INDIAN INSTITUTE OF INFORMATION TECHNOLOGY ALLAHABAD
Dept. of Information Technology

Geospatial Data Analysis of Prayagraj

Group Members

1.Shubham Kumar Bhokta
BTech IT - IIIT Allahabad
iit20200007@iiita.ac.in
(IIT2020007)

2.Anoop Kumar(IIT2020020)
B.Tech IT - IIIT Allahabad
iit2020020@iiita.ac.in

3.Shashikant Thakur(IIT2020024)
B.Tech IT - IIIT Allahabad
iit2020024@iiita.ac.in

Project Supervisor
Prof. Triloki Pant

I. *Abstract*

This report delves into the application of Sentinel-2 satellite imagery for the classification of land areas in Prayagraj into four distinct categories. Leveraging remote sensing techniques and machine learning algorithms, the study aimed to accurately delineate land cover types across the region. The methodology encompassed several key steps, including preprocessing of Sentinel-2 multispectral images, feature extraction, and classification utilizing supervised learning techniques such as Support Vector Machine (SVM) and Random Forest (RF). The performance of each classification algorithm was rigorously evaluated using accuracy assessment metrics. Additionally, the results were compared with ground truth data to validate the accuracy of the classification process.

The study highlights the effectiveness of Sentinel-2 imagery coupled with machine learning algorithms in accurately classifying land areas in Prayagraj. The classification process facilitated the categorization of land into four main classes, namely urban areas, agricultural land, water bodies, and vegetation cover. By leveraging spectral information captured by Sentinel-2, distinct patterns associated with different land cover types were identified and utilized for classification purposes.

The findings of this study have significant implications for various domains, including urban planning, environmental monitoring, and sustainable development in Prayagraj. Accurate land classification enables policymakers and urban planners to make informed decisions regarding land use management, infrastructure development, and natural resource conservation. Moreover, the insights derived from this study can aid in assessing the impact of urbanization on the environment and identifying areas susceptible to environmental degradation.

Introduction:

Urbanization is a defining characteristic of the contemporary world, driving the need for effective management of land resources within cities, particularly in light of escalating environmental concerns. Satellite imagery has emerged as a powerful tool in this endeavor, offering extensive coverage and high-resolution data for analyzing urban land areas. This project report focuses on leveraging satellite data to classify land areas within cities, employing advanced machine learning algorithms for precise categorization.

Urban land classification serves a multitude of purposes, including urban planning, environmental monitoring, and resource management. Accurate identification of different land cover types, such as residential, commercial, industrial, green spaces, and water bodies, is essential for informed decision-making regarding land use, infrastructure development, and conservation efforts. Satellite data presents a cost-effective and scalable solution for this task compared to traditional ground-based surveys, thanks to its ability to capture detailed spatial information over vast areas.

The project utilizes satellite imagery and sophisticated machine learning techniques to automatically classify land areas within urban environments. By extracting relevant features from satellite data, such as spectral signatures, texture, and spatial patterns, the classification model aims to accurately distinguish between various land cover classes. Integration of machine learning algorithms, including supervised and unsupervised methods, empowers the model to learn patterns from labeled training data and generalize its classification capabilities to unseen areas.

Moreover, the project delves into the challenges associated with land classification using satellite data, including issues such as cloud cover, seasonal variations, and spectral confusion between land cover classes. Overcoming these challenges requires meticulous preprocessing of satellite imagery, feature selection, and algorithm optimization to enhance classification accuracy and robustness.

In summary, this project contributes to the burgeoning field of research on utilizing satellite data and machine learning techniques for urban land classification. By shedding light on the methodology, challenges, and potential applications of land classification within cities using satellite data, this study aims to provide stakeholders with invaluable insights for making informed decisions and fostering sustainable urban development.

II. Literature Review

S.No	*Authors*	*Paper title*	*Description*	*Methodology*	*Result*
1.	M.Cavur, H.S.Duzgun, S.Kemec , D.C. Demirkan	LAND USE AND LAND COVER CLASSIFICATION OF SENTINEL 2-A: ST PETERSBURG CASE STUDY	The paper presents a methodology for land use and land cover (LULC) classification using Sentinel-2A satellite imagery in a case study of St Petersburg. It employs an object-oriented approach with Supported Vector Machines (SVM) for accurate classification. The study demonstrates the suitability of Sentinel-2 data for LULC mapping with an overall accuracy of 83.64% and a kappa coefficient of 0.802.	Data Collection involved acquiring Sentinel 2A satellite images from ESA Sentinel Online. These images were then Preprocessed by resampling them to 10 meters using SNAP software. In the First-Level Classification stage, NDVI and NDWI were created and used to classify the images into water, vegetation, bare land, and built-up areas employing SVM. Subsequently, Second-Level Classification and Class Combination were performed to refine and merge classes, followed by the application of SVM for detailed land use land cover mapping. Accuracy Assessment was carried out by checking classification accuracy with ground truth points and calculating overall accuracy and Kappa coefficient.	Overall accuracy of 83.64%
2.	M. Majidi Nezhad1 , A. Heydari1 , L. Fusilli , G. Laneve	Land Cover Classification by using Sentinel-2 Images: A case study in the city of Rome	The paper presents a study on land cover classification using Sentinel-2 images in the city of Rome, Italy. It compares the performance of two classification algorithms, Maximum Likelihood (ML) and Support Vector Machine (SVM), in mapping urban, forest, water, agriculture, and empty land classes. The study employs ENVI software for classification and evaluates accuracy using confusion matrices. Results indicate that the ML method generally outperforms SVM, with improvements attributed to post-classification processes such as Sieve Classes, Clump Classes, and Majority/Minority Analysis. The study highlights the importance of accurate land cover mapping for various applications, emphasizing the relevance of satellite-based observations like Sentinel-2 in	The methodology involves using Sentinel-2 images to classify land cover in Rome using Maximum Likelihood (ML) and Support Vector Machine (SVM) algorithms. It includes preprocessing of images, selection of training samples, application of classification algorithms, and assessment of accuracy using a confusion matrix. Post-classification processing techniques are also used to improve accuracy.	Accuracy : MLE : 92.7% SVM : 85.25%

S.No	Authors	Paper title	Description	Methodology	Result
			urban management and environmental monitoring.		
3.	A. Sekertekin, A.M. Marangoz, and H. Akcin	PIXEL-BASED CLASSIFICATION ANALYSIS OF LAND USE LAND COVER USING SENTINEL-2 AND LANDSAT-8 DATA	The paper investigates the accuracy of land use and land cover (LULC) classifications using Sentinel-2 and Landsat-8 satellite data in Zonguldak, Turkey. It compares the classification results derived from these datasets and evaluates their effectiveness in representing LULC features. The study employs pixel-based Maximum Likelihood supervised classification and conducts accuracy assessments using kappa statistics. Results indicate that Sentinel-2 generally provides more accurate LULC images compared to Landsat-8, although Landsat-8 performs better in certain areas. The research underscores the importance of accurate LULC data for sustainable land management and urban planning, highlighting the role of remote sensing technologies in monitoring Earth's changes.	The methodology employed in the study involved selecting Zonguldak city and its surroundings as the study area, acquiring Sentinel-2 MSI and Landsat-8 OLI satellite imagery, preprocessing the data including band stacking and pan-sharpening, conducting pixel-based Maximum Likelihood Classification (MLC) using training samples collected for various land cover classes, assessing classification accuracy through stratified random points, and comparing the results between Sentinel-2 and Landsat-8 datasets. This comprehensive approach aimed to determine which dataset provided better Land Use Land Cover (LULC) classification results for the study area.	Overall Accuracy: Landsat-8 LULC: 83.91% Sentinel-2 LULC: 88.74%
4.	Muhammad Nasar Ahmad, Zhenfeng Shao, Akib Javed	Mapping impervious surface area increase and urban pluvial fooding using Sentinel Application Platform (SNAP) and remote sensing data	The paper investigates the correlation between the increase in impervious surface area (ISA) and urban pluvial flooding (UPF) in three major cities of Pakistan: Islamabad, Lahore, and Karachi. Using remote sensing data from Landsat and Sentinel-1, the study maps the growth of ISA from 1992 to 2022 and identifies areas prone to UPF. The findings reveal significant ISA expansion in all cities, with Lahore experiencing the highest overall accuracy in mapping9. Moreover, Sentinel-1 data coupled with the SNAP tool effectively detect flooded areas, indicating a direct link between ISA growth and UPF occurrence. The research underscores the importance of monitoring ISA for urban planning and recommends modeling-based solutions for	The methodology used Landsat data for impervious surface area extraction and Sentinel-1 SAR data for urban pluvial flooding mapping. Landsat data underwent pre-processing steps including cloud removal and water body delineation, while Sentinel-1 data was pre-processed using SNAP. Analysis involved spectral indices and thresholding techniques for ISA extraction, and visual interpretation for UPF mapping. The study utilized Google Earth Engine for Landsat data processing and SNAP for Sentinel-1 data processing.	Overall Accuracy : Lahore: 0.93 Karachi: 0.86 Islamabad: 0.85

S.No	*Authors*	*Paper title*	*Description*	*Methodology*	*Result*
			identifying high-risk areas prone to UPF events.		
5.	ALGhaliya Nasser Mohammed Al-Rubkhi	Land Use Change Analysis and Modeling Using Open Source (QGIS)	The paper examines Land Use and Land Cover (LULC) changes between 2000 and 2010, using QGIS and GIS techniques, with a focus on urbanization trends in the Bawshar Wilayat, Oman. It utilizes the MOLUSCE plugin for mapping changes and transition matrix analysis, along with Cellular Automata Simulation for future projections. The study aims to understand the drivers of LULC changes and predict future trends, considering factors like population growth. Previous studies on LULC changes are also reviewed for context.	The methodology employed in this study involved the utilization of GIS data, including land use maps and spatial variables, to analyze land cover changes in Boasher willayat. The MOLUSCE Plugin in ArcMap was utilized, employing methods such as Artificial Neural Network (ANN), Logistic Regression (LR), and Cellular Automata Simulation to model land use transition potential. Data preparation included converting vector to raster formats, ensuring uniform coordinate systems, pixel sizes, and fixed extents. The process flow involved database development, correlation evaluation, area change analysis, transition potential modeling, Cellular Automata simulation, and validation. QGIS, ArcMap, and Microsoft Excel were utilized for data processing, analysis, and visualization. This methodology provided a comprehensive approach to understanding land use dynamics and urbanization trends in the study area.	Accuracy of ANN: 88.96% , Accuracy of Logistic Regression: 92.79%

S.No	Authors	Paper title	Description	Methodology	Result
06	Simona Niculescu, Antoine Billey, Halima Talab-Ou-Ali	: "Random Forest Classification for Vegetation Monitoring in Pays de Brest, France" l	The paper discusses a study on the coastal areas in the Pays du Brest, focusing on the classification of satellite images from Sentinel-1, Sentinel-2, and SPOT-6 using the Random Forest algorithm. The study utilizes time series analysis and stacking of results from different satellite images and vegetation indices. It also involves the comparison of image classifications and results between different satellite sources. The study site is in the coastal zone, and the images used for analysis were from the year 2015. Additionally, the paper mentions the availability of specific Sentinel-2 images and their respective acquisition dates and correction levels.	The article discusses the use of Random Forest classification with Sentinel-1 and Sentinel-2 images for vegetation monitoring in the Pays de Brest, France. The study site is the coastal areas in the Pays du Brest, and the methodology involves pixel-by-pixel classification of images from Sentinel-1, Sentinel-2, and SPOT-6 using the Random Forest algorithm. The study includes the use of various indices of vegetation from the optical images and the results of SPOT images. The article also mentions the dates of acquisition and the level of correction for the selected Sentinel-2 images.	Overall Accuracy: 0.96

III. Datasets

Sentinel-2 Satellite:

Sentinel-2 is an Earth observation mission from the Copernicus Programme, which is operated by the European Space Agency (ESA). This mission primarily focuses on monitoring the variability in land surface conditions. Sentinel-2 is critical for environmental monitoring, agricultural forecasting, and managing natural disasters among other applications.

Launched in two parts, Sentinel-2A on June 23, 2015, and Sentinel-2B on March 7, 2017, the twin satellites enhance data coverage and frequency, providing revisit times of five days at the equator and two to three days at mid-latitudes. This capability is crucial for monitoring rapid changes in Earth's surface caused by natural disasters such as floods or forest fires, and for observing seasonal changes in vegetation.

The Sentinel-2 satellites orbit the earth at an altitude of approximately 786 kilometers, in a sun-synchronous orbit that passes over the equator at the same local solar time on every pass. This consistent timing is ideal for capturing images with similar lighting, which is beneficial for comparing changes over time.

Each satellite is equipped with a Multi-Spectral Instrument (MSI) that captures 13 spectral bands ranging from the visible and the near infrared to the shortwave infrared at different spatial resolutions — 10 meters, 20 meters, and 60 meters. This range of spectral bands allows for detailed observations that can be used for various applications including vegetation classification and monitoring, soil and water cover and observation, inland waterways and coastal areas monitoring, and observation of natural disaster zones.

The data collected by Sentinel-2 is openly and freely accessible to users and researchers, supporting a wide range of applications. For instance, in agriculture, Sentinel-2 data helps in crop monitoring and management by enabling the assessment of crop health, the estimation of crop yields, and the detection of nutrient deficiencies. In forestry, the data aids in forest mapping, disease monitoring, and deforestation tracking.

Additionally, the high resolution of Sentinel-2's imagery makes it useful in urban planning and management. It helps city planners monitor land use changes, manage urban sprawl, and assess water bodies and green spaces within urban areas. Environmental monitoring is another critical application, where Sentinel-2 contributes to tracking changes in land cover and monitoring coastal ecosystems, helping in efforts towards sustainable management of natural resources.

Overall, Sentinel-2 stands as a pillar of the Copernicus Programme's observation fleet, offering unprecedented data accuracy and availability that empower a multitude of sectors and contribute significantly to our understanding and management of the Earth's surface.

IV. SNAP (Sentinel Application Platform)

The Sentinel Application Platform (SNAP) is a versatile software designed by the European Space Agency (ESA) to help users manage and analyze data from the Earth observation satellites within the Copernicus Programme, particularly the Sentinel missions. SNAP works on different operating systems like Windows, macOS, and Linux, making it accessible to many users. The platform is equipped with a range of tools that enable users to process satellite images, such as correcting data errors, analyzing specific details in images, and visualizing complex data in a more understandable form. SNAP supports automation, which means it can handle large amounts of data quickly by running repeated tasks without manual input. It is especially user-friendly, offering both graphical interfaces for casual users and command-line options for more technical users.

SNAP is not only flexible but also customizable, allowing users to add new features or tailor it to specific needs through additional plugins or toolboxes. These extensions make SNAP useful for various applications beyond just satellite data processing. For example, it is widely used in agriculture to monitor crop health and in environmental monitoring to observe changes in forests and water bodies. Educational institutions also use SNAP for teaching and research, helping students and researchers analyze environmental trends and impacts. Future improvements to SNAP may include better integration with cloud technologies, which will allow it to process even larger datasets more efficiently. Overall, SNAP is a key tool in turning satellite data into practical insights used in science, policy-making, and daily management of the environment.

Advantages of using SNAP for land classification

There are several advantages to using SNAP for land classification, including:

- **Open source:** SNAP is an open-source software platform, which means that it is free to use and modify.
- **Cross-platform:** SNAP is available for Windows, macOS, and Linux.
- **Easy to use:** SNAP has a user-friendly interface that makes it easy to perform land classification tasks.
- **Powerful:** SNAP offers a wide range of tools and algorithms for land classification.
- **Accurate:** SNAP has been shown to produce accurate land classification results.

Sentinel 2 Bands [11]

We are using sentinel-2A[6] data

Sentinel-2 carries the Multispectral Imager (MSI). This sensor delivers 13 spectral bands ranging from 10 to 60-meter pixel size.

Its blue (B2), green (B3), red (B4), and near-infrared (B8) channels have a 10-meter resolution.

Next, its red edge (B5), near-infrared NIR (B6, B7, and B8A), and short-wave infrared SWIR (B11 and B12) have a ground sampling distance of 20 meters.

Finally, its coastal aerosol (B1) and cirrus band (B10) have a 60-meter pixel size.

This is a short table for band details in Sentinel-2.

Band	**Resolution**	**Central Wavelength**	**Description**
B1	60 m	443 nm	Ultra Blue
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10m	665 nm	Red
B5	20m	705 nm	VNIR
B6	20m	740nm	VNIR
B7	20m	783 nm	VNIR
B8	10m	842 nm	VNIR
B8a	20m	865 nm	VNIR
B9	60m	940nm	SWIR
B10	60m	1375 nm	SWIR
B11	20m	1610 nm	SWIR
B12	20m	2190 nm	SWIR

Band details:

Band 1 (aerosol)

Resolution = 60m/px

Central Wavelength = 443nm

Bandwidth = 20nm

Use: For aerosol detection.

Band 2 (blue)

Resolution = 10m/px

Central Wavelength = 490nm

Bandwidth = 65nm

Use: Band 2 is useful for soil and vegetation discrimination, forest type mapping and identifying man-made features. It is scattered by the atmosphere, it illuminates material in shadows better than longer wavelengths, and it penetrates clear water better than other colors. It is absorbed by chlorophyll, which results in darker plants.

Band 3 (green)

Resolution = 10m/px

Central Wavelength = 560nm

Bandwidth = 35nm

Use: It gives excellent contrast between clear and turbid (muddy) water, and penetrates clear water fairly well. It helps in highlighting oil on water surfaces, and vegetation. It reflects green light stronger than any other visible color. Man-made features are still visible.

Band 4 (red)

Resolution = 10m/px

Central Wavelength = 665nm

Bandwidth = 30nm

Use: It is strongly reflected by dead foliage and is useful for identifying vegetation types, soils and urban (city and town) areas. It has limited water penetration and doesn't reflect well from live foliage with chlorophyll.

Band 5 (red edge)

Resolution = 20m/px

Central Wavelength = 705nm

Bandwidth = 15nm

Use: For classifying vegetation.

Band 6

Resolution = 20m/px

Central Wavelength = 740nm

Bandwidth = 15nm

Use: For classifying vegetation.

Band 7

Resolution = 20m/px

Central Wavelength = 783nm

Bandwidth = 20nm

Use: For classifying vegetation.

Band 8 (NIR)

Resolution = 10m/px

Central Wavelength = 842nm

Bandwidth = 115nm

Use: The near infrared band is good for mapping shorelines and biomass content, as well as at detecting and analyzing vegetation.

Band 8A

Resolution = 20m/px

Central Wavelength = 865nm

Bandwidth = 20nm

Use: For classifying vegetation.

Band 9

Resolution = 60m/px

Central Wavelength = 945nm

Bandwidth = 20nm

Use: It is good for detecting water vapour.

Band 10

Resolution = 60m/px

Central Wavelength = 1375nm

Bandwidth = 30nm

Use: For cirrus cloud detection.

Band 11 (SWIR 1)

Resolution = 20m/px

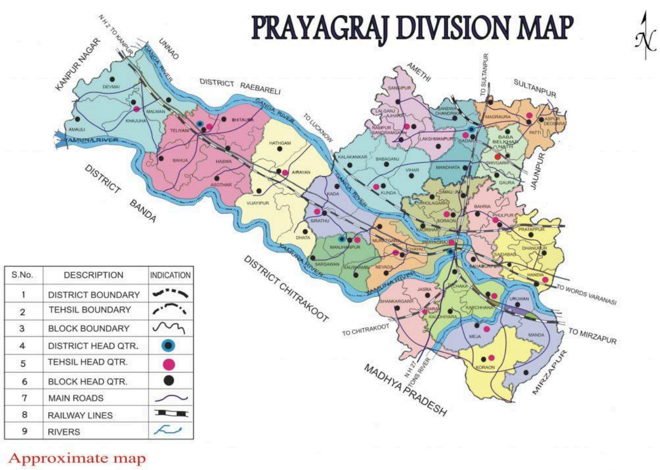
Central Wavelength = 1610nm

Bandwidth = 90nm

Use: It is useful for measuring the moisture content of soil and vegetation, and it provides good contrast between different types of vegetation. It helps differentiate between snow and clouds. On the other hand, it has limited cloud penetration.

Band 12 (SWIR 2)
Resolution = 20m/px
Central Wavelength = 2190nm
Bandwidth = 180nm
Use: It is useful for measuring the moisture content of soil and vegetation, and it provides good contrast between different types of vegetation. It helps differentiate between snow and clouds. On the other hand, it has limited cloud penetration.

Dataset details:



Prayagraj Division Map

We have used Sentinel-2 satellite data for 6 years (2019-2024) to analyze the changes happening in 4 classes of vegetation, built up, barren land and water. Data of month april-may for each year so that cloud cover is minimal during summer, and the weather condition is similar for each year.

Product uri	Processing level	Product type	Processing baseline	Generation time
S2B_MSIL2A_20190509T050659_N0500_R019_T44RNP_20221210T102053.SAFE	Level-2A	S2MSI2A	05.00	2022-12-10T10:20:53.000000Z
S2B_MSIL2A_20200403T050649_N0500_R019_T44RNP_20230612T173457.SAFE	Level-2A	S2MSI2A	05.00	2023-06-12T17:34:57.000000Z
S2B_MSIL2A_20210508T050649_N0500_R019_T44RNP_20230227T040224.SAFE	Level-2A	S2MSI2A	05.00	2023-02-27T04:02:24.000000Z
S2B_MSIL2A_20220513T050649_N0400_R019_T44RNP_20220513T081832.SAFE	Level-2A	S2MSI2A	04.00	2022-05-13T08:18:32.000000Z
S2A_MSIL2A_20230513T050651_N0509_R019_T44RNP_20230513T090759	Level-2A	S2MSI2A	05.09	2023-05-13T09:07:59.000000Z

S2B_MSIL2A_20240402T050649_N0510_R019_T44RNP_20240402T073359.SAFE	Level-2A	S2MSI2A	05.10	2024-04-02T07:33:59.000000Z

Ground Truth Data

We have collected data points from Several regions of IIITA. This dataset is used for reference purposes during classification and calculating the accuracy of the model..

S No.	Latitude	Longitutde	Latitude	Longitutde	location	landmark	class	subclass	photo
1	25.433039	81.770139	25° 25' 58.92" N	81° 46' 12.5" W	jhalwa	iiit gate 4	builtup	building	1.jpg
2	25.432832	81.770068	25° 25' 58.14" N	81° 46' 12.2" W	jhalwa	iiit gate 4	builtup	building	2.jpg
3	25.430588	81.77203	25° 25' 50.28" N	81° 46' 19.3" W	iiit allahabad	cc2	builtup	building	4.jpg
6	25.430534	81.77101	25° 25' 49.44" N	81° 46' 15.6" W	iiit allahabad	admin building	builtup	building	6.jpg
8	25.430913	81.770507	25° 25' 50.88" N	81° 46' 13.8" W	iiit allahabad	admin building	builtup	building	8.jpg
9	25.430478	81.770784	25° 25' 49.74" N	81° 46' 14.8" W	iiit allahabad	admin building	builtup	building	9.jpg
12	25.43039	81.769705	25° 25' 49.08" N	81° 46' 10.94" W	iiit allahabad	front of auditorium	builtup	road	12.jpg
14	25.430839	81.76927	25° 25' 51.04" N	81° 46' 9.37" W	iiit allahabad	auditorium	builtup	building	14.jpg
15	25.430586	81.769228	25° 25' 50.16" N	81° 46' 9.22" W	iiit allahabad	front of director house	builtup	road	15.jpg
16	25.431357	81.769408	25° 25' 52.92" N	81° 46' 9.85" W	iiit allahabad	left of auditorium	builtup	building	16.jpg
18	25.432171	81.770029	25° 25' 55.00" N	81° 46' 12.11" W	iiit allahabad	cc3	builtup	building	18.jpg
20	25.432802	81.769506	25° 25' 58.08" N	81° 46' 10.14" W	iiit allahabad	gate 4	builtup	road	20.jpg
21	25.429031	81.771462	25° 25' 44.12" N	81° 46' 17.77" W	iiit allahabad	swimming pool	builtup	building	21.jpg
23	25.429066	81.771974	25° 25' 44.64" N	81° 46' 19.79" W	iiit allahabad	tennis court	builtup	concrete surface	23.jpg
26	25.42814	81.773186	25° 25' 41.28" N	81° 46' 23.71" W	iiit allahabad	health care	builtup	building	26.jpg
28	25.428782	81.774252	25° 25' 43.32" N	81° 46' 27.31" W	iiit allahabad	behind main ground	builtup	building	28.jpg
29	25.428094	81.774584	25° 25' 41.84" N	81° 46' 28.5" W	iiit allahabad	professor house	builtup	building	29.jpg
30	25.427943	81.775184	25° 25' 40.36" N	81° 46' 30.50" W	iiit allahabad	i-block	builtup	building	30.jpg
32	25.42866	81.774945	25° 25' 43.16" N	81° 46' 29.82" W	iiit allahabad	building	builtup	building	32.jpg
33	25.429225	81.77512	25° 25' 45.68" N	81° 46' 30.48" W	iiit allahabad	building	builtup	building	33.jpg
34	25.429395	81.774875	25° 25' 45.92" N	81° 46' 29.54" W	iiit allahabad	building	builtup	building	34.jpg
38	25.430135	81.77493	25° 25' 48.20" N	81° 46' 29.74" W	iiit allahabad	vh2	builtup	building	38.jpg
35	25.429832	81.775562	25° 25' 47.64" N	81° 46' 32.03" W	iiit allahabad	visitor hostel 3	builtup	building	35.jpg
40	25.430572	81.775189	25° 25' 50.16" N	81° 46' 30.88" W	iiit allahabad	vh1	builtup	building	40.jpg
45	25.430285	81.773948	25° 25' 49.12" N	81° 46' 26.21" W	iiit allahabad	gh	builtup	building	45.jpg
42	25.430462	81.774792	25° 25' 49.72" N	81° 46' 29.25" W	iiit allahabad	front of vh1	builtup	road	42.jpg
47	25.430459	81.7731	25° 25' 49.84" N	81° 46' 23.15" W	iiit allahabad	water tank	builtup	building	47.jpg
53	25.431756	81.771477	25° 25' 54.32" N	81° 46' 17.31" W	iiit allahabad	library	builtup	building	53.jpg

Ref: Ground Truth Dataset

Normalized Difference Vegetation Index (NDVI):

The Normalized Difference Vegetation Index (NDVI) stands as a cornerstone in remote sensing, serving as the most widely used vegetation index for monitoring greenery on a global scale. Alongside NDVI, other commonly employed vegetation indices include the Enhanced Vegetation Index (EVI), Perpendicular Vegetation Index (PVI), and Ratio Vegetation Index (RVI). However, NDVI remains the preferred choice for its simplicity and effectiveness in capturing vegetation dynamics.

Healthy vegetation exhibits distinct spectral signatures owing to its unique biochemical composition. Chlorophyll, the primary pigment responsible for photosynthesis, absorbs radiation in the Blue (0.4 - 0.5 µm) and Red (0.6 - 0.7 µm) spectra while reflecting radiation in the Green (0.5 – 0.6 µm) spectrum. Consequently, healthy vegetation appears green to the human eye. Moreover, healthy plants possess a remarkable ability to reflect radiation in the Near Infrared (NIR) spectrum, typically ranging from 0.7 to 1.3 µm. This phenomenon is attributed to the internal structure of plant leaves, which contributes to the high reflectance observed in the NIR spectrum.

The formulation of NDVI capitalizes on these distinctive spectral properties of healthy vegetation, specifically leveraging the contrast between NIR and Red bands. The NDVI formula is elegantly simple:

NDVI = (NIR – Red) / (NIR + Red)

For sentinel-2 data:

The formula in terms of bands will be:

NDVI = (B8 - B4) / (B8 + B4)

This formula enables the calculation of NDVI values for each pixel in a remotely sensed image, providing insights into the spatial distribution and health of vegetation cover. High NDVI values typically indicate dense, healthy vegetation, while low NDVI values are associated with sparse vegetation or non-vegetated surfaces.

NDVI plays a pivotal role in various applications, including agriculture, forestry, environmental monitoring, and land use planning. In agriculture, NDVI is used to assess crop health, monitor vegetation vigor, and optimize agricultural management practices such as irrigation and fertilization. In forestry, NDVI helps in monitoring forest health, detecting deforestation, and assessing biomass and carbon stocks. Environmental monitoring applications of NDVI include monitoring land degradation, assessing habitat quality, and tracking changes in ecosystem dynamics. Furthermore, NDVI is valuable in urban planning for assessing green space distribution, monitoring urban vegetation health, and evaluating the effectiveness of urban greening initiatives.

While NDVI is a powerful tool for vegetation monitoring, it is not without limitations. NDVI values may saturate in areas with dense vegetation, leading to difficulty in distinguishing between different levels of vegetation density. Additionally, NDVI may be influenced by factors such as soil background, atmospheric conditions, and sensor characteristics, which need to be carefully accounted for in data interpretation.

In conclusion, NDVI stands as a versatile and indispensable tool for monitoring vegetation dynamics and assessing ecosystem health. Its simplicity, effectiveness, and wide applicability make it a valuable asset for a diverse range of environmental and agricultural applications, contributing to our understanding of global vegetation patterns and dynamics.

Normalized Difference Water Index (NDWI):

The Normalized Difference Water Index (NDWI) serves as a pivotal tool for analyzing water bodies using remote sensing imagery. This index exploits the spectral characteristics captured by the Green and Near Infrared (NIR) bands, which are sensitive to water features. By leveraging these bands, NDWI efficiently enhances water information in most cases, making it a valuable asset for water body analysis.
Water bodies exhibit distinct spectral properties, with low reflectance across most wavelengths except for the visible portion of the electromagnetic spectrum. Liquid water typically reflects more strongly in the Blue spectrum compared to the Green and Red spectra. Clear water, in particular, demonstrates the greatest reflectance in the Blue portion of the spectrum, resulting in its characteristic blue appearance. In contrast, turbid water exhibits higher reflectance across the visible spectrum. Notably, water bodies do not reflect in the Near Infrared (NIR) and beyond.
The NDWI was initially developed by Gao (1996) to enhance water-related features in landscapes. It utilizes the Near Infrared (NIR) and Short-Wave Infrared (SWIR) bands to differentiate water bodies from other land cover types. The formula for calculating NDWI is straightforward:

NDWI = (NIR – SWIR) / (NIR + SWIR)
For sentinel-2 data:
NDWI = (B8 - B12) / (B8 + B12)

However, the results obtained from the original NDWI formula may suffer from poor quality due to its sensitivity to build-up land, resulting in overestimation of water bodies. To address this limitation, Xu (2005) proposed a modified version of NDWI, known as the Modified Normalized Difference Water Index (MNDWI). This index utilizes the Green and SWIR bands, yielding improved performance in distinguishing water bodies from other land cover types.

MNDWI = (Green – SWIR) / (Green + SWIR)
For sentinel-2 data:
MNDWI = (B3 - B12) / (B3 + B12)

Similar to NDWI, the MNDWI values range between -1 and 1, with water bodies typically exhibiting values greater than 0.5. Vegetation features typically yield smaller values, facilitating their differentiation from water bodies. Built-up features typically yield positive values ranging from 0 to 0.2.
In summary, NDWI and MNDWI are valuable tools for water body analysis using remote sensing imagery. Their ability to efficiently enhance water information and differentiate water bodies from other land cover types makes them indispensable assets for various applications, including environmental monitoring, hydrological modeling, and land use planning.

Built Up Index :

Analysis of built-up areas is crucial for urban planning and monitoring urban expansion. Several indices, such as Built-up Index (BU), Urban Index (UI), Index-based Built-up Index (IBI), and Enhanced Built-up and Bareness Index (EBBI), have been developed for this purpose, each with its own formula and calculation method. These indices exploit the spectral characteristics of different land cover types to identify built-up areas effectively.

Built-up areas and bare soil exhibit higher reflectance in the Shortwave Infrared (SWIR) spectrum compared to the Near-Infrared (NIR). Conversely, water bodies do not reflect in the infrared spectrum. Additionally, green surfaces tend to have higher reflectance in the NIR spectrum than in the SWIR spectrum. Among these indices, the Built-up Index (BU) stands out as a robust tool for analyzing urban patterns.

BU is calculated using the Normalized Difference Built-up Index (NDBI) and the Normalized Difference Vegetation Index (NDVI). The BU index produces a binary image where higher positive values indicate built-up and barren areas. This allows BU to automatically map built-up areas with accuracy and efficiency.

BU = NDBI - NDVI

NDVI, on the other hand, measures vegetation density and health based on the difference between NIR and Red band reflectance.

NDBI = (SWIR – NIR) / (SWIR + NIR)

Unlike traditional image classification techniques, which involve complex operations and require meticulous analyst input, NDBI calculation is straightforward and easy to derive. Image classification techniques, such as supervised and unsupervised classification, involve the compilation of composite bands and the application of numerous operations to obtain the final result. The accuracy of these techniques is heavily reliant on the expertise of the image analyst and the methodological approach employed.

In contrast, NDBI calculation offers a simpler alternative for mapping built-up areas. By directly computing the NDBI using spectral bands, analysts can quickly derive accurate results without the need for extensive processing. This simplicity not only reduces the time and effort required but also minimizes the potential for subjective bias inherent in traditional classification methods.

Furthermore, the binary nature of BU output simplifies interpretation and facilitates automated mapping of built-up areas. Higher positive BU values directly indicate the presence of built-up and barren areas, allowing for rapid identification and analysis.

In conclusion, the Built-up Index (BU), derived from the Normalized Difference Built-up Index (NDBI) and Normalized Difference Vegetation Index (NDVI), offers a simple yet effective approach to analyzing built-up areas. Its straightforward calculation method and binary output make it a valuable tool for urban planning and monitoring, offering an efficient alternative to traditional image classification techniques.

Bareness Index :

The Normalized Difference Barren Index (NDBaI) is a critical tool in remote sensing for identifying and mapping barren or non-vegetated areas. It serves as a valuable index for assessing land cover dynamics, particularly in arid and semi-arid regions, where barren land cover types are prevalent. NDBaI exploits the spectral properties of remote sensing data to distinguish barren areas from other land cover classes, providing insights into land degradation, desertification processes, and ecosystem health.

The formulation of NDBaI is based on the contrast between the Near-Infrared (NIR) and Short-Wave Infrared (SWIR) bands, which exhibit distinct spectral responses for barren or non-vegetated surfaces. Unlike vegetated areas, barren surfaces typically have low reflectance in the NIR spectrum and relatively higher reflectance in the SWIR spectrum. This spectral behavior is attributed to the absence of chlorophyll and the presence of bare soil or rock surfaces, which tend to reflect more strongly in the SWIR region.

The formula for calculating NDBaI is expressed as follows:

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

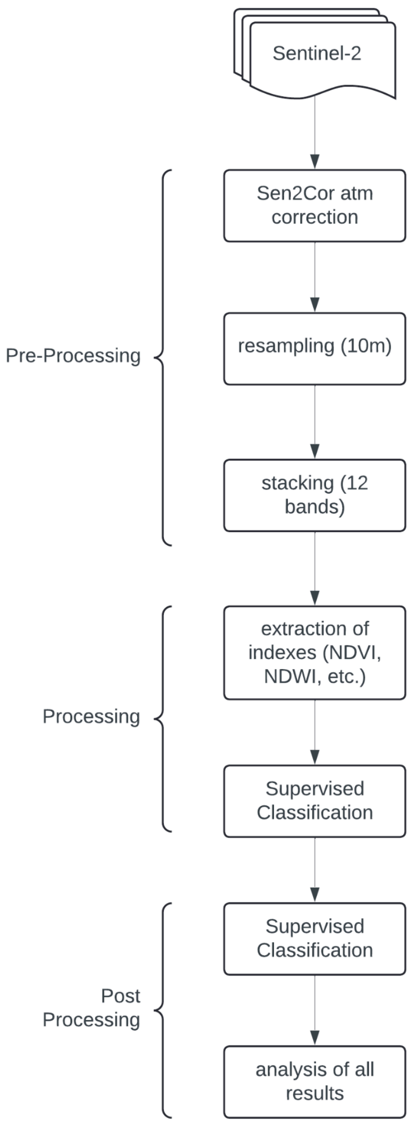
For Sentinel-2 data, the formula in terms of bands is:

$$NDBaI = (B8+B11)/(B8-B11)$$

NDBaI values typically range from -1 to 1, with higher positive values indicating barren or non-vegetated areas and lower or negative values representing vegetated areas. By analyzing the spatial distribution of NDBaI values across an area of interest, analysts can effectively delineate barren land cover types and monitor changes in barren areas over time. The interpretation of NDBaI values is straightforward: areas with higher positive NDBaI values are likely to be barren or non-vegetated, while areas with lower or negative values are more likely to be vegetated. This binary nature of NDBaI makes it a useful tool for mapping barren areas and distinguishing them from other land cover classes, such as vegetation, water bodies, and built-up areas.
NDBaI finds wide-ranging applications in environmental monitoring, land management, and natural resource management. In arid and semi-arid regions, where barren land cover types are prevalent, NDBaI can be used to assess the extent of desertification, monitor changes in land cover, and identify areas at risk of land degradation. Moreover, NDBaI can aid in the identification of potential sites for land restoration or reclamation efforts, helping to mitigate the impacts of land degradation and promote sustainable land management practices.
Furthermore, NDBaI can be integrated with other remote sensing indices, such as NDVI and NDBI, to provide a more comprehensive understanding of land cover dynamics. For example, combining NDBaI with NDVI allows for the differentiation between barren areas and vegetated areas, while integrating NDBaI with NDBI facilitates the identification of built-up areas and non-vegetated areas. Such multi-index approaches enable analysts to gain deeper insights into land cover patterns and processes, enhancing their ability to monitor and manage land resources effectively.

In conclusion, the Normalized Difference Barren Index (NDBaI) is a valuable tool for mapping barren or non-vegetated areas using remote sensing imagery. Its simplicity, effectiveness, and wide-ranging applications make it an indispensable asset for environmental monitoring, land management, and natural resource management. By leveraging the spectral properties of remote sensing data, NDBaI provides valuable insights into land cover dynamics, helping to inform decision-making and promote sustainable land management practices.

• III. METHODOLOGY



1. Data Acquisition:

Collected free Sentinel-2 satellite data with minimal cloud cover for the desired research(prayagraj) and clip it in required region of prayagraj, from **(25.7372, 81.12611)** to (**25.2302, 82.03166),** region and time frame **(2019-2024)**. Sentinel-2's 10m, 20m and 60m spatial resolution is appropriate for land cover classification.

2. Ground Truth Collection:

Obtain reliable ground truth data for the four designated land cover classes.

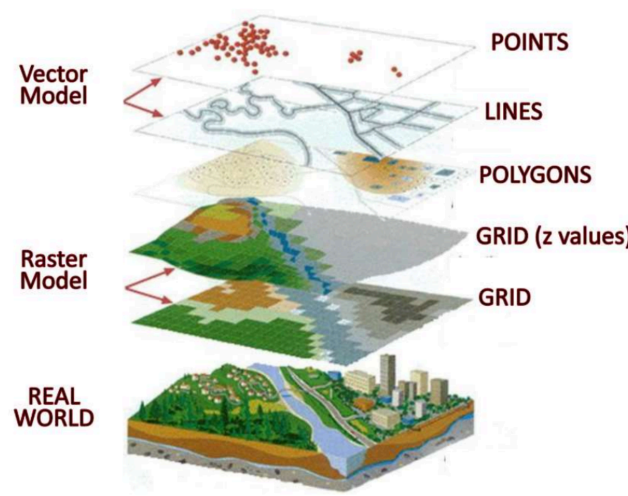
- First we did Survey at the IIITA campus, equipped with phone and GPS devices. The purpose of the survey was to collect ground truth points, representing different types of land cover found on the campus. Then we did a survey in the region around IIITA, in Kalindipuram, then we did a survey in Civil Lines, and Sangam region. This ground truth dataset will be used as a reference dataset.

3. Image Preprocessing:

Prepare the imagery by applying pre-processing operations to improve interpretation and analysis. This includes:

- Atmospheric Correction: Use well-established algorithms like Sen2Cor to correct for atmospheric effects.

4. Clipping: Methodology



Ref: [Data Visualization with GIS: An Introduction | PPT]

5. Classification Algorithm:

Select and apply classification algorithms to classify the pre-processed imagery.

- Algorithm Selection: Evaluate different algorithms like K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM), or neural networks to determine the optimal approach.
- Training: Split the reference dataset into training and validation sets. Train the chosen algorithm with the training data.
- Hyperparameter Tuning: Adjust algorithm parameters to optimize performance.

6. Model Evaluation:

Assess the accuracy and performance of the classification algorithms.

- Validation: Use the validation set to evaluate model performance.
- Accuracy Metrics: Calculate metrics like overall accuracy, misclassification rate and weighted f1-score.

7. Visualization:

Generate visual representations of the classification results.

- Thematic Maps: Create thematic maps for each land cover class, highlighting the spatial distribution.
- False Color Composites: Generate false-color composites to visualise spectral differences and aid interpretation.
- Change Detection Maps: Visualize change areas by overlaying classification results from different years.

8. Classification of Previous Years' Data:

- Repeat the image pre-processing, training data collection, and classification steps for each year in the time period of 2019-2024.

9. Analysis:

Analyze changes in land cover over time.

- Post-Classification Comparison: Compare the classification results year-on-year to identify areas of land cover change.
- Change Metrics: Calculate metrics like percent change, magnitude of change, and transition matrices to quantify the nature and extent of land cover changes.

IV. Results

Type	**2020**	**2021**	**2022**	**2023**	2024
Vegetation	36.150	42.937	45.498	29.911	34.942
Built Up	7.812	8.962	9.716	10.050	10.320
Barren Land	51.6626	45.448	42.823	54.811	52.893
Water	4.375	2.654	1.963	5.228	1.837

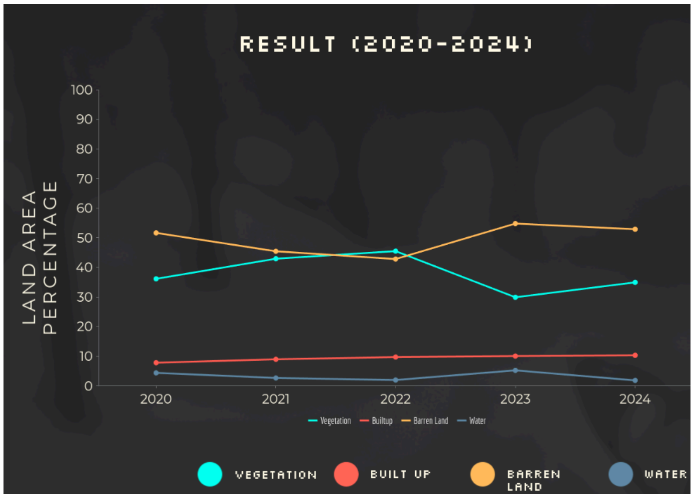
Confusion Matrix (2024):

TARGET / OUTPUT	Vegetation	Built Up	Barren Land	Water	SUM
Vegetation	108 22.50%	5 1.04%	7 1.46%	0 0.00%	120 90.00% 10.00%
Built Up	4 0.83%	102 21.25%	14 2.92%	0 0.00%	120 85.00% 15.00%
Barren Land	1 0.21%	3 0.63%	110 22.92%	6 1.25%	120 91.67% 8.33%
Water	1 0.21%	2 0.42%	4 0.83%	113 23.54%	120 94.17% 5.83%
SUM	114 94.74% 5.26%	112 91.07% 8.93%	135 81.48% 18.52%	119 94.96% 5.04%	433 / 480 90.21% 9.79%

Accuracy : 0.9021

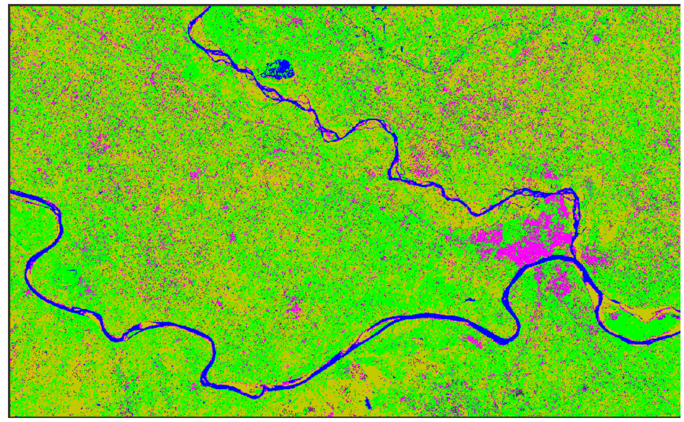
Misclassification Rate : 0.0979

Weighted-F1 : 0.9015



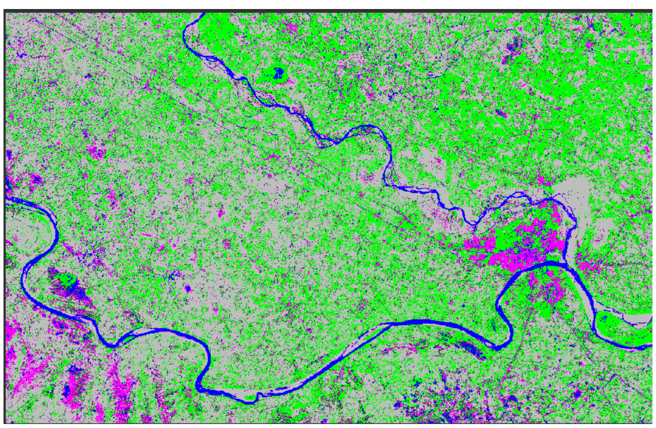
Trend in previous year data

2021



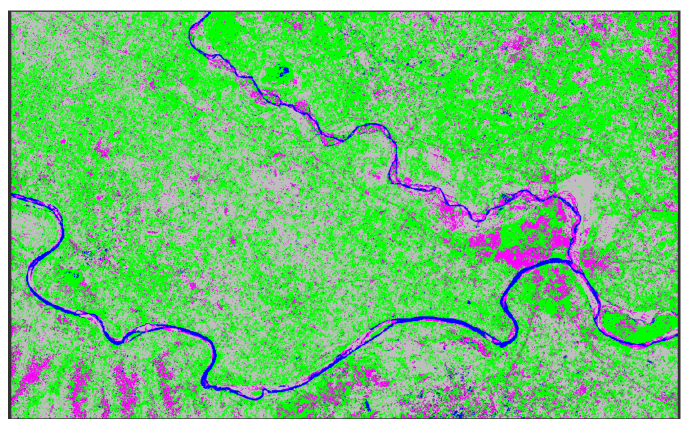
Label	Colour	Value	Frequency	Description
no data		-1	0.000%	no data
Barren land		0	45.448%	
Built-Up Area		1	8.962%	
Vegetation		2	42.937%	
Water		3	2.654%	

2020



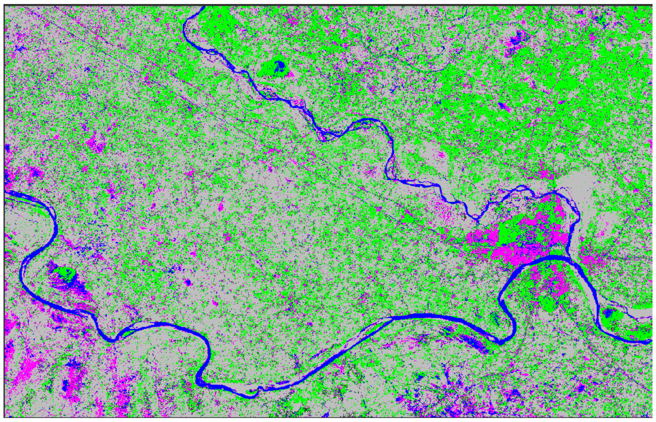
Label	Colour	Value	Frequency	Description
no data		-1	0.000%	no data
Barren Land		0	51.662%	
Built-Up Area		1	7.812%	
Vegetation		2	36.150%	
Water		3	4.375%	

2022



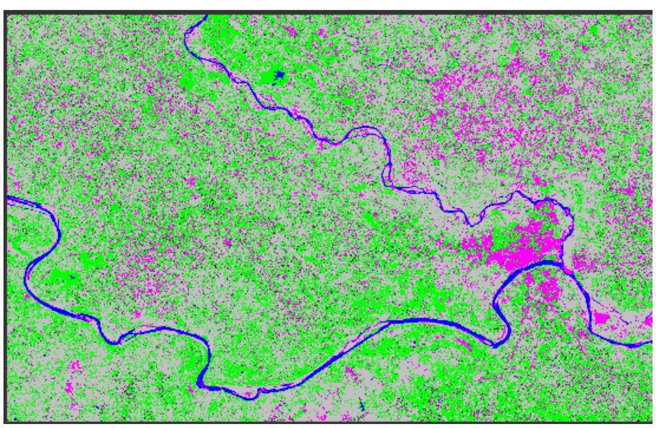
Label	Colour	Value	Frequency	Description
no data		-1	0.000%	no data
Barren Land		0	42.823%	
Vegetation		1	45.498%	
Water		2	1.963%	
Built-Up Area		3	9.716%	

2023



Label	Colour	Value	Frequency	Description
no data		-1	0.000%	no data
Barren Land		0	54.811%	
Built-Up Area		1	10.050%	
Vegetation		2	29.911%	
Water		3	5.228%	

2024



Label	Colour	Value	Frequency	Description
no data		-1	0.000%	no data
Barren Land		0	53.893%	
Built-Up Area		1	11.327%	
Vegetation		2	32.942%	
Water		3	1.837%	

V. Conclusions

In conclusion, this study demonstrates the efficacy of Sentinel-2 satellite imagery combined with machine learning algorithms for precise land classification in Prayagraj. By employing sophisticated remote sensing techniques and advanced algorithms such as Support Vector Machine and Random Forest, the research successfully delineated land cover types into four distinct categories: urban areas, agricultural land, water bodies, and vegetation cover.

The results underscore the significance of leveraging spectral information captured by Sentinel-2 to identify and classify diverse land cover patterns accurately. This capability holds immense potential for various sectors, including urban planning, environmental monitoring, and sustainable development initiatives in Prayagraj.

The implications of this study are far-reaching, offering valuable insights for policymakers and urban planners. Accurate land classification facilitates informed decision-making regarding land use management, infrastructure development, and natural resource conservation. Furthermore, it enables the assessment of urbanization's impact on the environment and aids in identifying areas vulnerable to environmental degradation.

Overall, the findings contribute to advancing our understanding of land dynamics in Prayagraj and provide a robust framework for future research and decision support systems aimed at fostering sustainable development and environmental stewardship in the region.

References

1. M. N. Ahmad, Z. Shao, A. Javed (2023)"Mapping Impervious Surface Area Increase and Urban Pluvial Flooding Using Sentinel Application Platform (SNAP) and Remote Sensing Data."Environ Sci Pollut Res 30, 125741–125758 **Paper Link**

2. M. Cavur, H. S. Duzgun, S. Kemec, D. C. Demirkan,(2019) "Land Use and Land Cover Classification of Sentinel-2A: St Petersburg Case Study."International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-1/W2, 13–16 **Paper Link**

3. M. Majidi Nezhad, A. Heydari, L. Fusilli, G. Laneve,(2019) "Land Cover Classification by Using Sentinel-2 Images: A Case Study in the City of Rome."In Proceedings of the 4th World Congress on Civil, Structural, and Environmental Engineering (ICEPTP19) (pp. 1–April). Environmental Science. DOI: 10.11159/ICEPTP19.158 **Paper Link**

4. A. Sekertekin, A. M. Marangoz, H. Akcin, (2017)"Pixel-Based Classification Analysis of Land Use Land

Cover Using Sentinel-2 and Landsat-8 Data." . International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, (pp. 91-93) XLII-4/W6 **Paper Link**

5. A. N. M. Al-Rubkhi,(2017) "Land Use Change Analysis and Modeling Using Open Source (QGIS).Case Study: Boasher Willayat" Paper Link

6.Y. Zha, J. Gao & S. Ni (2003) Use of normalized difference built-up index in automatically mapping urban areas from TM imagery, International Journal of Remote Sensing, 24:3, 583-594, DOI: 10.1080/01431160304987

7.Zhao, H. – Chen, X. (2005): Use of normalized difference bareness index in quickly mapping bare areas from TM/ETM+. Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. IEEE International. 3: 1666-1668.

8.Gao. "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space." 1996.

9.Crippen, R.E. (1990). Calculating the vegetation index faster: Remote sensing of Envirommnent, 34, 71-73.

10. Level-2A Algorithm Theoretical Basis Document

11. Sentinel-2 Bands

12. Niculescu, S., Billey, A., & Talab-Ouali, H. (Year). Random Forest Classification using Sentinel-1 and Sentinel-2 series for vegetation monitoring in the Pays de Brest (France). Remote Sensing for Agriculture, Ecosystems, and Hydrology XX, Christopher M. U. Neale & Antonino Maltese (Eds.), Proc. of SPIE Vol. 10783 (pp. 1078305-1). SPIE. doi:10.1117/12.2325546
Paper Link