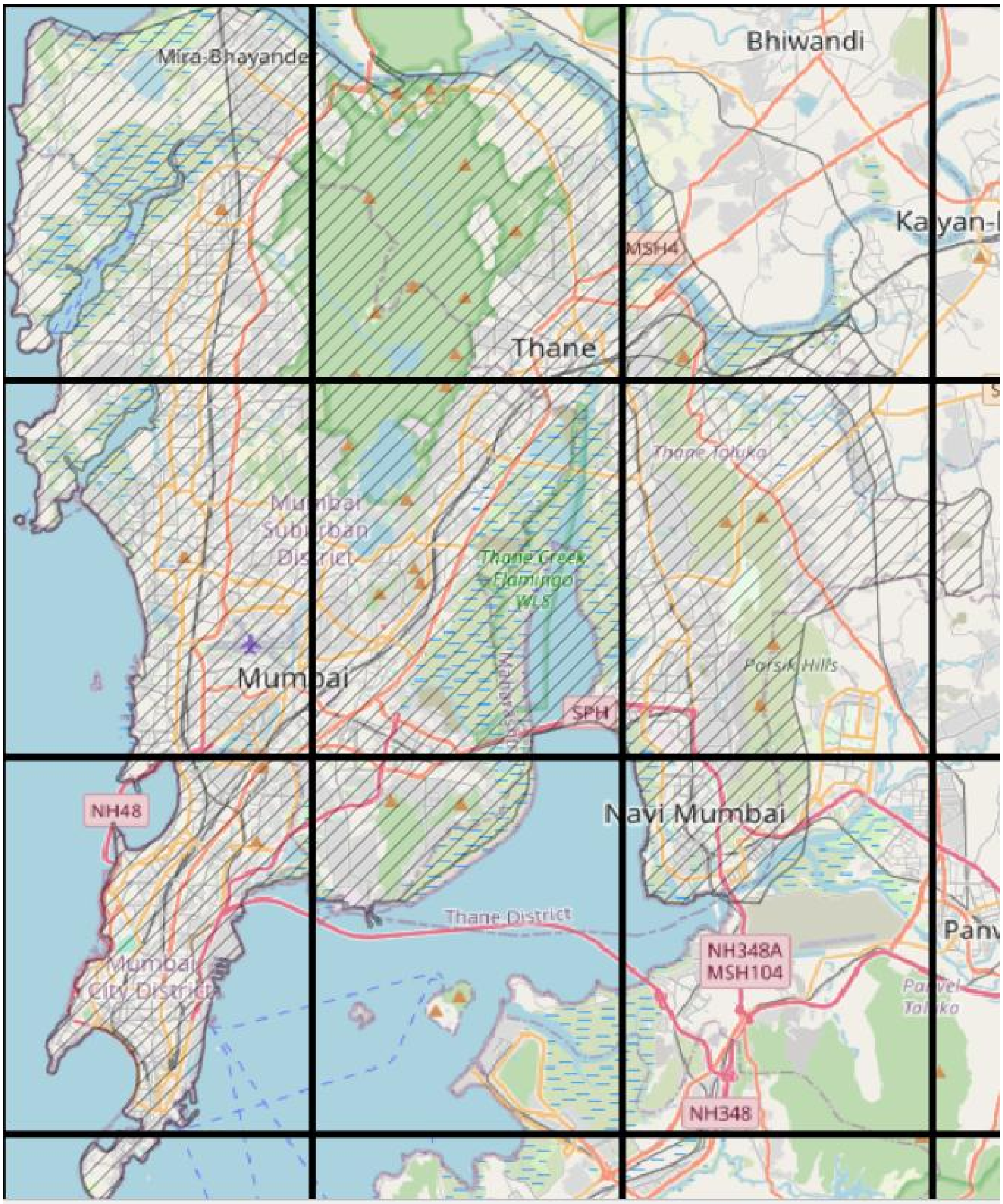
NDVI Dataset:

	Tile11	Tile12	Tile13	Tile21	Tile22	Tile23	Tile31	Tile32	Tile33
2014	0.1559828	0.1729263	0.1148708	0.1224606	0.1164066	0.1225268	0.09697006	0.1656725	0.093279
2015	0.152343	0.1818141	0.1258913	0.1240975	0.1355193	0.1272452	0.1056613	0.1806233	0.1081881
2016	0.1516478	0.1710769	0.12734851	0.11280299	0.118985	0.11930899	0.07932594	0.1711468	0.09241157
2017	0.1325373	0.1599971	0.1079512	0.09689338	0.09804141	0.1032044	0.06516312	0.12839128	0.0759551
2018	0.1631191	0.1844998	0.12565766	0.12499835	0.125722	0.1351223	0.09001273	0.180371	0.09424168
2019	0.1528368	0.1631997	0.10432066	0.1175454	0.1131071	0.11501967	0.09218461	0.168771	0.09625275
2020	0.1653688	0.1659653	0.1210838	0.12243753	0.1206684	0.1283117	0.09703571	0.1677421	0.100653
2021	0.148784	0.170896	0.1286279	0.1174705	0.1127226	0.1275976	0.0905047	0.1536074	0.1008836
2022	0.1702084	0.1804485	0.146979	0.1394302	0.1291002	0.1496581	0.1067139	0.1687079	0.1127953
2023	0.1505594	0.1641133	0.11815319	0.11172422	0.1190742	0.12297163	0.08618033	0.1452245	0.09953609
2024	0.1648122	0.1865116	0.1424807	0.12144908	0.1341325	0.1412168	0.08618909	0.1707248	0.1104235

NDBI Dataset:

	Tile11	Tile12	Tile13	Tile21	Tile22	Tile23	Tile31	Tile32	Tile33
2014	-0.042495	-0.047079	-0.024293	-0.021726	-0.042216	-0.0013141	-0.0105039	-0.0596729	-0.0209086
2015	-0.040241	-0.052171	-0.021274	-0.021856	-0.057762	-0.0043975	-0.0152411	-0.0723650	-0.0273552
2016	-0.043182	-0.043883	-0.023562	-0.018926	-0.046923	0.00096504	-0.0225305	-0.0681488	-0.0205190
2017	-0.038401	-0.041604	-0.027149	-0.020275	-0.047011	-0.0059558	-0.0125082	-0.057517	-0.0315288
2018	-0.057607	-0.065687	-0.040386	-0.035413	-0.058626	-0.0148372	-0.0244325	-0.0889686	-0.0409768
2019	-0.046695	-0.045133	-0.021431	-0.023958	-0.047211	-0.0026875	-0.0162537	-0.0713715	-0.0317307
2020	-0.050178	-0.041366	-0.026949	-0.022402	-0.043424	-0.0030884	-0.0138329	-0.0638220	-0.0251580
2021	-0.03568	-0.041194	-0.030270	-0.022265	-0.039796	-0.0040572	-0.013729	-0.0566912	-0.0257343
2022	-0.050936	-0.046425	-0.039801	-0.031378	-0.047365	-0.0174554	-0.0151938	-0.0610637	-0.03341425
2023	-0.038740	-0.043225	-0.029912	-0.021098	-0.044672	-0.0092113	-0.0130876	-0.0493589	-0.029869286
2024	-0.055650	-0.061276	-0.044188	-0.035653	-0.058256	-0.0213531	-0.0290587	-0.0784353	-0.04481466

Data	NDVI and NDBI indexes of Mumbai and neighbouring regions
Start Year	2014
End Year	2024
Number of observation	99 each of NDVI and NDBI images
Frequency	Quarterly

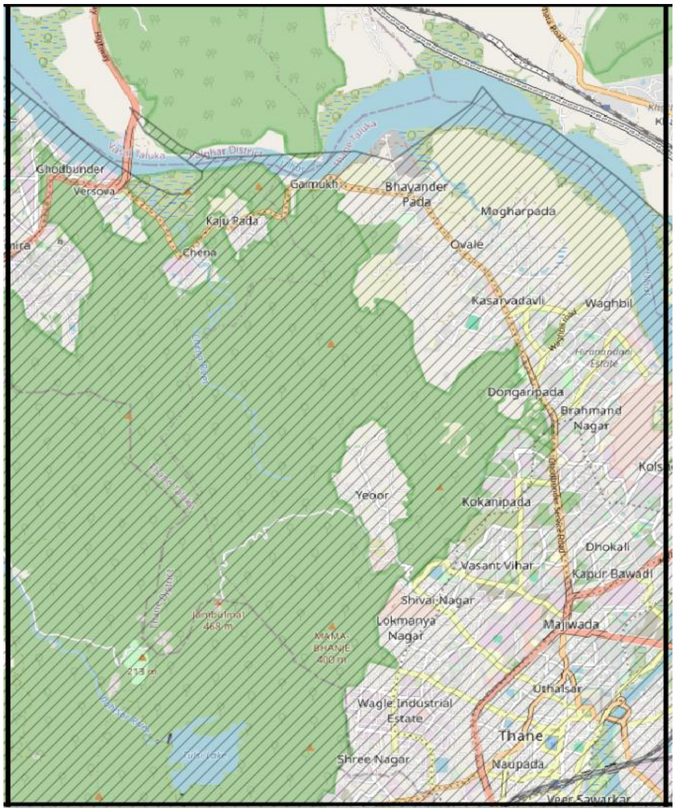


Variable Description:

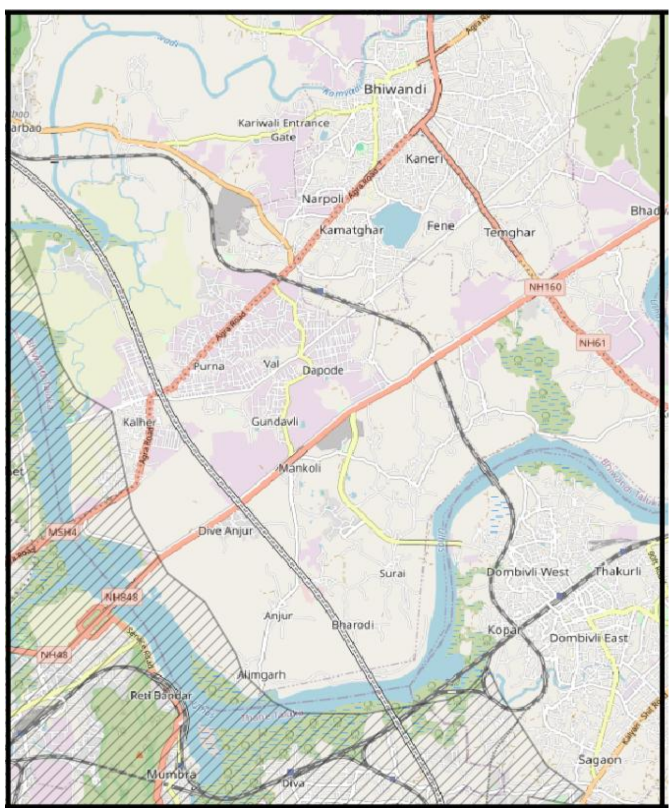
Tile11	Area of Mira-Bhayander and (Borivali & Kandivali) suburb of Mumbai
Tile12	Area of (Versova & Shreenagar) from Mumbai and Kapur bawadi from Thane
Tile13	Area of (Mumbra & Kalwa) from thane
Tile21	Area of (Madh, Bandra & Malad) from Mumbai
Tile22	Area of (Mulund, Chandivali & Govandi) from Mumbai
Tile23	Area of (Ghansoli, Sanpada , Ghatkopar) form Mumbai
Tile31	Area of (Koliwada, Dadar, Wadala) from Mumbai
Tile32	Area of (Chembur & Cheeta Camp) form Mumbai
Tile33	Area of (Belapur & Nerul) From Navi Mumbai

<u>Tile11</u> :

<u>Tile12 :</u>

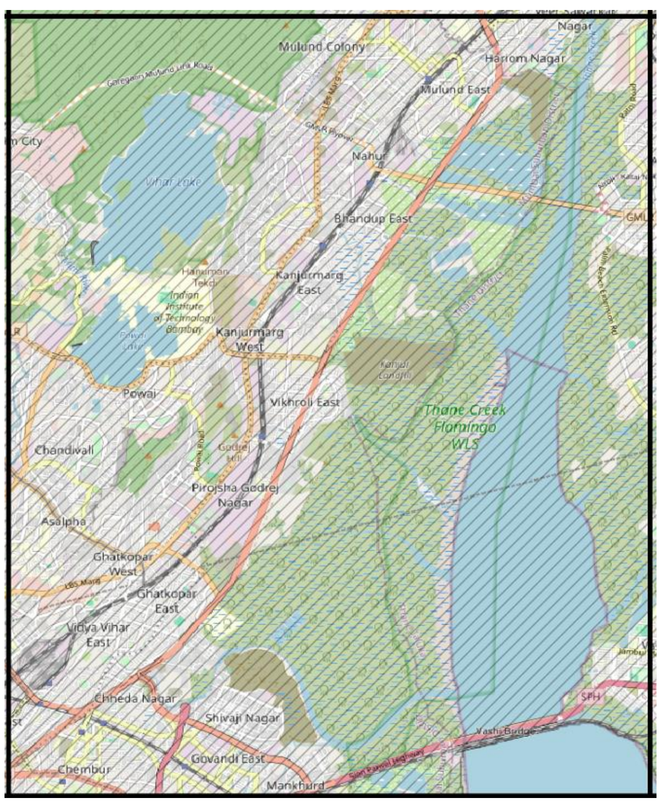


Tile13 :

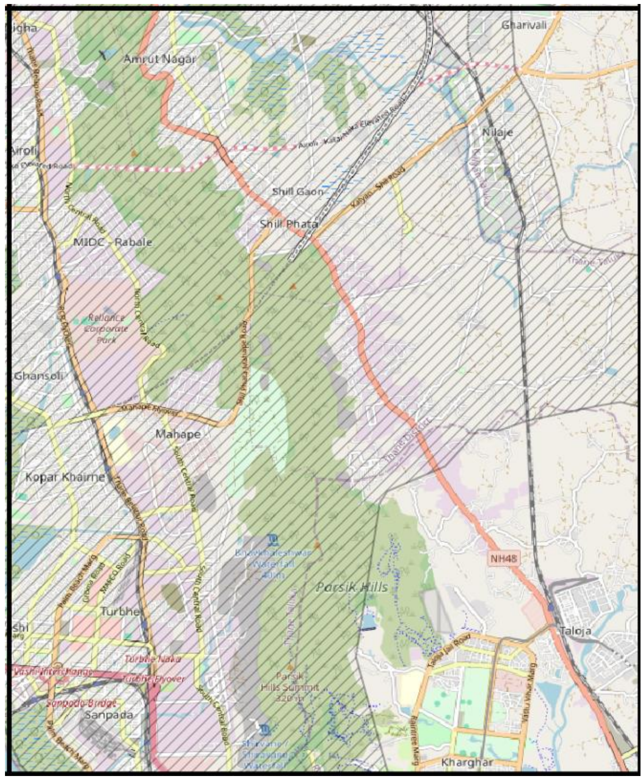


Tile21 :

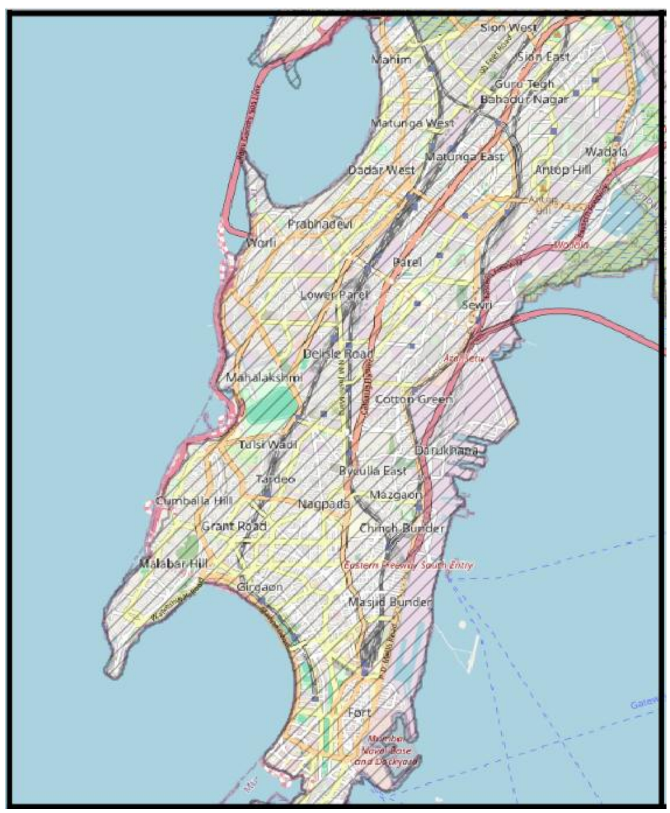
Tile22 :



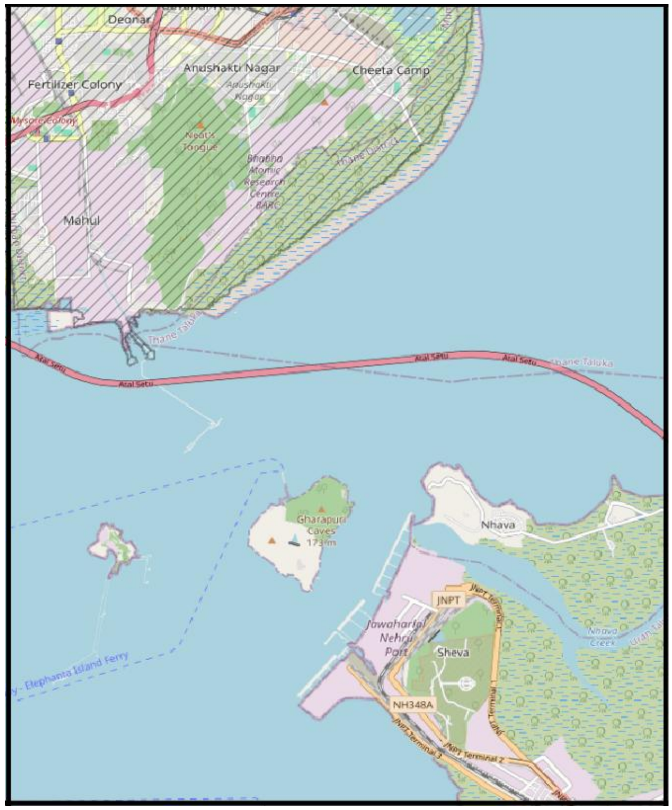
Tile23 :



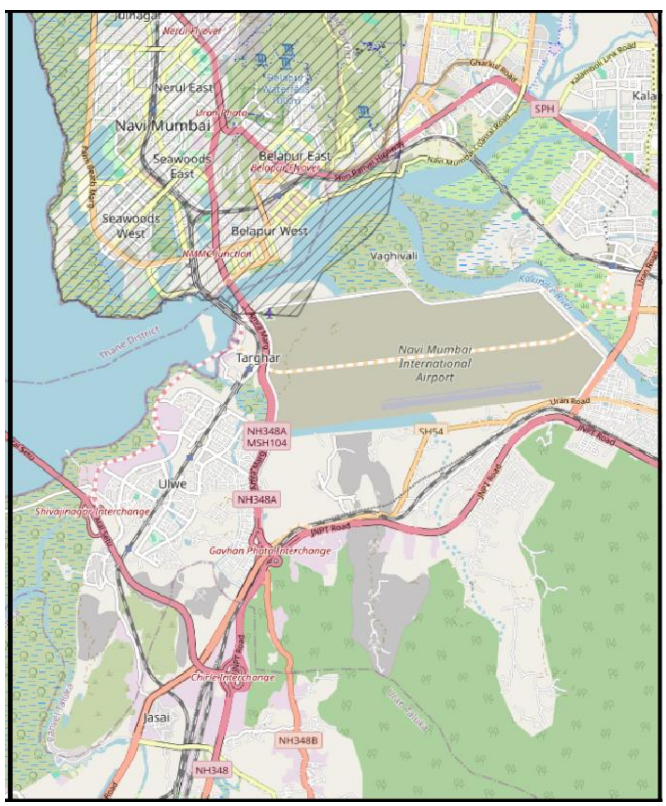
Tile31 :



Tile32 :



Tile33 :



Methods Used :

1. **<u>Trend Analysis</u> :**

For tile(i,j) we fitted a linear model (for studying trend) of the form: --

$$y_{t_{ij}} = a_{ij} + b_{ij}t + e_{t_{ij}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(*)$$

i = 1,2,3
j = 1,2,3
$y_{t_{ij}}$= NDVI/NDBI value for the Tile (i,j) for the year t
t = 2014, 2015, … ,2024
a_{ij}, b_{ij} = intercept and slope parameters for the tile (i,j)
$e_{t_{ij}}$ = error term corresponding to year t and tile (I,j)
The assumptions related to this model are: --

i) $e_{t_{ij}} \sim N(0,\sigma^2)$, for all t=2014, 2015, … ,2024
ii) The errors are not auto correlated,
iii) t should be variable enough,
iv) The value of t is fixed in repeated sample.

Checking the assumptions :

i. **<u>Normality assumption</u>** :

<u>Shapiro-Wilk Test for Normality</u> : ($e_{t_{ij}} \sim N(0,\sigma^2)$)

Null hypothesis: W is equal to 1 ($e_{t_{ij}}$ are normally distributed)
Alternative hypothesis: W is not equal to 1 ($e_{t_{ij}}$ are normally distributed)

Level of sig. is α.

where W is Shapiro Wilk test statistic which is defined as follows: --
If $X_1, X_2, \ldots\ldots\ldots, X_n$ be the n sample observations and $X_{(1)}, X_{(3)}, \ldots\ldots\ldots, X_{(n)}$ be the n corresponding order statistic, then W is given by

$$W = \frac{(\sum_{i=1}^{n} a_i X_{(i)})^2}{(\sum_{i=1}^{n} (X_{(i)} - \bar{X})^2}$$

Where $X_{(i)}$ is the i^{th} order statistic, i.e., the i^{th} smallest number in the sample, and $\bar{X}$ is the sample mean. The coefficients a_i are given by:

$$(a_1, a_2, \ldots, a_n) = \frac{m^T V^{-1}}{C}$$

Where C is a vector norm:

$$C = \|V^{-1}m\| = (m^T V^{-1} V^{-1} m)^{1/2}$$

And the vector m:

$$m = (m_1, \ldots, m_n)^T$$

is made of the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution; finally, V is the covariance matrix of those normal order statistics.

▶ In practice, the test is simple to apply on a computer using R. Namely, let X = (X_1, ..., X_n) be the data vector, represented in R if entered individually as c (X_1, ..., X_n).

▶ In R, one simply types Shapiro. Test(X).

▶ As output, the value of the test statistic W and a p-value are obtained. If the p-value is less than, say, the conventional level 0.05, then one rejects the normality hypothesis, otherwise one doesn't reject it.

ii. No Autocorrelation Test:

Durbin-Watson Test:

The Durbin-Watson Statistic is a test statistic used to detect the presence of autocorrelation at lag 1 in the residuals from a regression analysis.

Durbin-Watson test statistic is:

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Where T is the number of observations. For large T, d is approximately equal to 2(1-$\hat{\rho}$), where $\hat{\rho}$ is the sample autocorrelation of the residuals at lag 1.

The mechanics of the Durbin–Watson test are as follows:

1. Run the OLS regression and obtain the residuals.
2. Compute d from the equation.
3. For the given sample size and given number of explanatory variables, find out the critical d_L and d_U values.
4. Now follow the decision rules given in Table.

Decision rules:

Null Hypothesis	Decision	If
No positive autocorrelation	Reject	$0 < d < d_L$
No positive autocorrelation	No decision	$d_L \le d \le d_U$

No negative autocorrelation	Reject	4 - d_L< d < 4
No negative autocorrelation	No decision	4 - $d_U \leq d \leq$ 4 - d_L
No autocorrelation, positive or negative	Do not reject	$d_U < d <$ 4 - d_U

One can use the following modified d test:

Given the level of significance α,

1. H_0: ρ = 0 versus H_1: ρ > 0. Reject H_0 at α level if d < d_U. That is, there is statistically significant positive autocorrelation.
2. H_0: ρ = 0 versus H_1: ρ < d_U, that is, there is statistically significant evidence of negative autocorrelation.
3. H_0: ρ = 0 versus H_1: ρ = 0. Reject H_0 at 2α level if d < d_U or (4– d) < d_U, that is, there is statistically significant evidence of autocorrelation, positive or negative.

iii. <u>Correlation:</u>

Correlation is the quantitative measure of the relationship between the two variables, if the change in one variable affects a change in other variable. Than the variables are said to be correlated. If the two variables deviate in the same direction, i.e., if the increase (or decrease) in one results in a corresponding increase (or decrease) in the other, directions, i.e., if increase (or decrease) in one results in the corresponding decrease (or increase) in the other, correlation is said to be diverse or negative.

Say (x1,y1);(x2,y2);………..;(xn,yn) be n paired dataset. then the sample correlation coefficient is defined by,

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$
$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{[\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2]^{\frac{1}{2}}}$$

Where $\bar{x}$ and $\bar{y}$ are the sample means.

Result and Discussion :

1) Plotting the NDBI and NDVI data for different tiles over time we get: --

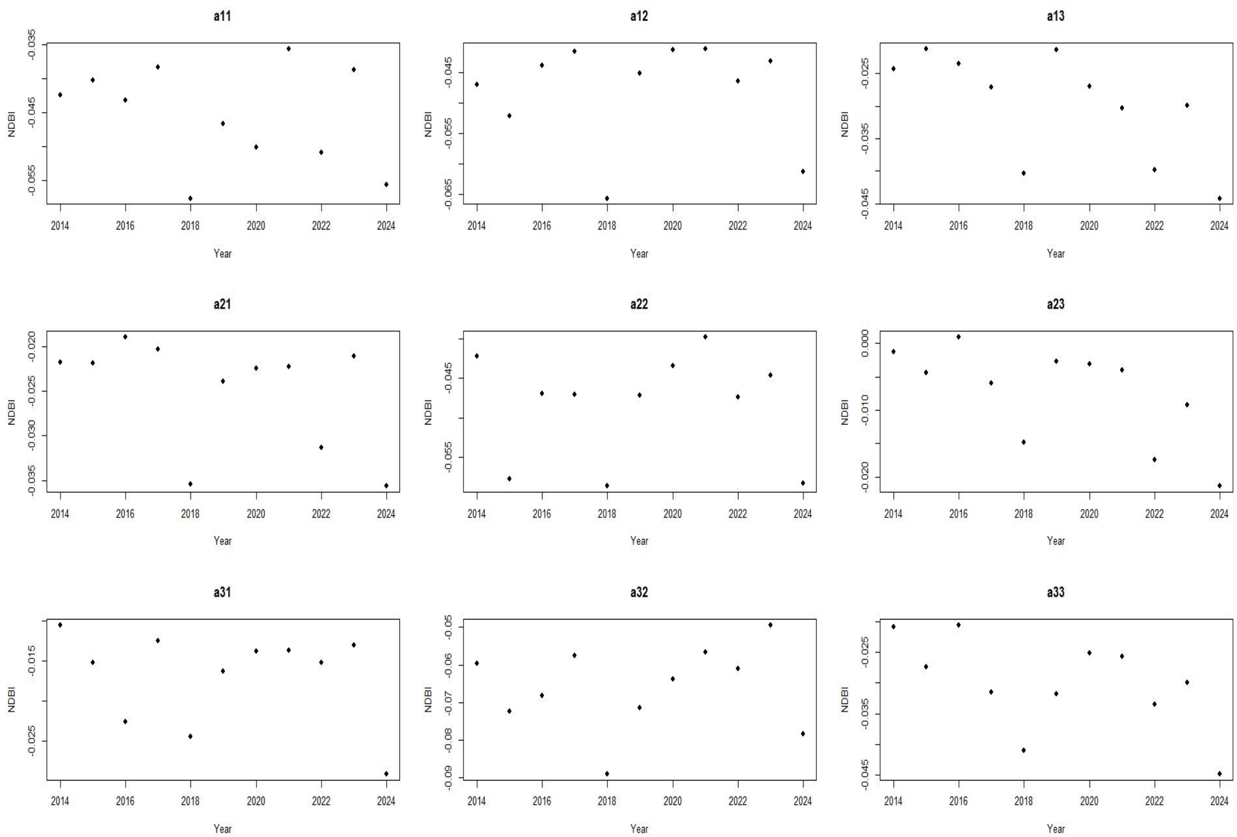


Fig: -- plot for NDBI

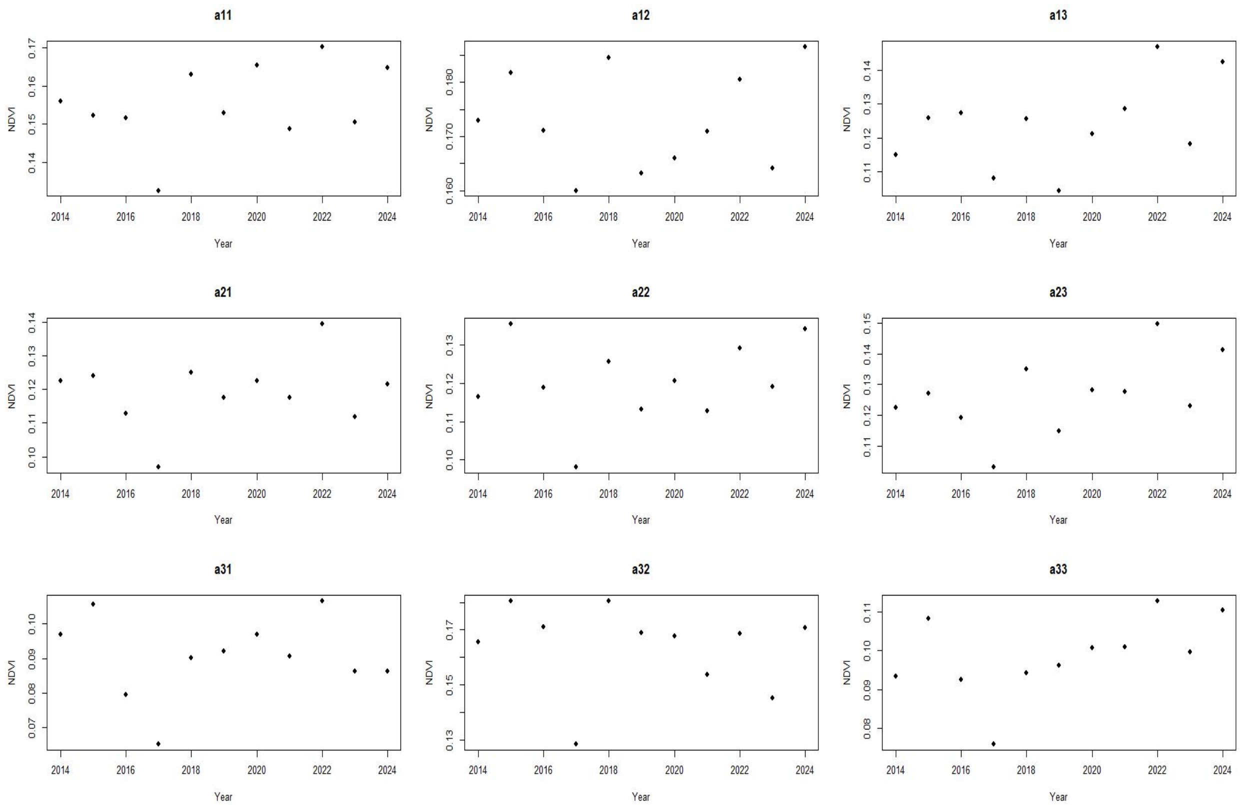


Fig: -- plot for NDVI

From the above plots it is not certain that there is no trend in the NDVI and NDBI values for different tiles over time.

2) Trend Analysis :

The model (*) is fitted for both NDVI and NDBI values separately. Therefore, there are two models of the form (*), one for NDVI and another for NDBI.

Now, at first, we are to check the assumptions of the model (*) for both NDVI and NDBI separately because if any one of the assumptions does not hold, then the inferences from the data obtained by running the model may go wrong.

Checking the assumptions :

a) Normality assumption :

Shapiro-Wilk Test :

For both the NDVI and NDBI models separately we are interested in testing,
Null hypothesis: W is equal to 1 ($e_{t_{ij}}$ are normally distributed)
against
Alternative hypothesis: W is not equal to 1 ($e_{t_{ij}}$ are normally distributed)

The following values are obtained :

Tile	W	p-value	W	p-value
11	0.92644	0.376	0.98131	0.973
12	0.91323	0.2662	0.81901	0.01674
13	0.95587	0.7193	0.96817	0.8675
21	0.96542	0.837	0.93918	0.5109
22	0.98253	0.9786	0.85202	0.04535
23	0.97506	0.9325	0.90596	0.2183
31	0.94455	0.5754	0.82039	0.01745
32	0.85539	0.05017	0.94454	0.5753
33	0.92843	0.3952	0.92	0.3186

For both the cases, p-value of the Shapiro-Wilk test for all the tiles are less than 0.01, therefore at 1% level of significance null hypothesis cannot be rejected. Thus, for both the models it can be concluded that the errors are normally distributed.

b) No Auto correlation Assumption :

Durbin-Watson Test :

For both the NDVI and NDBI models separately we are interested in testing
Null hypothesis: The errors are not auto correlated
against
Alternative hypothesis: The errors are auto correlated.

The following values are obtained: --

	NDVI		NDBI	
Tile	d	p-value	d	p-value
11	3.0801	0.09636	8473	0.2432
12	2.6796	0.4022	2.1508	0.9014
13	2.5262	0.5822	2.6156	0.4738
21	2.4871	0.6324	2.6018	0.4898
22	2.4569	0.6721	2.0167	0.7184
23	2.8966	0.205	2.2858	0.908
31	1.6939	0.3484	2.0541	0.7683
32	2.8025	0.2814	2.1896	0.9558
33	1.9095	0.582	1.7058	0.3597

For both the cases, p-value of the Durbin-Watson test for all the tiles are less than 0.01, therefore at 1% level of significance null hypothesis cannot be rejected. Thus, for both the models it can be concluded that the errors are not autocorrelated.

Therefore, the assumptions hold true for the models of both NDVI and NDBI separately.

Now, we are interested in finding whether there is any significant trend in the NDBI and NDVI values for different tiles. Therefore, we are interested in testing whether the slope parameter for both the NDVI and NDBI models(separately) are significantly different from 0 or not.

Thus, for both the NDBI model and NDVI model we are interested in testing

Null hypothesis: The slope parameter is not significant (b_{ij} = 0)

against

Alternative hypothesis: The slope parameter is not significant ($b_{ij} \neq 0$)

At level of significance 1%

$$\textbf{Test Statistic} = \frac{\hat{b}_{ij} - b_{ij}}{\boldsymbol{S.E}(\hat{b}_{ij})} \sim t_{n-2} \text{ under } H_0$$

The following values are obtained :

a) **For NDBI model**

Tile	$\hat{b}_{ij}$	$\boldsymbol{S.E}(\hat{b}_{ij})$	t- value	p-value	Decision
11	-0.00064	0.000705	-0.90487	0.389117	Do not reject null hypothesis
12	-0.00016	0.000832	-0.1931	0.85117	Do not reject null hypothesis
13	-0.0016	0.000608	-2.62434	0.027614	Do not reject null hypothesis
21	-0.00086	0.000542	-1.59168	0.14592	Do not reject null hypothesis
22	4.25E-06	0.000672	0.006325	0.995091	Do not reject null hypothesis
23	-0.00145	0.000545	-2.65609	0.026213	Do not reject null hypothesis
31	-0.00049	0.00056	-0.87711	0.403234	Do not reject null hypothesis
32	0.000421	0.00112	0.375646	0.715881	Do not reject null hypothesis
33	-0.00128	0.000634	-2.02124	0.073974	Do not reject null hypothesis

Decision: p-value for all the tiles is less than 0.01 suggesting that we cannot reject null hypothesis. Therefore, we cannot conclude that there is any trend in the NDBI values for any tiles.

b) **For NDVI model**

Tile	$\hat{b}_{ij}$	$S.E(\hat{b}_{ij})$	t- value	p-value	Decision
11	0.0011585	0.000974	1.189374	0.264723	Do not reject null hypothesis
12	0.0002591	0.000922	0.28094	0.785108	Do not reject null hypothesis
13	0.0018433	0.001148	1.605171	0.142919	Do not reject null hypothesis
21	0.0005811	0.001033	0.562674	0.587397	Do not reject null hypothesis
22	0.0007046	0.001053	0.668931	0.520327	Do not reject null hypothesis
23	0.0019034	0.001095	1.737515	0.116304	Do not reject null hypothesis
31	7.31E-05	0.001182	0.061831	0.952049	Do not reject null hypothesis
32	-0.00078	0.001545	-0.50514	0.625596	Do not reject null hypothesis
33	0.0015321	0.000893	1.715974	0.120302	Do not reject null hypothesis

Decision: p-value for all the tiles is less than 0.01 suggesting that we cannot reject null hypothesis. Therefore, we cannot conclude that there is any trend in the NDVI values for any tiles.

3) Scatter plot for NDVI and NDBI for different years: --

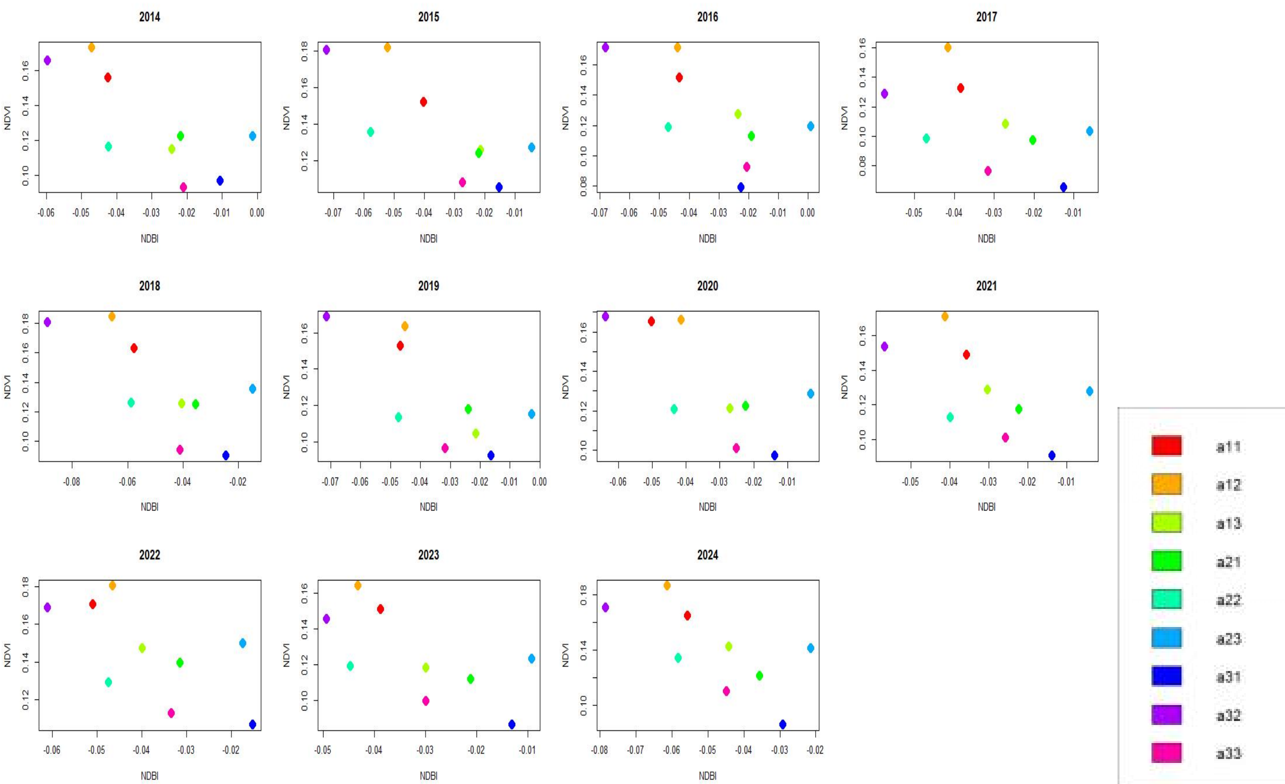


Fig: --Scatter plot for NDBI and NDVI

From the above scatter it seems that for each year the NDVI and NDBI values moves in opposite direction.

4) Correlation between NDVI and NDBI values for each year: --

First of all, the Standard Deviation of the NDVI and NDBI values for each year across different tiles are found which are as follows: --

Year	NDBI	NDVI
2014	0.018922	0.029041
2015	0.022312	0.028148
2016	0.020451	0.032197
2017	0.016677	0.029213
2018	0.022679	0.034003
2019	0.020663	0.029106
2020	0.01901	0.027618
2021	0.01574	0.025657
2022	0.014272	0.025657
2023	0.017716	0.025009
2024	0.177165	0.031383

Since there is a variation among the NDBI values and the NDVI values respectively for each year across different tiles, the correlation between NDBI and NDVI values can be found out and are as follows:

Year	Correlation
2014	-0.7487045
2015	-0.7779217
2016	-0.6759314
2017	-0.5465218
2018	-0.7213805
2019	-0.7383303
2020	-0.7227017
2021	-0.5953006
2022	-0.633905
2023	-0.6561814
2024	-0.6704456

Thus, for each year the correlation between NDBI and NDVI values are found out to be negative which suggests that if NDBI increases then NDVI decreases and vice-versa that is, they move in opposite direction.

CONCLUSION :

In our present study, NDVI and NDBI values has been calculated using Landsat 8 OLI/TIRS data for the regions under consideration and the findings are as follows:

i) There is no significant trend in the respective values of NDVI and NDBI over time (2014-2024) for different regions under consideration.

ii) NDVI and NDBI have a strong negative correlation that is, if NDBI increases then NDVI decreases and vice-versa for time period 2014-2024 which suggests that the areas with higher built-up or urban density have lower vegetation cover and vice-versa.

Moreover, this project can be further implemented on other areas of INDIA and the trend of NDVI and NDBI values and their correlation can be studied for better understanding about the land use cover dynamics particularly urban areas and agricultural lands. Other factors like LST, NDWI etc. can also be considered for future research. Other than this the factors which can be considered for farther research are as follows: --

a) Increasing the number of tiles,

b) Panel analysis on the NDVI and NDBI data.

REFERENCES :

1. Hai Tao, Bassim Mohammed Hashim,Salim Heddam, Leonardo Goliatt, Mou Leong Tan, Zulfaqar Sa'adi, Iman Ahmadi Nafar, Mayadah W. Falah, Bijay Halder, Zaher Mundher Yaseen. Megacities' environmental assessment for Iraq region using satellite image and geo- spatial tools. This research studies the remote sensing-based on Landsat images that are used for investigating the vegetation circumstances of megacities of Iraq.

2. Shoukat Ali Shaha, Madeeha Kiranb, Aleena Nazirc, Shaharyar Hassan Ashrafani. EXPLORING NDVI AND NDBI RELATIONSHIP USING LANDSAT 8 OLI/TIRS IN KHANGARH TALUKA, GHOTKI. This study investigates the relationship between Normalized Difference Vegetation Index (NDVI) and Normalized Difference Built-up Index (NDBI) in Khangarh taluka.