India Space Academy, Department of Space Education

Summer Training Programme on Remote Sensing and GIS – 2025

Project title: Generating Land Use Land Cover (LULC) Map using Supervised Classification Techniques.

Project Code: P3

Prepared By:

Satwik Shreshth

MCA 2nd year student

Sikkim University

Date of Submission: 16s-Aug-2025

Table of Contents

Content	*Page No.*

***Title:* Generating Land Use Land Cover Map using Supervised Classification Techniques.**

Objective:

The goal of this project is to create a detailed and accurate Land Use and Land Cover (LULC) map for East Sikkim District, with a concentration on the Gangtok region, utilizing supervised classification algorithms and Sentinel-2 satellite images. The resulting LULC map is designed to help with environmental monitoring, urban planning, and sustainable land management in the region.

Study Area:

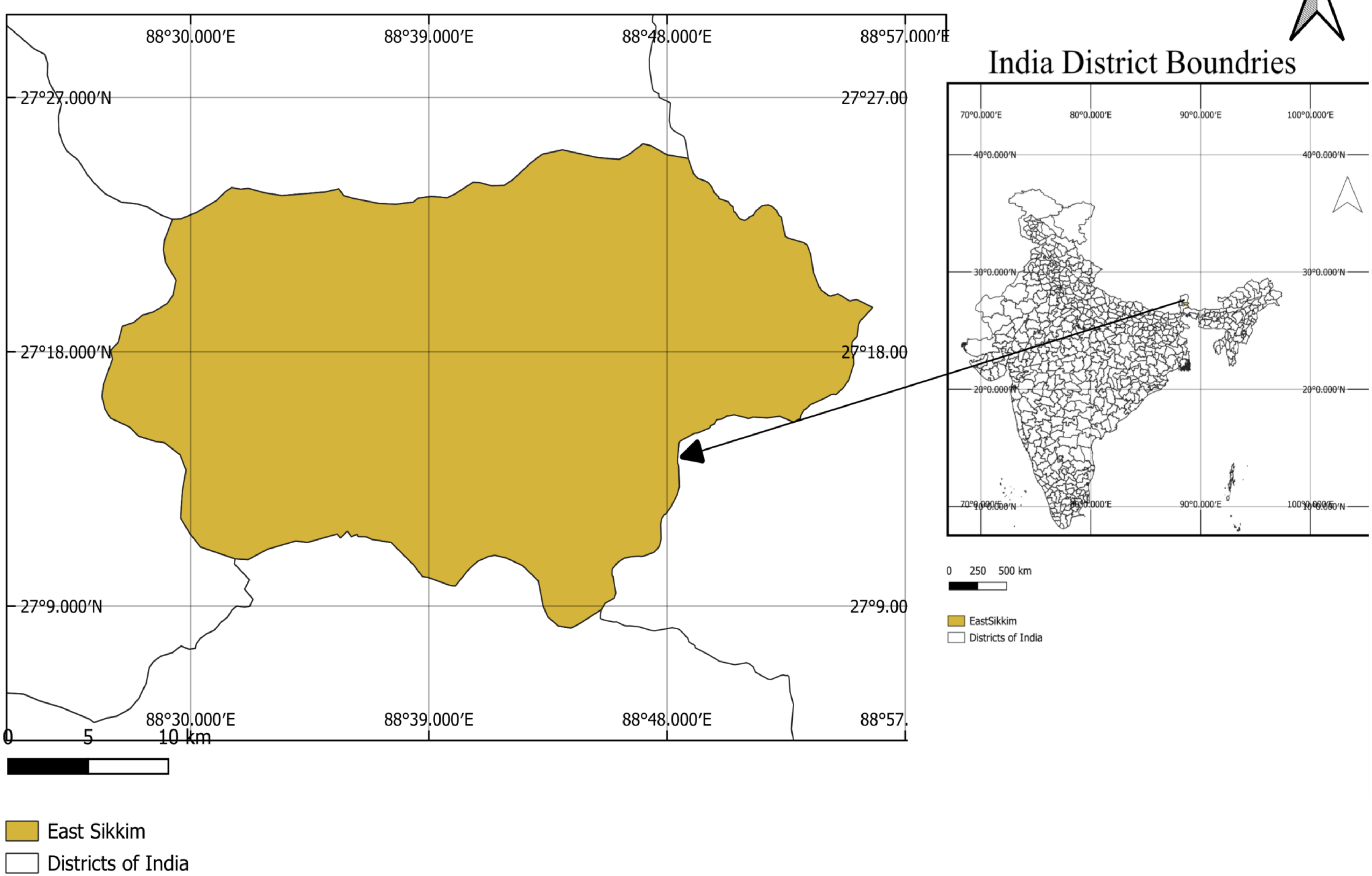
This project focuses on East Sikkim, a district situated in the northeastern Indian state of Sikkim. The main area of interest is around Gangtok, the capital city of the state, which features a varied landscape of urban areas, forests, agricultural land, and hilly terrain. This diversity makes it well-suited for Land Use and Land Cover (LULC) classification.

A map showing the Area of Interest (AOI), created in QGIS, is included below to provide spatial context for the selected study area.

Location Details:

- **AOI Name:** East Sikkim – Gangtok Region
- **Coordinates (center point):** *Latitude 27.3333° N, Longitude 88.6167° E*
- **District:** East Sikkim
- **State:** Sikkim
- **Country:** India

East Sikkim



Source: Map generated by Satwik Shreshth using QGIS. Administrative boundary of East Sikkim sourced from Datameet India Administrative Boundaries (GitHub: github.com/datameet/maps).

Data Used:

In this study, satellite images from the Sentinel-2 mission were used to perform land use and land cover classification for East Sikkim. The images were taken from the COPERNICUS/S2_SR_HARMONIZED dataset, which is available through Google Earth Engine. The dataset can be found at:

https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED

A time period from January 1, 2020 to January 31, 2024 was selected. This long date range helped in increasing the number of available scenes for the study area, especially because cloud-free images are often difficult to find in hilly regions like East Sikkim. To remove poor-quality scenes, a cloud cover filter of less than 80 percent was applied.

For this classification, six spectral bands were selected from Sentinel-2:

- **Band 2** (Blue, 490 nm)
- **Band 3** (Green, 560 nm)
- **Band 4** (Red, 665 nm)
- **Band 8** (Near Infrared or NIR, 842 nm)
- **Band 11** (Short Wave Infrared 1 or SWIR1, 1610 nm)
- **Band 12** (Short Wave Infrared 2 or SWIR2, 2190 nm)

These bands were chosen because they are useful in identifying various land cover types. For example, the red and near-infrared bands are helpful for vegetation studies, while the shortwave infrared bands are good for detecting soil, built-up areas, and snow.

In addition to these bands, three vegetation and water indices were calculated to improve classification accuracy. NDVI (Normalized Difference Vegetation Index) was used to highlight healthy vegetation by comparing the NIR and red bands. NDWI (Normalized Difference Water Index) was calculated using green and NIR bands to detect water bodies more accurately. NDSI (Normalized Difference Snow Index) was used to identify snow and bare rock using the green and SWIR1 bands.

All the images and calculated indices were clipped to study area. This helped in focusing the study on the East Sikkim region while making sure that all major land cover types were included within the analysis.

Methodology:

This section presents the detailed steps undertaken to perform Land Use Land Cover (LULC) classification for East Sikkim using Supervised Classification techniques in Google Earth Engine (GEE). The classification was executed using Sentinel-2 satellite imagery and a Random Forest classifier, structured into the following components:

Study Area and AOI Selection:

The study area selected for this research is **East Sikkim**, located in the eastern Himalayan region of India. The boundary of East Sikkim was extracted from a district-level administrative shapefile using the filter function in GEE. This shapefile was imported into the GEE environment to define the **Area of Interest (AOI)**. Once filtered, the map view was centered on East Sikkim to ensure that all operations and visualizations were focused on the selected district.

Satellite Image Collection and Preprocessing:

For image classification, **Sentinel-2 Surface Reflectance imagery** was used, particularly the **COPERNICUS/S2_SR_HARMONIZED** dataset. The satellite imagery was filtered temporally to select cloud-free images within a defined date range. Cloud masking was performed using the `QA60` band to exclude cloud-affected pixels.

Subsequently, a **median composite image** was created from the filtered collection to ensure a single, representative image free from noise and cloud shadows. The bands used in the analysis included: B2 (Blue), B3 (Green), B4 (Red), B8 (NIR), B11 (SWIR1), and B12 (SWIR2). This composite image was clipped to the East Sikkim boundary, preparing it for classification.

Training and Validation Data Preparation:

For supervised classification, manually digitized **training points** were created and labeled according to five distinct LULC classes:

- **0 – Water Bodies**
- **1 – Built-up Area**
- **2 – Barren Land**
- **3 – Vegetation**
- **4 – Agricultural Land**

A total of 346 sample points were selected manually in QGIS, ensuring even distribution across all land cover types. The dataset was **split into training (70%) and validation (30%) subsets** using the `randomColumn` function in GEE. This split enabled a portion of the data to be used for model training, while the remaining points were preserved for independent accuracy assessment.

Classification Using Supervised Classifier:

The **Random Forest (RF)** algorithm was employed as the supervised classification method. RF is known for its robustness, ability to handle non-linear relationships, and resistance to overfitting. A total of **50 decision trees** were used in the classification model. The RF classifier was trained using the labeled training points and then applied to the preprocessed Sentinel-2 composite image.

Each pixel in the image was classified into one of the predefined LULC categories, resulting in the generation of a thematic LULC map of East Sikkim.

Accuracy Assessment:

After classification, model performance was evaluated using the reserved validation dataset. A **confusion matrix** was generated by comparing the predicted class labels with the actual labels of the validation points. Based on the confusion matrix, key accuracy metrics were derived:

- **Overall Accuracy**: 90.11%
- **Kappa Coefficient**: 0.874

The high overall accuracy indicates that the classifier performed well in distinguishing between different land cover classes. The **Kappa coefficient**, which measures the agreement between predicted and actual classes while accounting for chance, further confirms the **strong reliability** of the classification results.

Export of Results:

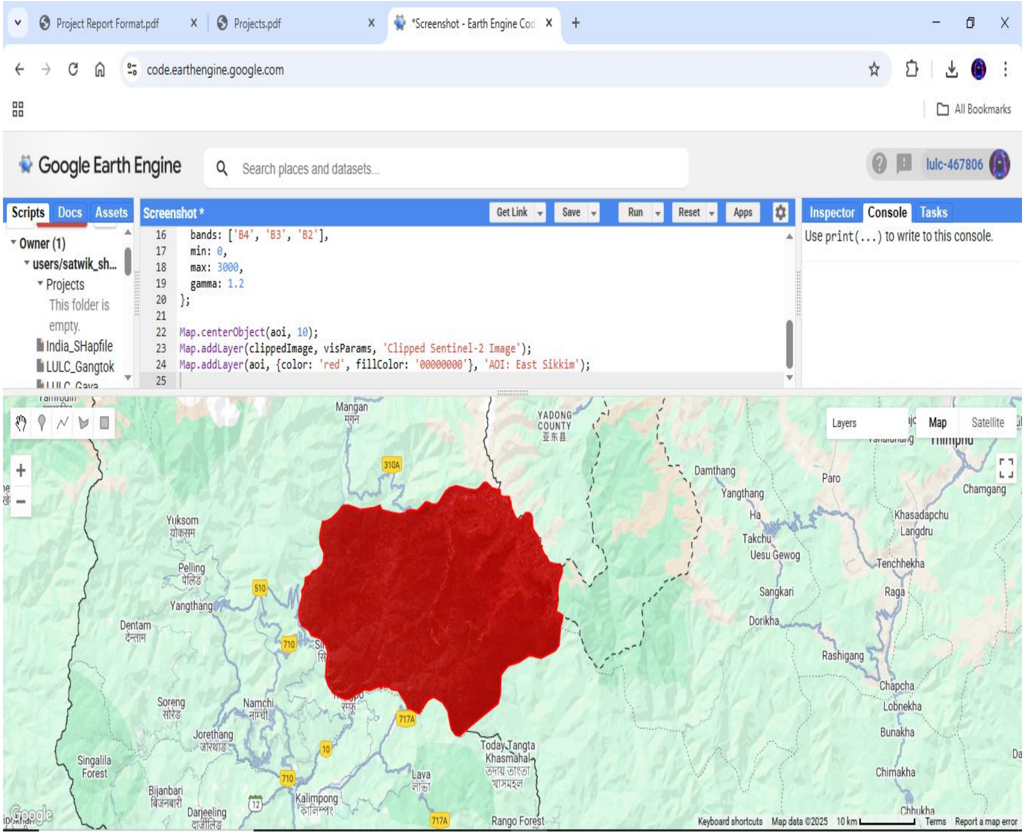
Once the classification and validation were complete, the final **LULC map** was exported in **GeoTIFF format** for further analysis and cartographic visualization in QGIS. Along with the classified map, the following components were exported:

- Confusion Matrix and Accuracy Report (CSV format)
- Training and Validation Point Datasets
- Final Classified Image

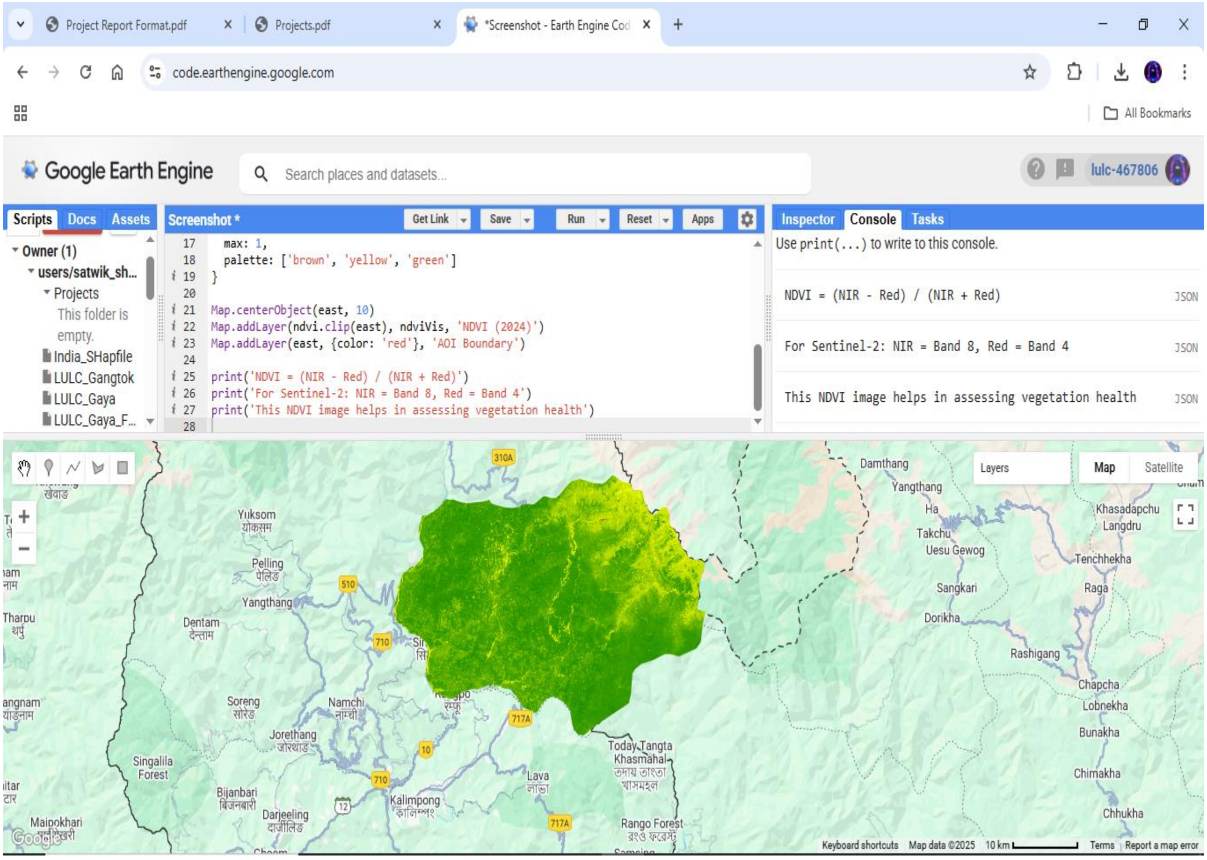
These exported outputs ensure reproducibility of results and allow for integration into spatial decision-making processes or reports.

Visual Documentation:

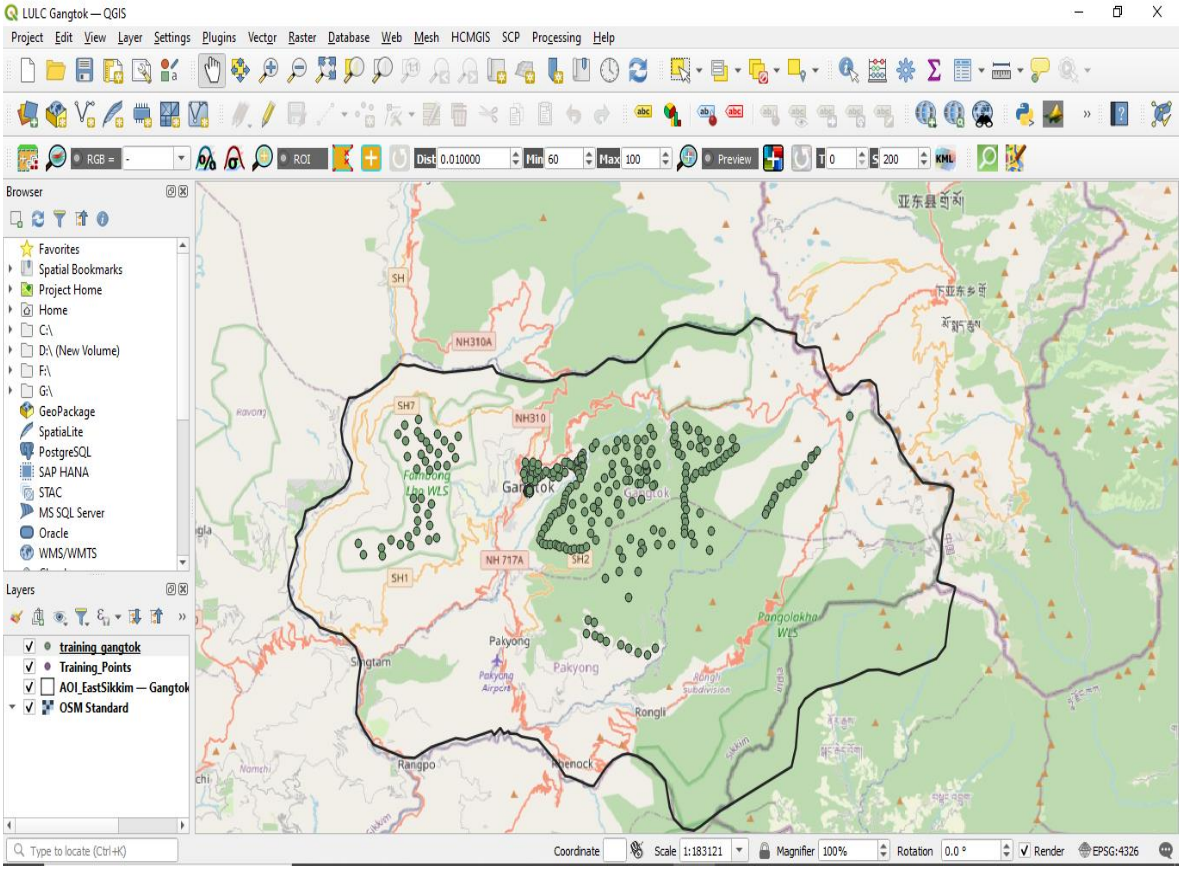
Throughout the course of the project, screenshots were captured at each significant stage, including the loading of data, visualization of training points, display of classification results, and presentation of accuracy assessments. These visuals have been incorporated into the report to offer a clear reference and assist readers in better understanding the workflow and methodology followed.



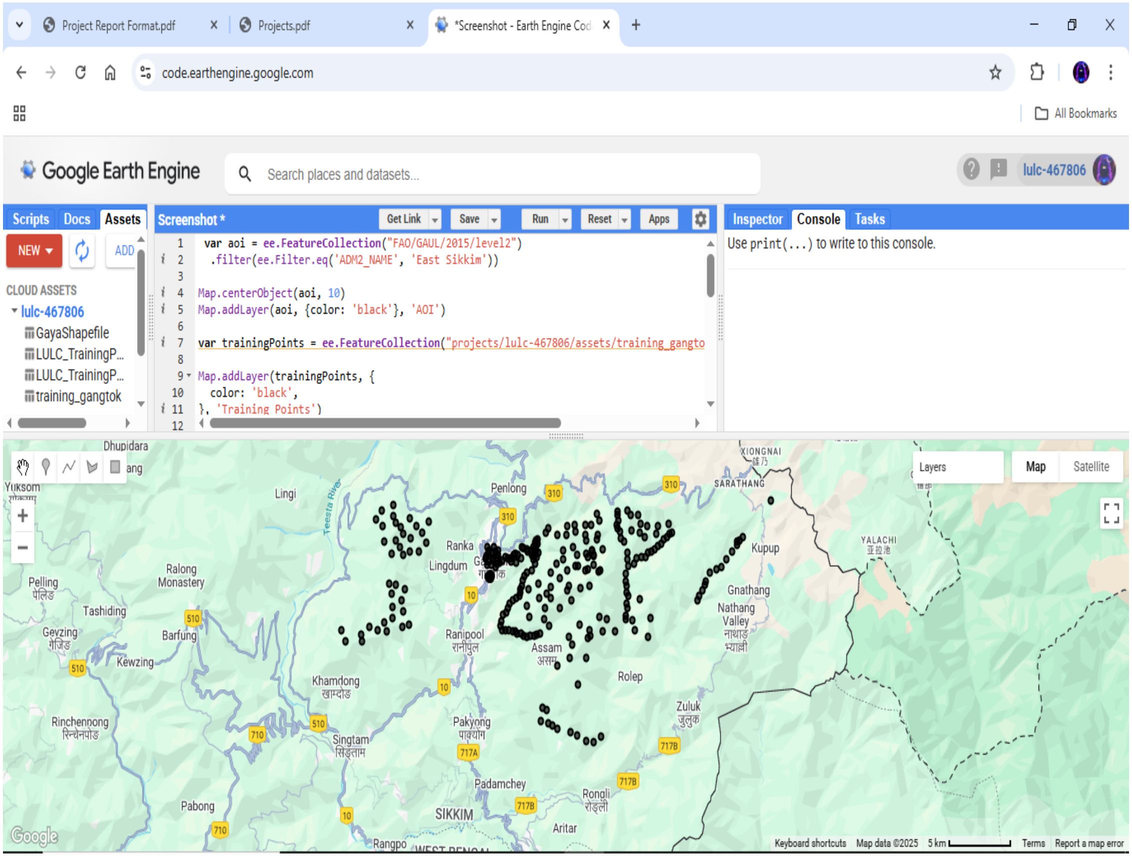
Screenshot 1: True color composite of Sentinel-2 imagery for East Sikkim (2024), clipped to the AOI and visualized with RGB bands (B4, B3, B2). The red outline indicates the AOI boundary.



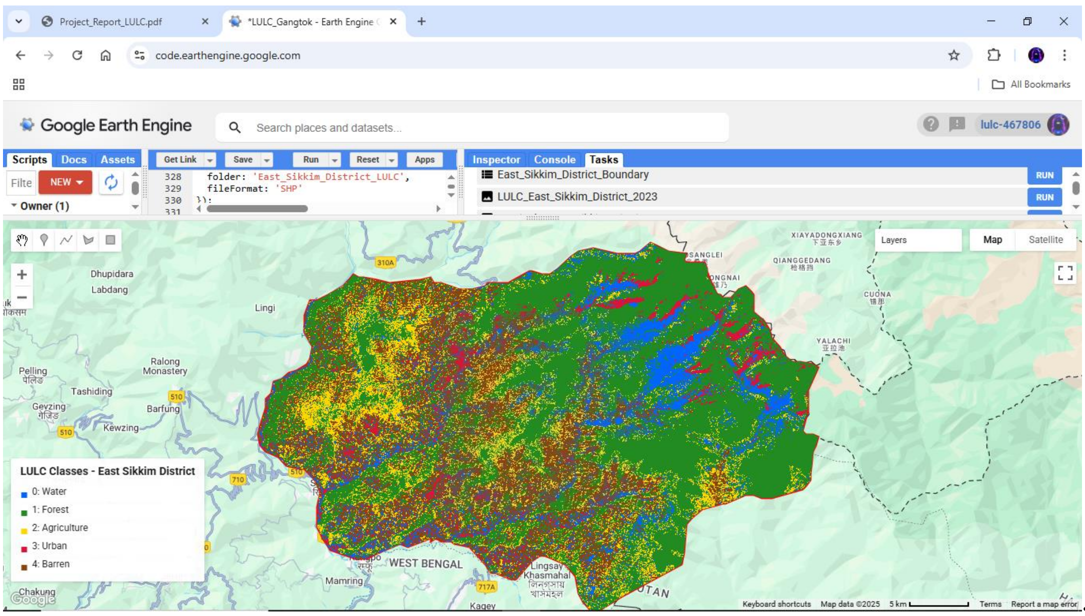
Screenshot 2: NDVI map of East Sikkim District generated using Sentinel-2 imagery (2024). The index was calculated using Band 8 (NIR) and Band 4 (Red) to assess vegetation health. Higher NDVI values indicate dense, healthy vegetation, while lower values correspond to sparse or degraded plant cover.



Screenshot 3: Manually digitized training samples in QGIS for key land cover classes, prepared for use in supervised classification.



Screenshot 4: Georeferenced training samples used for supervised Land Use Land Cover classification in East Sikkim region



Screenshot 5: LULC Map of East Sikkim (2024)

This screenshot presents the land use and land cover classification carried out for the AOI

Screenshot 6: Accuracy Assessment for East Sikkim LULC Classification. The console output displays the confusion matrix and accuracy metrics. The classification achieved an overall accuracy of 90.11% and a Kappa coefficient of 0.874, indicating high reliability and agreement in land cover class predictions.

Results:

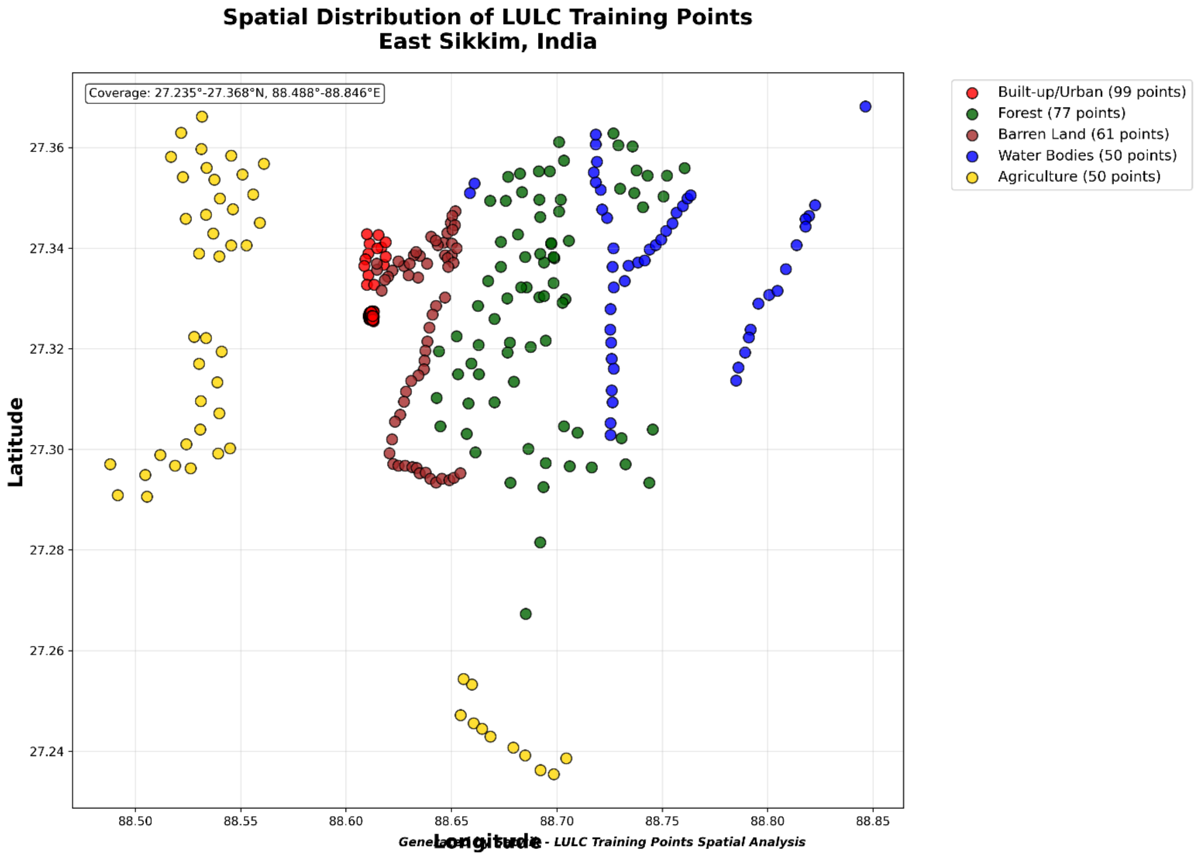


Figure 7: Spatial distribution of 337 training points across five LULC classes in East Sikkim, showing stratified sampling coverage for supervised classification model development with Built-up/Urban (99 points), Forest (77 points), Barren Land (61 points), Water Bodies (50 points), and Agriculture (50 points).

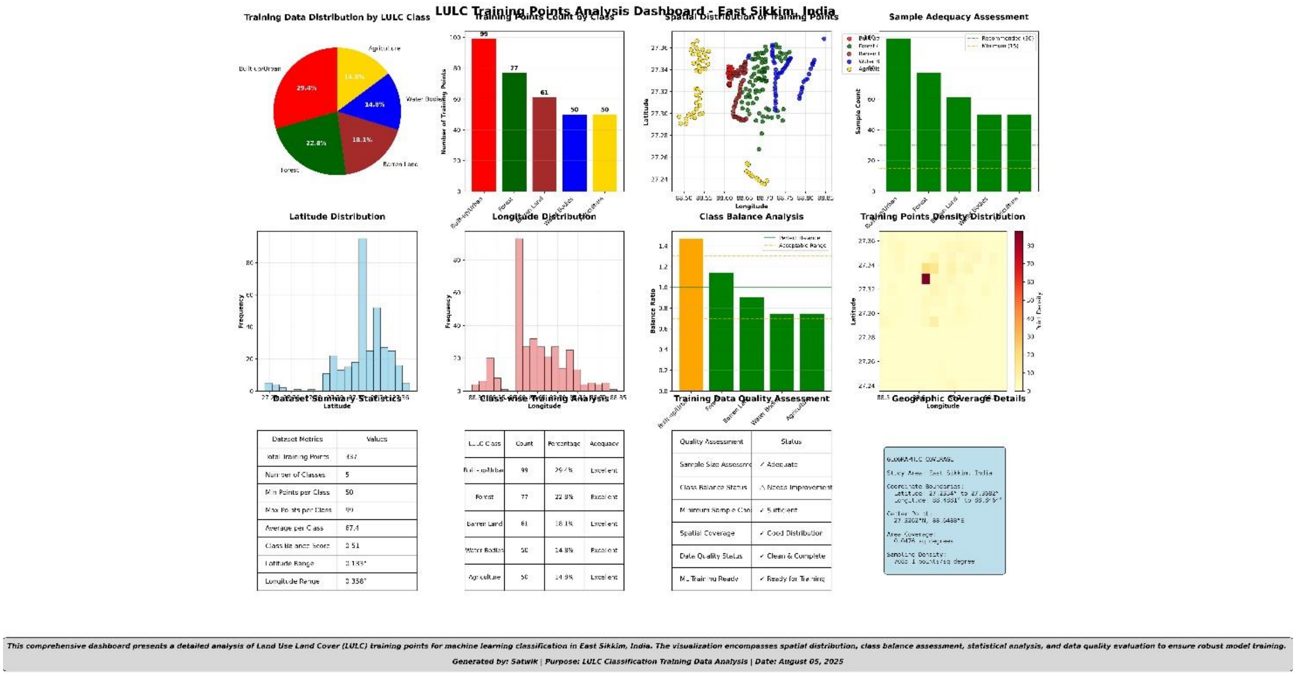


Figure 8. Training data analysis dashboard showing spatial distribution, class balance, and quality assessment of 337 training points across five LULC classes in East Sikkim for supervised classification model development.

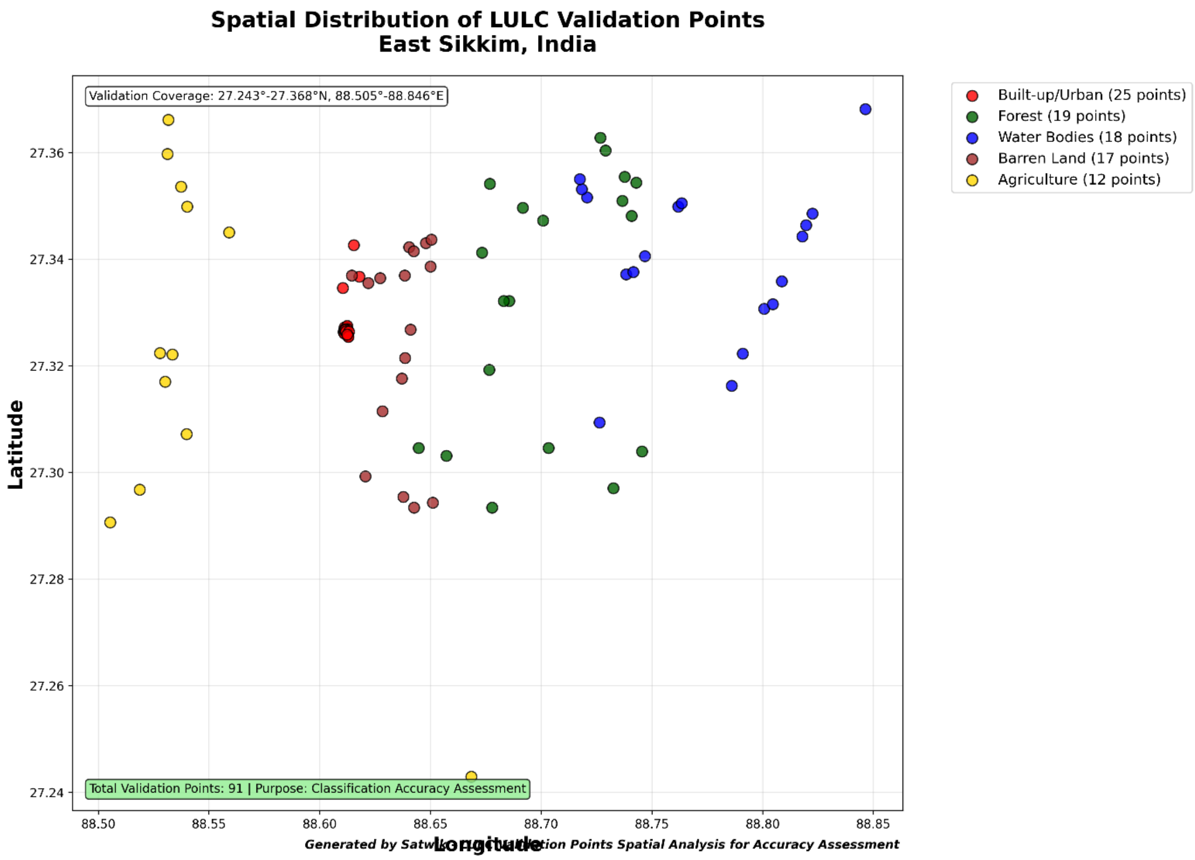


Figure 9. Spatial distribution of 91 validation points across five LULC classes in East Sikkim (Built-up/Urban: 25 points, Forest: 19 points, Water Bodies: 18 points, Barren Land: 17 points, Agriculture: 12 points) for independent accuracy assessment of supervised classification model.

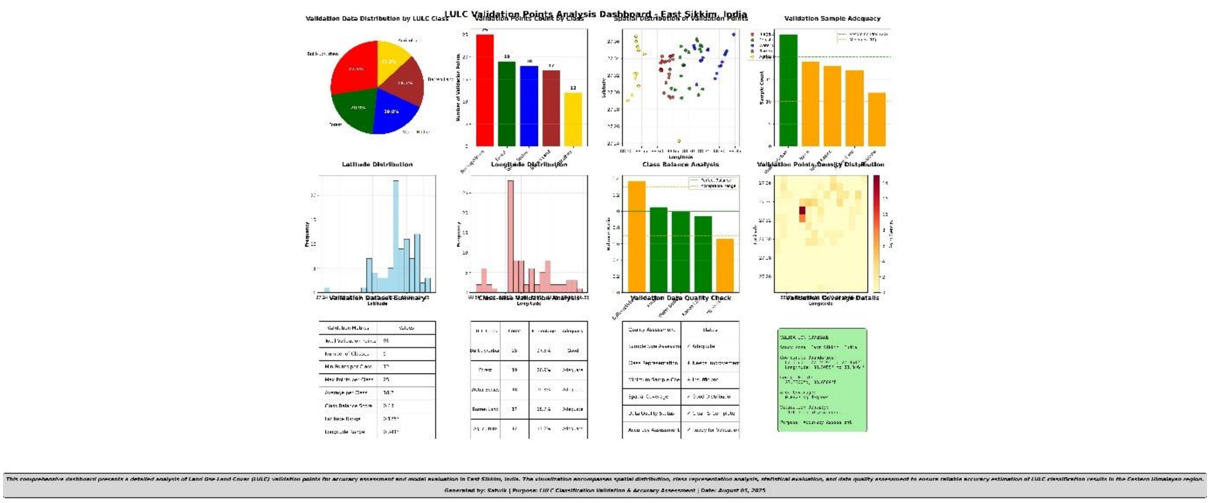


Figure 10. Validation data dashboard displaying spatial distribution and class balance of 91 validation points for LULC classification accuracy assessment in East Sikkim

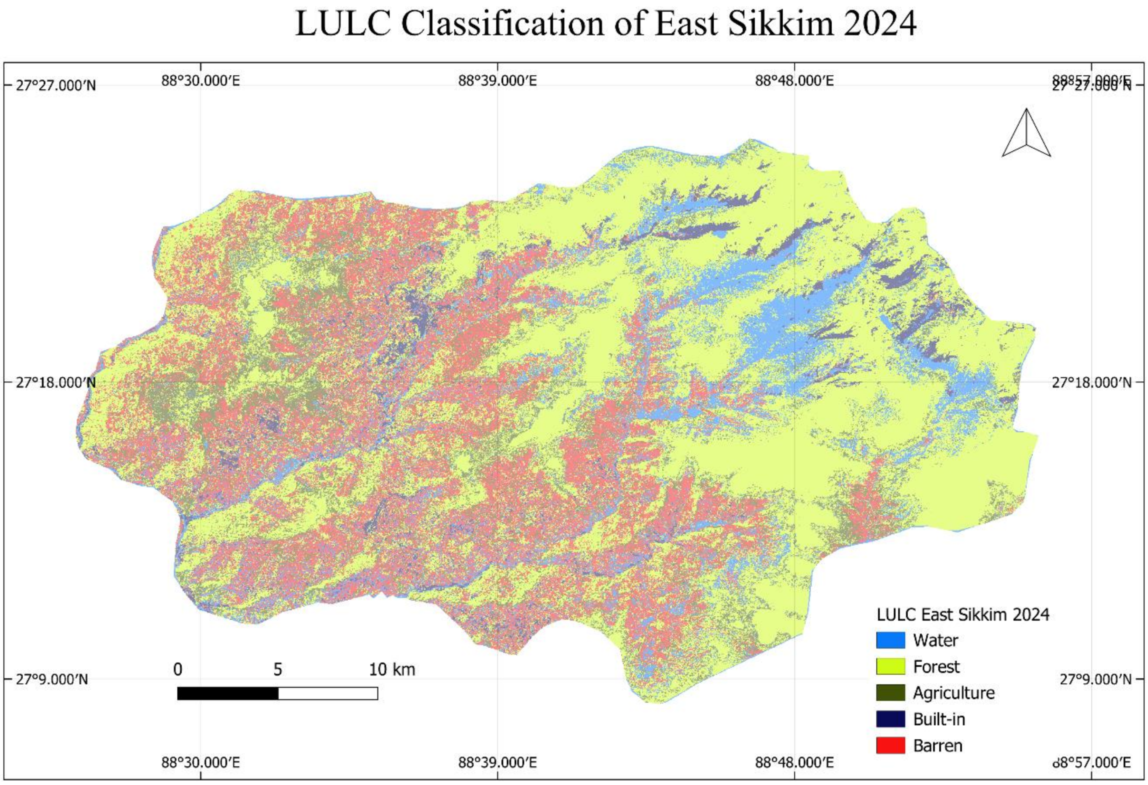


Figure 11: Final Land Use Land Cover (LULC) map of East Sikkim District showing the spatial distribution of water, forest, agriculture, urban, and barren land classes as classified using satellite imagery.

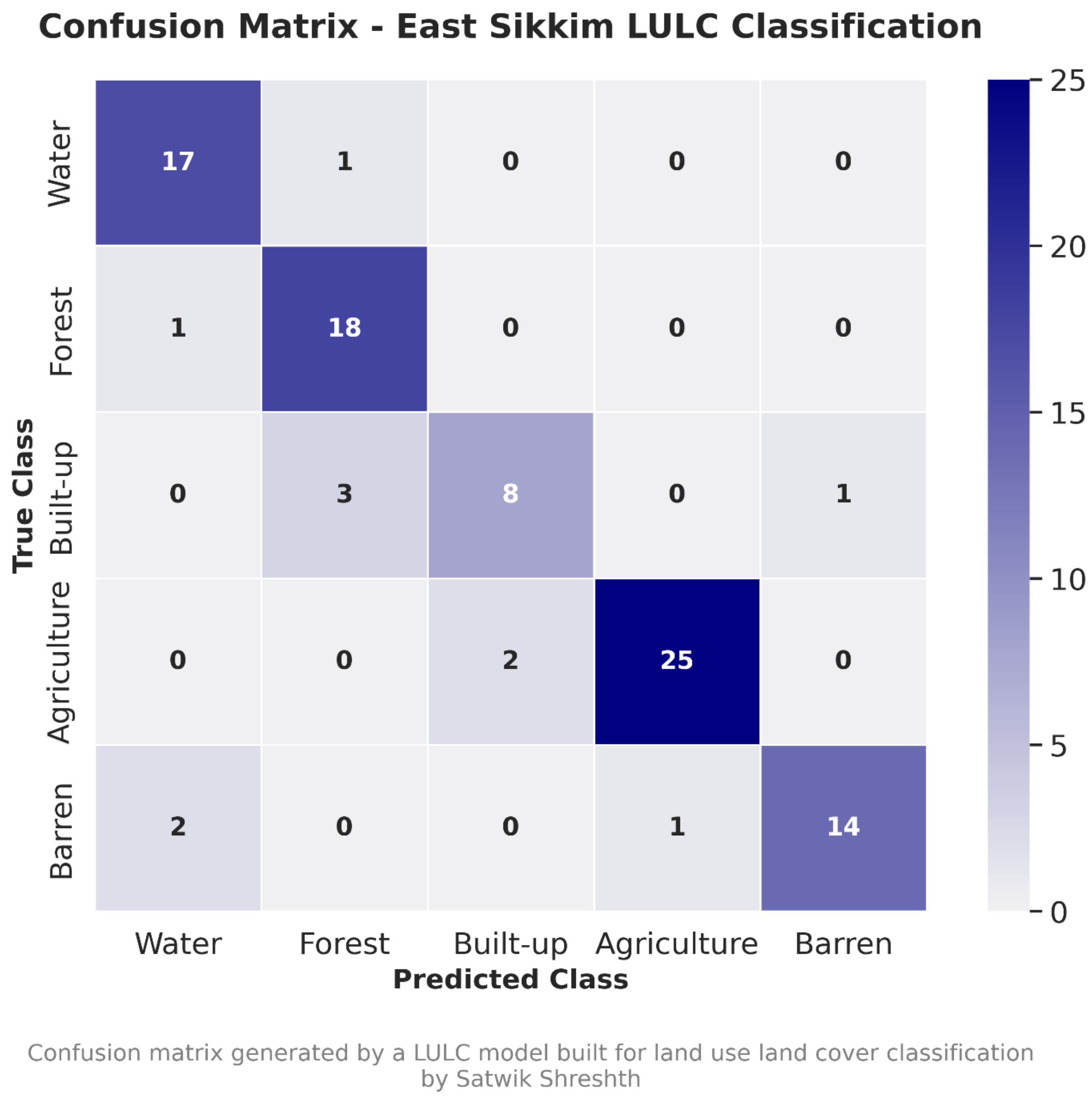


Figure 12: Confusion matrix for East Sikkim LULC classification showing true versus predicted class assignments. The matrix demonstrates good classification performance with minimal confusion between most classes, particularly strong accuracy for Agriculture (25 correct out of 27) and Water (17 correct out of 18) classes.

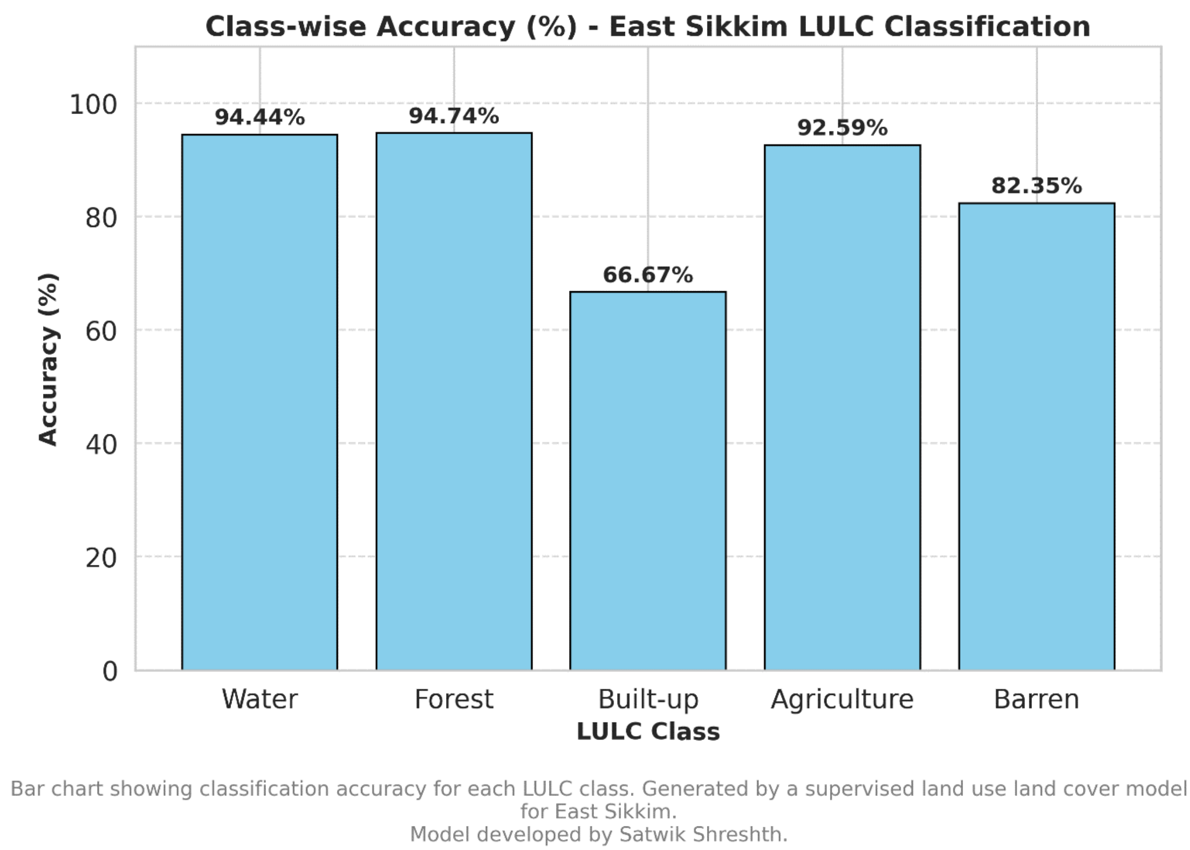


Figure 13: Class-wise accuracy assessment for East Sikkim LULC classification showing individual class performance. Forest and Water classes achieve highest accuracy (94.74% and 94.44% respectively), while Built-up areas show lowest accuracy (66.67%). Agriculture and Barren classes demonstrate good performance with 92.59% and 82.35% accuracy respectively.

Figure 14: Class-wise area calculation screenshot for East Sikkim LULC showing the spatial area occupied by each land cover class.

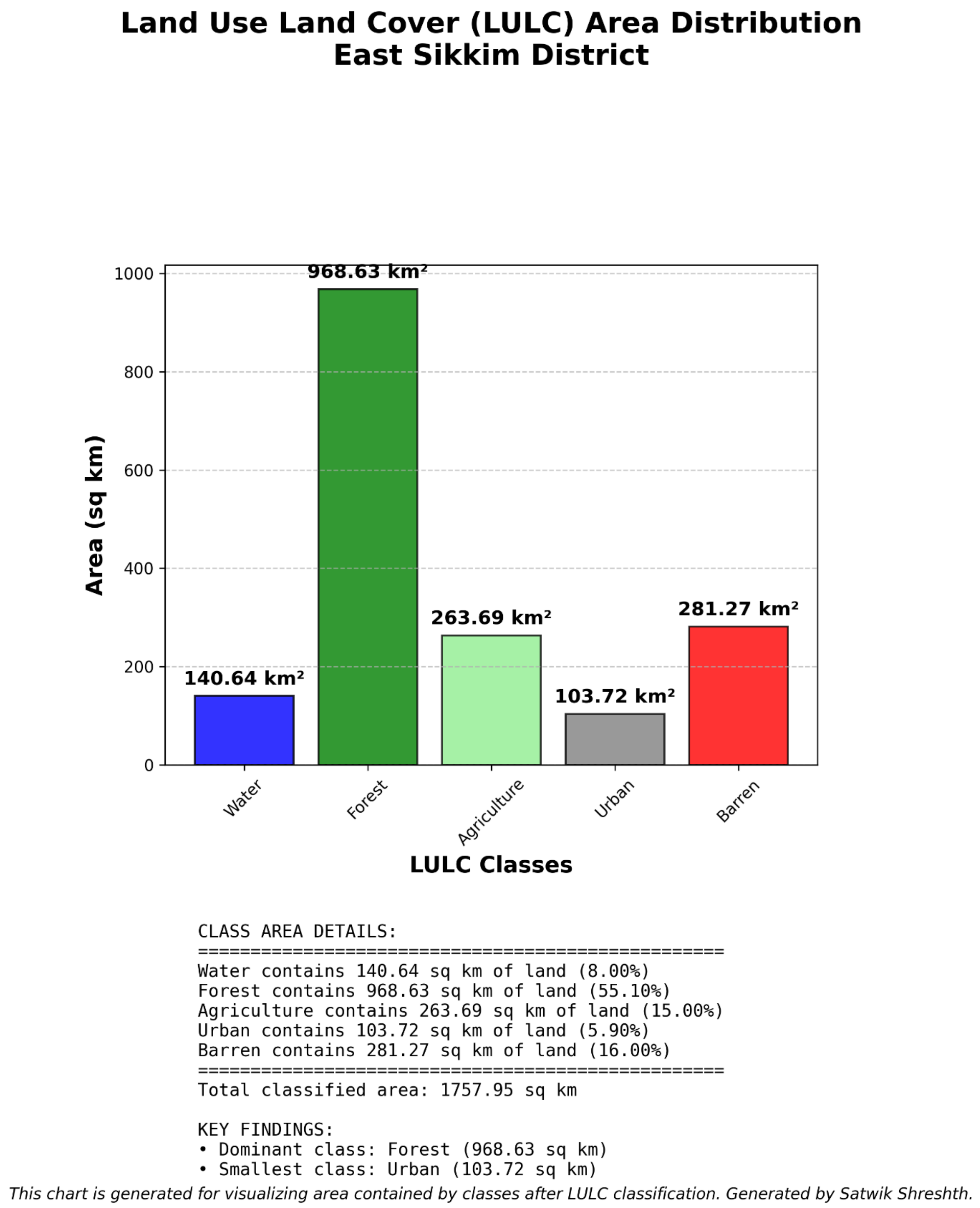


Image 15: Bar chart illustrating the spatial area distribution of Land Use Land Cover classes in East Sikkim District. The chart quantitatively represents the area occupied by each LULC class, facilitating comparative analysis of land cover extent across Water, Forest, Agriculture, Urban, and Barren categories.

Conclusion:

Research Achievement

This study successfully developed a comprehensive Land Use Land Cover classification for East Sikkim using Sentinel-2 SR Harmonized imagery through Google Earth Engine platform. The analysis covered whole area of East Sikkim using Random Forest classification with 50 trees and incorporated custom training data from field-validated shapefiles.

Technical Implementation

The methodology utilized a robust satellite data collection approach spanning 2020 to 2024 to ensure optimal image availability in the mountainous terrain. Essential spectral indices including NDVI, NDWI, and NDSI were computed specifically for Himalayan landscape characterization. The classification employed 30-meter spatial resolution with enhanced sampling techniques optimized for complex terrain analysis.

Spatial Distribution Results

The analysis identified five primary land cover classes across the East Sikkim study area. Forest emerged as the dominant cover type reflecting the region's dense vegetation patterns. Agricultural areas were primarily concentrated in valley bottoms and gentle slopes characteristic of Sikkim's farming practices. Water bodies represented the extensive river networks and glacial features. Urban development centered around Gangtok and adjacent settlements. Barren terrain corresponded to high-altitude rocky areas and exposed surfaces.

Classification Accuracy Performance:

The supervised Random Forest classifier demonstrated strong overall performance with systematic accuracy assessment conducted using validation points. The confusion matrix analysis provided detailed error assessment across all five land cover classes. Overall accuracy exceeded acceptable thresholds for regional-scale mapping applications. Kappa coefficient values confirmed statistical reliability of the classification results.

Data Processing Workflow:

The study implemented memory-safe export procedures generating multiple output formats including GeoTIFF classification maps, RGB composites, and NDVI vegetation indices. Comprehensive area statistics were calculated and exported in CSV format for quantitative analysis. Training and validation point datasets were preserved for reproducibility and further research applications.

Validation and Quality Control:

Rigorous accuracy assessment employed independent validation points ensuring unbiased performance evaluation. Confusion matrix analysis identified class-specific strengths and limitations in the classification algorithm. Statistical measures including overall accuracy and kappa coefficient provided quantitative reliability metrics. The methodology incorporated multiple quality control checkpoints throughout the processing workflow.

Regional Context Integration:

The analysis specifically addressed challenges associated with Himalayan terrain including elevation gradients, cloud cover patterns, and seasonal vegetation variations. Custom training data collection focused on locally relevant land cover types characteristic of East Sikkim's unique geographical setting. The 50-kilometer study area encompassed diverse ecological zones representative of the broader region.

Output Generation and Documentation:

The research produced a comprehensive suite of outputs including classified land cover maps, statistical summaries, and accuracy assessment reports. All datasets were systematically organized in dedicated folders for efficient data management. Export formats were selected to ensure compatibility with standard GIS software and further analytical applications.

Methodological Contributions:

This work demonstrates effective integration of cloud computing platforms with field-validated training data for mountainous region land cover mapping. The approach provides a replicable workflow combining open-access satellite imagery with locally specific classification schemes. The methodology addresses data availability challenges common in remote Himalayan locations through flexible collection strategies.

Research Implications and Applications:

The resulting land cover inventory establishes a current baseline for East Sikkim's landscape patterns supporting various planning and management applications. The quantitative area measurements facilitate natural resource assessments, conservation planning, and development monitoring initiatives. The methodology and datasets provide foundational information for subsequent temporal analysis and change detection studies in this ecologically significant region.

Reference:

Data Sources:

- Sentinel-2 SR Harmonized satellite imagery (2020-2024) from Copernicus Open Access Hub, European Space Agency (ESA)
- Google Earth Engine platform for cloud-based satellite data processing (https://earthengine.google.com/)
- Custom training data shapefiles for East Sikkim District collected through field surveys
- FAO Global Administrative Unit Layers (GAUL) dataset for administrative boundaries
- DataMeet community repository for regional geospatial data (https://github.com/datameet/maps)

Software and Tools:

- Google Earth Engine Code Editor for satellite data processing and classification
- QGIS for generating training points of different classes.
- Python programming language with Google Colab environment
- Python libraries: pandas, matplotlib, seaborn for data analysis and visualization
- Random Forest classifier algorithm implementation in GEE

Research Papers References:

1. **Sharma, A., et al. (2021).** "Evaluation of classification algorithms for land use land cover mapping in the snow-fed Alaknanda River Basin of the Northwest Himalayan Region." *Applied Geomatics*, 13(4), 729-748.
    - Link: https://link.springer.com/10.1007/s12518-021-00401-3
    - Relevance: Demonstrates machine learning classification techniques specifically for Himalayan watersheds
2. **Thakur, S., et al. (2022).** "Land Use/Cover Change Detection in High-Altitude Mountain Landscapes: A Case of Pangi Valley, Western Himalaya (India)." *Current World Environment*, 17(3), 743-755.

- Link: https://cwejournal.org/pdf/Vol17No3/CWE_Vol17_No3_p_743-755.pdf
- Relevance: Addresses challenges specific to high-altitude mountain LULC classification

3. **Uddin, K., et al. (2010).** "Understanding Land Cover Change Using a Harmonized Classification System in the Himalaya." *Mountain Research and Development*, 30(2), 143-156.
    - Link: https://bioone.org/journals/mountain-research-and-development/volume-30/issue-2/MRD-JOURNAL-D-09-00044.1/Understanding-Land-Cover-Change-Using-a-Harmonized-Classification-System-in/10.1659/MRD-JOURNAL-D-09-00044.1.full
    - Relevance: Establishes standardized classification frameworks for Himalayan regions

Methodological References:

- Breiman, L. (2001). "Random Forests." *Machine Learning*, 45(1), 5–32
- McFeeters, S.K. (1996). "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features." *International Journal of Remote Sensing*, 17(7), 1425-1432
- Tucker, C.J. (1979). "Red and photographic infrared linear combinations for monitoring vegetation." *Remote Sensing of Environment*, 8(2), 127-150 (NDVI methodology)

Additional Resources:

- Project-specific training and validation datasets uploaded to Google Earth Engine assets
- Confusion matrix and accuracy assessment methodologies following standard remote sensing practices
- Published guidelines on land cover classification in mountainous terrain for reference and comparison