Indian Institute of Information Technology Allahabad

Prayagraj (UP) India

Geospatial Data Analysis of Prayagraj

Group Members:

Shubham Kumar (IIT2020007)

Anoop Kumar (IIT2020020)

Shashikant Thakur (IIT2020024)

Supervised By :

Dr. Triloki Pant

Index

1. Topic
2. Literature Review
3. Dataset Description
4. Methodology
5. Result & Discussion
6. References

Timeline

C1
Literature Review
Reading Research Paper

C1
Making Methodology
Finalising methodology for the task

C1
Ground Truth Collection
Collecting data by visiting different sites

C1-C2
Training and evaluation of models
Training the models on raster

C1-C2
Analysis of 2020 to 2024 data
Processing on previous years data and performing change detection

C2-C3
Enlarging ROI radius
Performing previous steps on larger ROI

TOPIC

- **Geospatial data analysis**
  - It involves interpreting data containing spatial or geographic elements using GIS and analytical techniques.
- **Change detection in prayagraj region**, center point taken as IIITA, in the previous five years(2020 - 2024).

Objective of the project

- Objective of our project is to classify the land area of the city of Prayagraj into following four classes: Urban area, Barren land area, Vegetation area, Water bodies
- Classification will be done with algorithms random forest. On the best performing algorithm on current data, classification of previous 5 year data will be done.
- Perform change detection for previous five year(2020 - 2024).

Literature Review

LAND USE AND LAND COVER CLASSIFICATION OF SENTINEL 2-A: ST PETERSBURG CASE STUDY

- Objective: Classify St. Petersburg's land use and cover with Sentinel-2A imagery and SVM.
- Methodology: Use ESA data, SNAP preprocessing, and SVM for classification,ArcGIS Desktop Software used.
- Findings: Achieve 83.64% accuracy, proving Sentinel-2A and SVM's effectiveness.
- Significance: Highlights Sentinel-2A's value for precise urban mapping.
- Recommendations: Further refine algorithms and data integration for improved mapping.

Literature Review

Classification by using Sentinel-2 Images: A case study in the city of Rome

- Objective: Improve land cover classification accuracy in Rome using Sentinel-2 images.
- Methodology: Employ MLE and SVM algorithms to classify urban, forest, water, agriculture, and empty land.
- Study Area: Rome, with diverse terrain and distinctive building distribution.
- Results: Accuracy :
    - MLE : 92.7%
    - SVM : 85.25%
- Implications: Accurate land cover mapping crucial for environmental planning and management.

Literature Review

[Mapping impervious surface area increase and urban pluvial flooding using Sentinel Application Platform (SNAP) and remote sensing data]

- Objective: To map impervious surface area growth in major Pakistani cities and assess its impact on urban pluvial flooding.

- Methodology: Utilized GISAI on Google Earth Engine for impervious surface mapping and Sentinel-1 GRD data for urban flood extent detection.

- Findings: Significant increases in impervious surface area were observed in Islamabad, Lahore, and Karachi, correlating with heightened urban pluvial flooding occurrences.

- Significance: Provides crucial insights into urbanization dynamics and flood vulnerability, aiding urban planning and flood risk management in Pakistani cities.

- Results: Overall Accuracy : Lahore: 0.93 Karachi: 0.86 Islamabad: 0.85

- Recommendations: Implement modeling-based solutions for identifying high-risk flood areas and continuously monitor impervious surface growth using remote sensing for enhanced urban resilience.

Literature Review

PIXEL-BASED CLASSIFICATION ANALYSIS OF LAND USE LAND COVER USING SENTINEL-2 AND LANDSAT-8 DATA

- Objective: Compare accuracy of LULC classifications from Sentinel-2 and Landsat-8 in Zonguldak
- Methodology: Utilized MLC supervised classification with common bands, assessed accuracy using kappa statistics
- Findings: Sentinel-2 generally superior, Landsat-8 better in some Sea areas
- Significance: Crucial for sustainable land management and urban planning
- Results: Sentinel-2: 88.74% accuracy, kappa 0.85; Landsat-8: 83.91% accuracy, kappa 0.78
- Recommendations: Explore alternative classification methods, investigate temporal changes using multi-temporal imagery

Literature Review

[Land-Use-Change-Analysis-and-Modeling-Using-Open-Source-QGIS]

- Objective: Analyze land use changes in Boasher Willayat ,Oman (2000-2010) using QGIS.
- Methodology: Used ANN and Logistic Regression for Classification of Land Area.
- Findings: Urbanization increased while agriculture decreased, with further urban expansion forecasted.
- Significance: Demonstrated MOLUSCE's effectiveness in monitoring land use changes.
- Result : Accuracy of ANN: 88.96% , Accuracy of Logistic Regression: 92.79%
- Recommendations: Advocate for urban planning to manage rapid urbanization, emphasizing governmental intervention and open-source GIS tools' relevance.

Literature Review

Random Forest Classification using Sentinel-1 and Sentinel-2 series for vegetation monitoring in the Pays de Brest (France)

Objective: Assess the Random Forest algorithm efficacy in vegetation mapping using Sentinel and SPOT satellite data in the Pays de Brest area.

Methodology: Utilized Sentinel-1, Sentinel-2, and SPOT-6 satellite imagery from 2017; conducted preprocessing and calculated vegetation indices; implemented supervised Random Forest classification with time series stacking.

Findings: Achieved high accuracy in delineating vegetation classes, including summer and winter crops, forests, water bodies, grasslands, and sand dunes.

Significance: Highlights the potential of Random Forest and multi-source satellite data for accurate vegetation monitoring, crucial for environmental management and agricultural planning.

Recommendations: Further exploration of Random Forest application in different regions, incorporation of recent satellite data, and collaboration with stakeholders for land use planning and conservation strategies.

Result : Overall Accuracy: 0.96

Dataset Description

Sentinel-2 data has been used in our project

Sentinel-2 is equipped with an optical instrument payload that samples 13 spectral bands, including:
four bands at 10 m
six bands at 20 m
three bands at 60 m spatial resolution

Ref - https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2

Bands in sentinel-2

ROI(IIITA)

Band	Resolution	Central Wavelength	Description
B1	60 m	443 nm	Ultra Blue (Coastal and Aerosol)
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B8a	20 m	865 nm	Visible and Near Infrared (VNIR)
B9	60 m	940 nm	Short Wave Infrared (SWIR)
B10	60 m	1375 nm	Short Wave Infrared (SWIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)

Collection of ground truth data of IIITA (60 points) [6]

We have collected 250 ground truth locations from region of IIITA, near IIITA, Kalindipuram, Civil Lines, Sangam region

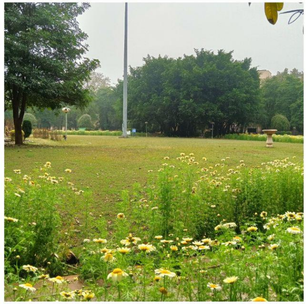
S No.	longitude	latitude	location	landmark	class	subclass	photo	remarks
1	25.433039	81.770139	jhalwa	iiit gate 4	builtup	building	1.jpg	
2	25.432832	81.770068	jhalwa	iiit gate 4	builtup	building	2.jpg	
3	25.433039	81.77203	iiit allahabad	cc2	builtup	building	3.jpg	
4	25.433039	81.771668	iiit allahabad	front of cc2	vegetation	grass	4.jpg	
5	25.433039	81.771677	iiit allahabad	left of cc2	vegetation	trees	5.jpg	
6	25.433039	81.77101	iiit allahabad	admin building	builtup	building	6.jpg	
7	25.433039	81.771131	iiit allahabad	admin building	vegetation	grass	7.jpg	
8	25.433039	81.770507	iiit allahabad	admin building	builtup	building	8.jpg	
9	25.433039	81.770784	iiit allahabad	admin building	builtup	building	9.jpg	
10	25.433039	81.770413	iiit allahabad	rock garden	vegetation	grass	10.jpg	
11	25.433039	81.769744	iiit allahabad	front of auditoriu	vegetation	trees	11.jpg	
12	25.433039	81.769705	iiit allahabad	front of auditoriu	builtup	road	12.jpg	
13	25.433039	81.769621	iiit allahabad	in front of directc	vegetation	trees	13.jpg	
14	25.433039	81.76927	iiit allahabad	auditorium	builtup	building	14.jpg	
15	25.433039	81.769228	iiit allahabad	front of director h	builtup	road	15.jpg	
16	25.433039	81.769408	iiit allahabad	left of auditorium	builtup	building	16.jpg	
17	25.433039	81.769463	iiit allahabad	left of auditorium	vegetation	trees	17.jpg	
18	25.433039	81.770029	iiit allahabad	cc3	builtup	building	18.jpg	
19	25.433039	81.769761	iiit allahabad	behind cc3	vegetation	trees	19.jpg	
20	25.433039	81.769506	iiit allahabad	gate 4	builtup	road	20.jpg	
21	25.433039	81.771462	iiit allahabad	swimming pool	builtup	building	21.jpg	
22	25.433039	81.771682	iiit allahabad	clock tower	vegetation	grass	22.jpg	

Sheet1









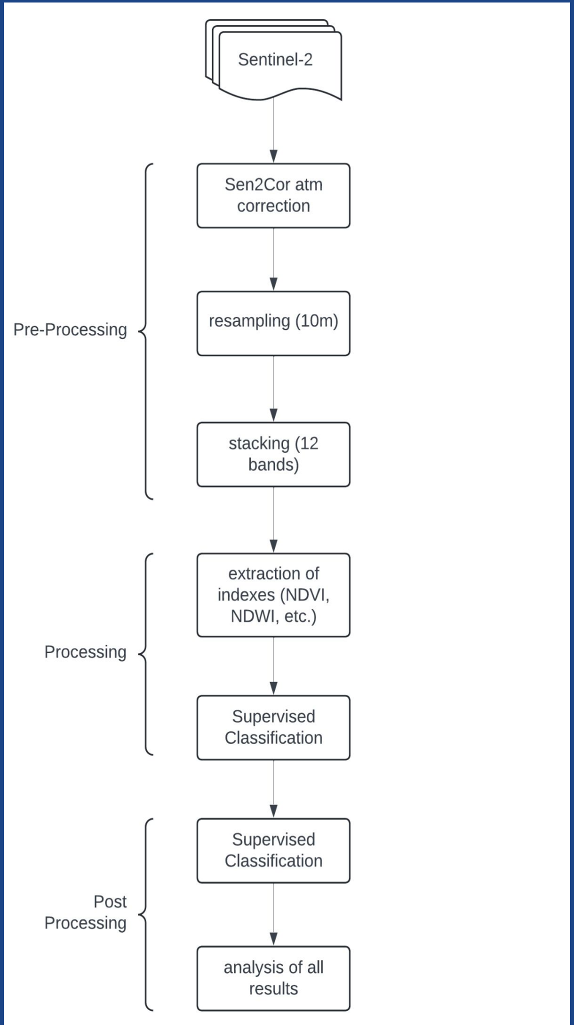









Flow Diagram



Methodology

- Data Acquisition: Collection of Sentinel-2 satellite data with minimum could cover.
- Image Preprocessing:
  - Perform image preprocessing tasks such as atmospheric corrections to enhance the quality of the imagery.
- Image Clipping: Clip the required ROI raster by overlaying vector over raster.
- Training Data Collection:
  - Collect ground truth data representing the four land cover classes (Barren Land, vegetation, built-up, water) through field surveys or existing land cover datasets.
- Classification Algorithm:
  - Split the reference dataset into training and validation sets to train and assess the performance of the classification algorithm.
  - Classification using Random Forest algorithm.
- Model evaluation : Evaluate classification by matrix like accuracy and f1-score.
- Visualization: Visualize the classified map for interpretation.
- Classification of Previous years data: Repeat above steps for previous 5 years data(2019-2023).
- Analysis of results for year on 2020 to 2024

Confusion Matrix

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%

(2024)

Results and Analysis

Accuracy : 0.9021

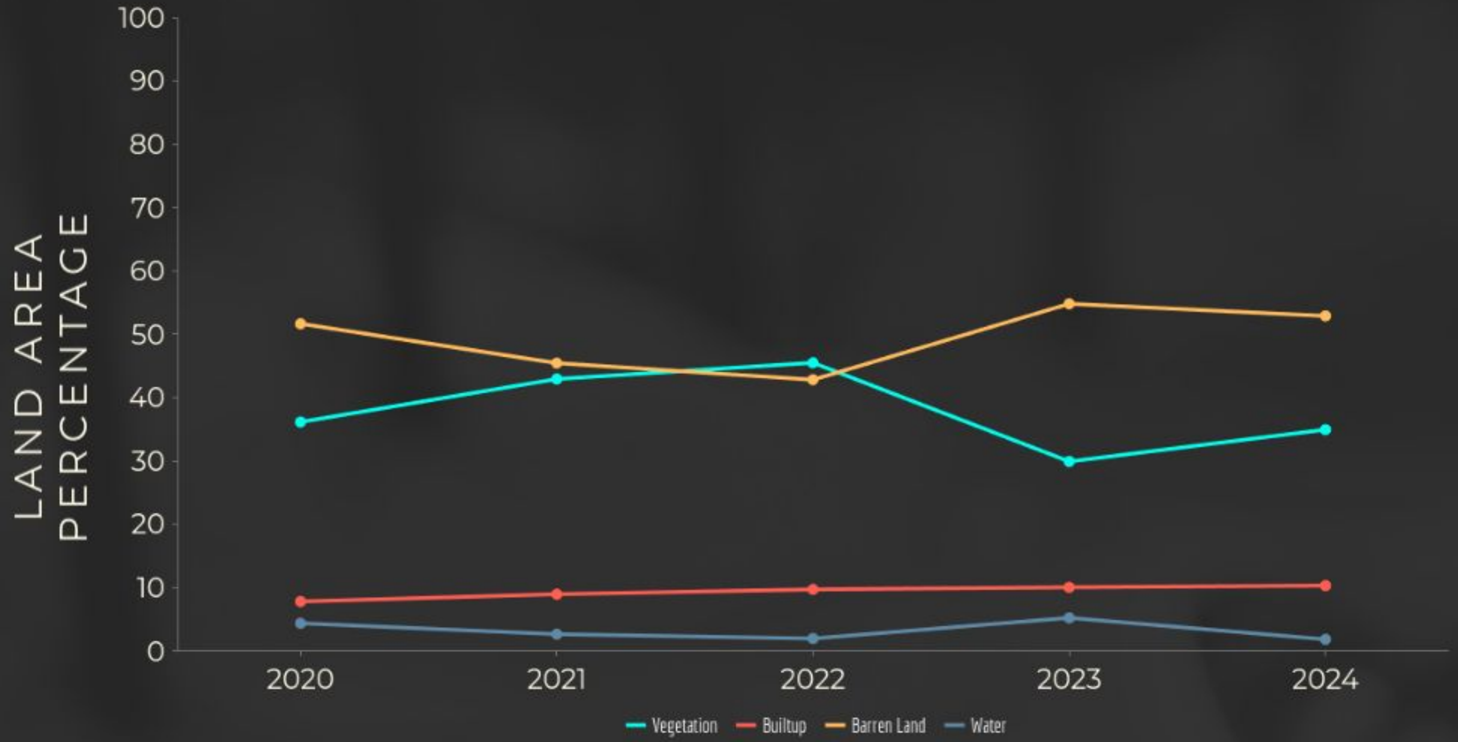
Misclassification Rate : 0.0979

Weighted-F1 : 0.9015

Results and Analysis

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

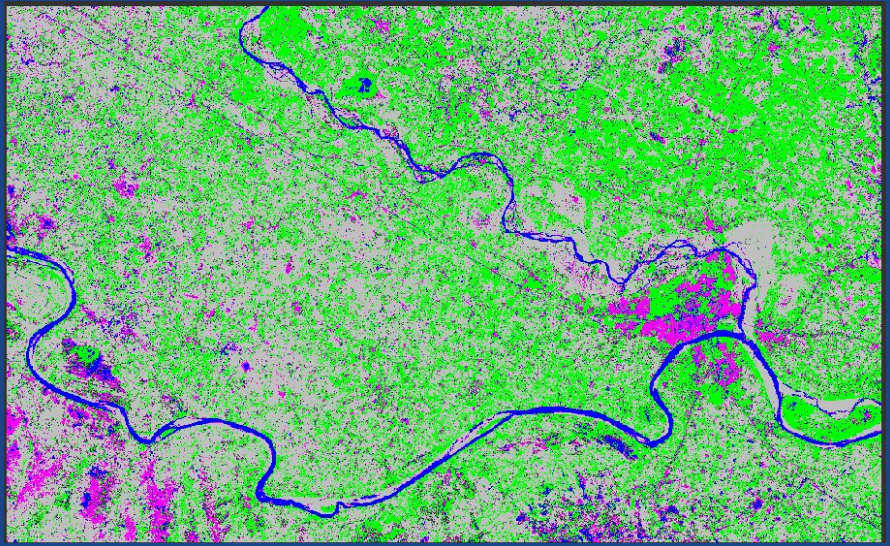
RESULT (2020-2024)



VEGETATION BUILT UP BARREN LAND WATER

Results and Analysis

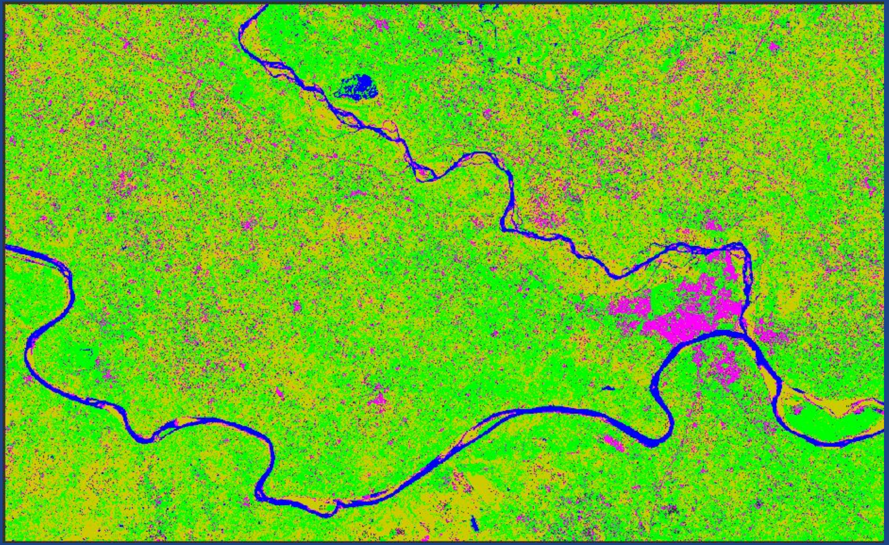
2020



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%	

Results and Analysis

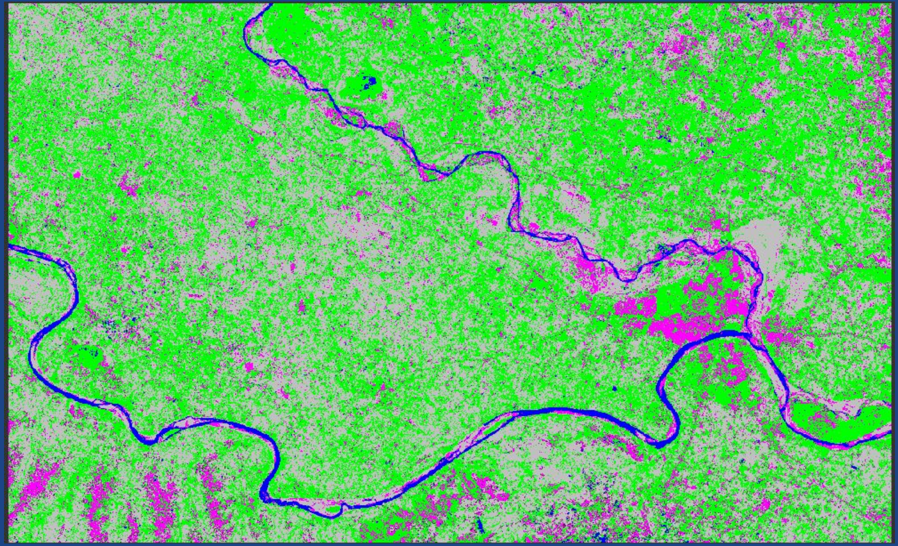
2021



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%	

Results and Analysis

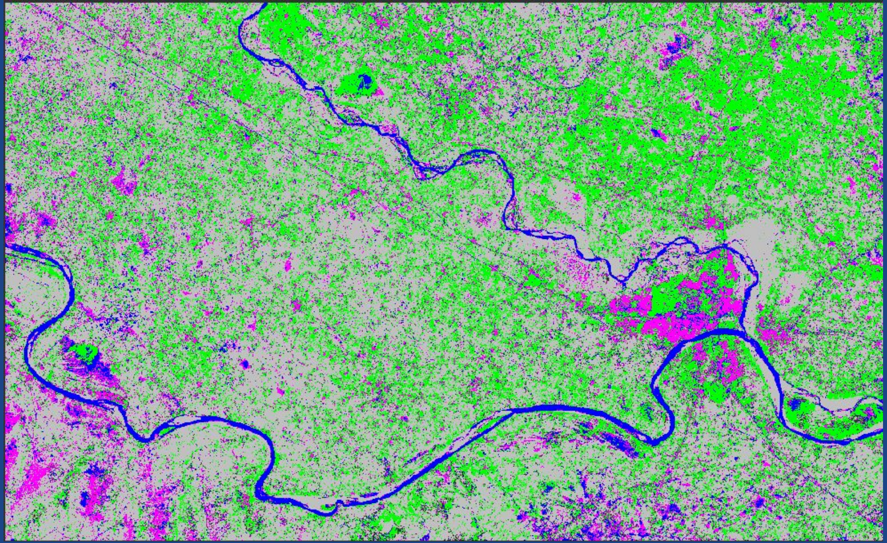
2022



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%	

Results and Analysis

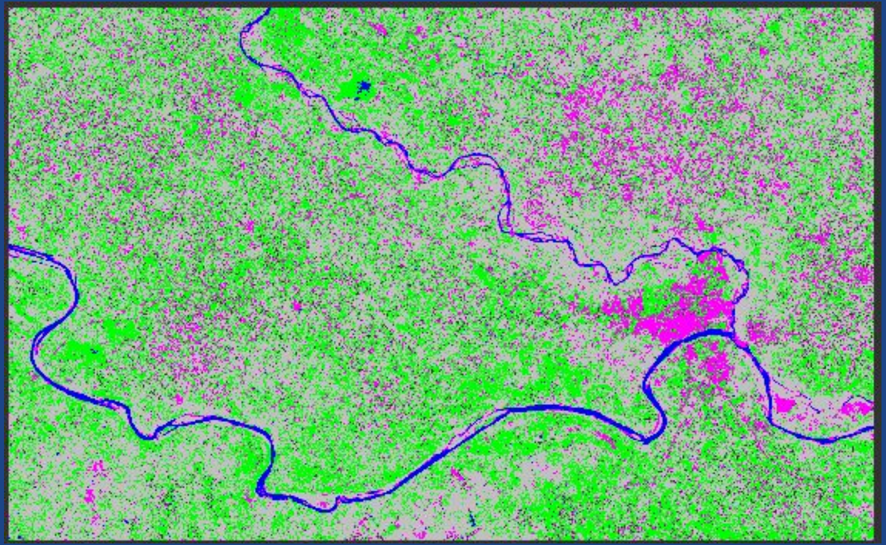
2023



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%	

Results and Analysis

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%	

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

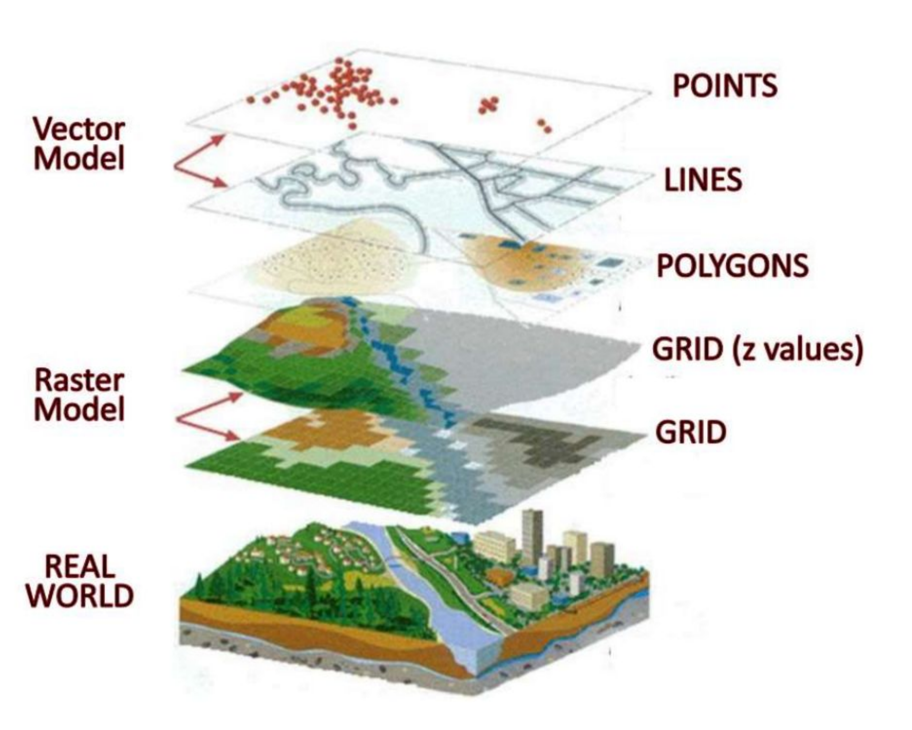
References

[06] **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

[7] Atmospheric / Topographic Correction for Satellite Imagery
[8] Atmospheric Correction in Sentinel-2
[9] Kappa Coefficient
[10] Vector Image
[11] Change Detection Algorithm

Thank You

Clipping: Methodology



[9]

Algorithm for change detection

(1) The changing image for both the images were calculated by the following Equation (3),

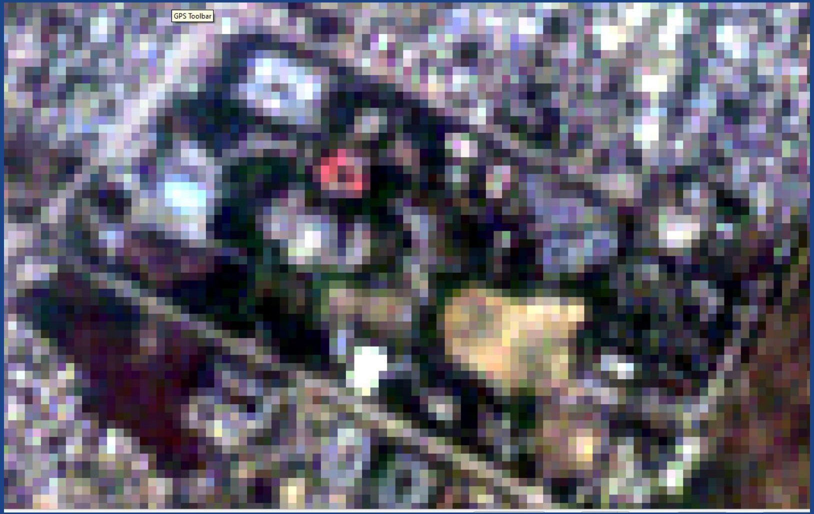
$$\Delta BV_{ijk} = BV_{ijk}\,(Img1) - BV_{ijk}\,(Img2) \tag{3}$$

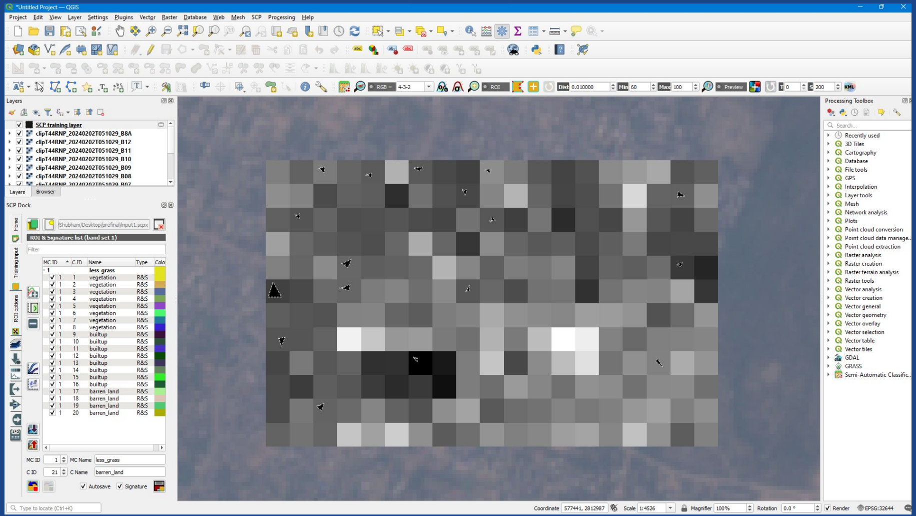
Change detection is useful n areas with natural calamities like forest fire, floods, etc. Since Allahabad is also a flood prone areas we want to analyse the changes that are happening in land area of Prayagraj. This will give us an insight about how the changes are happening in Prayagraj.

Semi-Automatic Classification Plugin (SCP)

1. SCP is a plugin for QGIS designed for semi-automatic image classification.
2. It simplifies the process of remote sensing data analysis and classification tasks.
3. SCP supports various classification algorithms and provides tools for accuracy assessment.
4. Users can follow a step-by-step workflow for loading, preprocessing, classifying, and evaluating results.
5. SCP enhances the efficiency and accuracy of image classification within the QGIS environment.

Progress………………………………………………………………………………


GPS Toolbar


*Untitled Project — QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh SCP Processing Help
Layers
SCP training layer
SCP Dock
ROI & Signature list (band set 1)
Processing Toolbox

Cohen Kappa Statistic

Measure The Performance of Classification Models

Kappa Statistic

$$k = \frac{2 * (TP * TN - FN * FP)}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)}$$

Assess the level of agreement between an actual and predicted

Predicted (rater 2) \ Actual (rater 1)	YES	NO	
YES	45 (**TP**)	15 (**FN**)	60
NO	25 (**FP**)	15 (**TN**)	40
	70	30	

Kappa Score Interpretation

Kappa	**Agreement**
<0	Less than chance agreement
0.01-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-0.99	Almost perfect agreement

$$k = \frac{2 * (45 * 15 - 15 * 25)}{(45 + 25) * (25 + 15) + (45 + 15) * (15 + 15)} = 0.13 \ (13\%)$$

Kappa Coefficient

Cohen's Kappa and F1 score serve different purposes in the evaluation of machine learning models, and each has its own advantages based on the specific requirements of a task. Here are some advantages of Cohen's Kappa over F1 score and vice versa:

Advantages of Cohen's Kappa:

Handles Class Imbalance:

- Kappa is less sensitive to class imbalance than F1 score. In situations where one class significantly outnumbers the others, Kappa can provide a more balanced measure of agreement.

Accounts for Chance Agreement:

- Kappa corrects for the agreement that could occur by chance alone. This makes it particularly useful when there is a random element in the classification, or when there might be agreement by luck.

Interpretability:

- Kappa is more interpretable as it directly measures agreement in terms of chance-corrected concordance. This can be useful in scenarios where the goal is to have reliable and consistent predictions across different raters or classifiers.

Applicability to Multiclass Problems:

- Cohen's Kappa can be extended to handle multi-class classification problems, providing a measure of overall agreement that considers chance agreement for multiple classes.

ALGORITHMS