ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-5/W2-2025
Unleashing the power of Geospatial & Frontier Technologies for a Sustainable Future – GIFTS Summit 2025,
1–3 September 2025, Symbiosis International University, Pune, India

Deep Learning Based Land Use Land Cover Classification On Multi-Sensor Remote Sensing Data

Advaith C A[1], Shefali Agrawal[2], Vinay Kumar[1], Shashi Kumar[1], Poonam S. Tiwari[1]

[1] Photogrammetry and Remote Sensing Department, Indian Institute of Remote Sensing,
Dehradun, India - (advaith.iirs, vinayiirs, sksinghiirs)@gmail.com, poonam@iirsddn.ac.in
[2] Geospatial Technology and Outreach Program, Indian Institute of Remote Sensing,
Dehradun, India - s_agrawal@iirs.gov.in

Keywords: Deep Learning, Remote Sensing, Land Use Land Cover Classification, Multi-Sensor Remote Sensing Data

Abstract

Accurate Land Use Land Cover (LULC) classification is critical to monitor urban expansion, resource planning, and environmental planning. This research investigates the combination of multi-modal remote sensing data—optical (Sentinel-2, PRISMA), Synthetic Aperture Radar (Sentinel-1), Digital Elevation Model (Cartosat-derived), and Global Human Settlement Layer (GHSL)—for LULC mapping using a deep learning model. A Vision Transformer-based model, SegFormer with Spatial Attention, was used to leverage the capabilities of the sensors optimally. The research was carried out over Dwarka, Delhi, selected due to its composite LULC patterns of urban, grassland, forest, and water bodies. The data stack was resampled to 10m resolution, segmented to 256×256 tiles, and augmented to improve model generalizability. Semi-manual annotations were employed for supervised training. The model was trained with a combination of cross-entropy and dice loss, and tested with precision, recall, F1-score, and overall accuracy. The results reveal the significant enhancement of classification performance with multi-sensor integration, where the model achieved an overall accuracy of 86.9%. The inclusion of each data source improved performance, especially when combining optical and SAR data with GHSL. This research illustrates the potential of transformer-based models in remote sensing applications, specifically in exploring multi-source satellite data for sophisticated LULC mapping. Future research can include large-scale training datasets, region-specific tuning, or architectural variations to enhance model adaptability and robustness.

1. Introduction

Accurate Land Use Land Cover (LULC) classification is essential for urban monitoring, environmental assessment, and resource planning. With increasing urbanization, the demand for timely and reliable LULC maps has grown significantly. Remote sensing provides a synoptic view of large areas and has been a key tool in generating such maps.

Traditional LULC mapping primarily uses optical satellite imagery, which may be limited in heterogeneous landscapes or under cloudy conditions. The advent of multi-modal remote sensing data—including Synthetic Aperture Radar (SAR), Digital Elevation Models (DEM), and human activity proxies like the Global Human Settlement Layer (GHSL)—offers new opportunities. These diverse data capture complementary surface characteristics such as spectral reflectance, surface structure, elevation, and built-up intensity.

Concurrently, deep learning methods have significantly advanced remote sensing image classification. Transformer-based models, especially SegFormer, have demonstrated strong performance in semantic segmentation by modeling long-range dependencies and spatial context.

This study explores the integration of multi-sensor data—Sentinel-1 (SAR), Sentinel-2 and PRISMA (optical), Cartosat-derived DEM, and GHSL—for LULC classification using a modified SegFormer with spatial attention. The Dwarka sub-region of Delhi was selected as the study area due to its diverse urban and natural land cover types. All datasets were resampled to a common resolution and prepared for model training using semi-manual annotations.

The proposed model was trained with a hybrid loss and evaluated using standard classification metrics. Results show that combining SAR and GHSL with optical data improves classification performance, achieving an overall accuracy of 86.9%. This work demonstrates the potential of transformer-based models for high-resolution, multi-sensor LULC mapping and lays the groundwork for scalable and robust classification frameworks.

2. Literature Review

Understanding land use land cover (LULC) classification using multi-sensor remote sensing data is critical for building scalable, deep learning-based models. Several key themes emerge in this domain: foundational methodologies, sensor fusion strategies, and recent advances in learning-based techniques.

2.1 Foundational Approaches:

Classical frameworks such as Anderson et al. (Anderson, 1971) and the NRSC LULC classification schemes form the basis of most schema designs. Richards (Richards and Jia, 2006) provides essential insights into satellite image preprocessing, including geometric correction and image registration, which are vital when working with multi-sensor datasets.

2.2 Multi-Sensor Data Fusion:

The integration of optical, SAR, DEM, and hyperspectral data has shown significant potential in improving classification accuracy. Studies such as (Parmegiani and Posolieri, 2002, Thurmond et al., 2006) highlight the value of combining topographic

and spectral information. Fusion techniques such as discriminant correlation analysis (Wang et al., 2020) and statistical feature selection over multi-resolution segmentation (Pelizari et al., 2018) enable more robust feature extraction. Additionally, spectral indices like NDVI, NDBI, and NDWI are widely used to enhance class separability and reduce dimensionality (Kaur and Pandey, 2022, Kebede et al., 2022).

2.3 Classification Techniques:

Machine learning models like SVM and Random Forest have been successfully applied to fused datasets (Huang et al., 2002, Uddin et al., 2021), often with region-based segmentation techniques (Valdivieso-Ros et al., 2023). Recent work leverages deep learning architectures such as U-Net with attention mechanisms to combine SAR, optical, and topographic features effectively (Chen et al., 2024). Super-resolution-based approaches (Han et al., 2024) and reproducible, open-source SAR-based classification pipelines (Hütt et al., 2020) also demonstrate growing emphasis on scalability and precision.

The reviewed literature confirms that multi-sensor data fusion, when combined with advanced classification methods, significantly improves LULC mapping performance. Continued progress in fusion strategies and learning frameworks will be key to developing generalizable, high-accuracy classification systems.

3. Methodology

3.1 Study Area

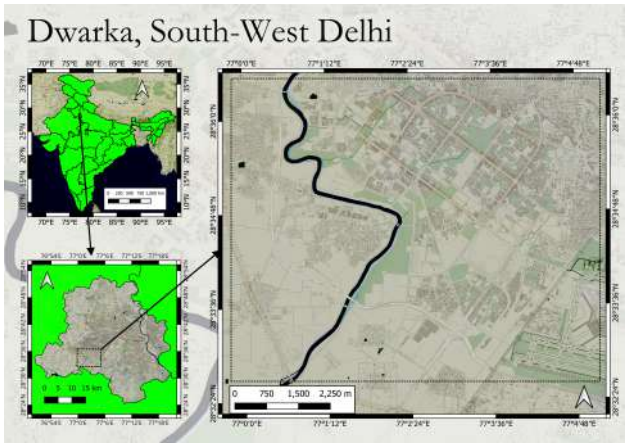


Figure 1. Study Area

The Study Area, as shown in Figure 1, is a region in Dwarka, New Delhi. It encompasses the Sectors in Dwarka, along with regions of Qutub Vihar, Goyla Khurd, Shyam Vihar, Chhawla, Kakrola, Palam Village, Raj Nagar and the Indhira Gandhi International Airport. It covers approximately $62km^2$ area. The bounding box coordinates of this area is given in the Table 1.

This region was chosen due to the variety of LULC classes observed here, including Forest Cover, Grasslands, Industrial Areas and Various Different types of Urban Residential Built-up areas. These characteristics that the study area possesses make it an ideal testing ground for the developed SegFormer (Xie et al., 2021) based LULC classification algorithm.

Table 1. Geographic Coordinates of Study Area

Corner	Longitude (°E)	Latitude (°N)
Lower Left	76.9963	28.5434
Upper Right	77.0864	28.6091

3.2 Datasets

A major part of the research work conducted was the integration of msrs data. Ergo, data spanning multiple sensors and satellites were used. They are summarized in the Table 2.

Table 2. Details of Satellite Data and Other Geospatial Products Used

Satellite / Product	Resolution	Year(s)
Sentinel-1 RTC γ^0	10 m	2024, 2016
Sentinel-2 L2A	10–60 m	2024, 2016
PRISMA L2D (Super Res.)	30 m (10 m)	2024, 2016
GHSL	10 m	2018
Cartosat DEM	2.5 m	—

This study employed multiple satellite and geospatial datasets to ensure comprehensive spatial and spectral coverage. Sentinel-1 SAR (RTC γ^0 product) provided weather-independent backscatter data at 10 m resolution, useful for detecting built-up surfaces. Sentinel-2 MSI imagery, with 10 m resolution in Bands 2, 3, 4, and 8 (Figure 2, 6), offered optical information critical for vegetation and general land cover mapping.

PRISMA hyperspectral data (originally at 30 m resolution) was super-resolved to 10 m and used to derive the Hyperspectral Built-up Index (HIBI) (Gaur et al., 2023), aiding urban area identification. Additional layers included the Global Human Settlement Layer (GHSL, 2018) for built-up density, and Cartosat-derived DEM (2.5 m) for topographic context.

The integration of these datasets enables a multi-modal perspective—combining spectral, structural, and elevation information—to support detailed land cover classification in complex urban environments, where single-source imagery may fall short.

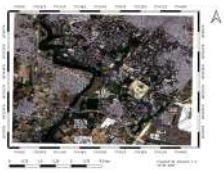


Figure 2. True Color Composite of the Sentinel-2 Scene.

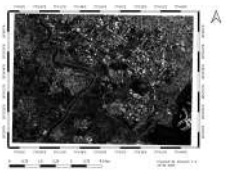


Figure 3. Sentinel-1 VV Polarized RTC Gamma Imagery.

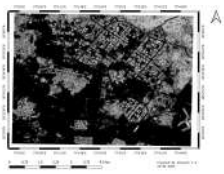


Figure 4. GHSL Data.

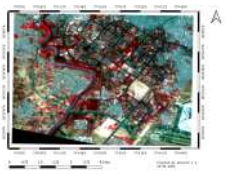


Figure 5. Super Resolved PRISMA Data.

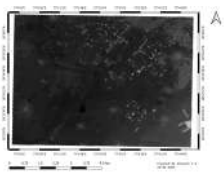


Figure 6. Cartosat DEM.

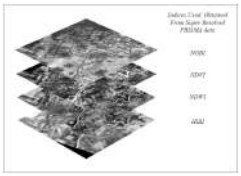


Figure 7. Indices Derived from PRISMA Super Resolved Data.

4. Methodology

This section presents the methodology followed for land cover classification using multi-sensor remote sensing (MSRS) data. The methodology is divided into two broad strategies: a traditional approach using Random Forest and an end-to-end deep learning model based on SegFormer with a Spatial Attention Module. The entire workflow is summarized in Figure 8.

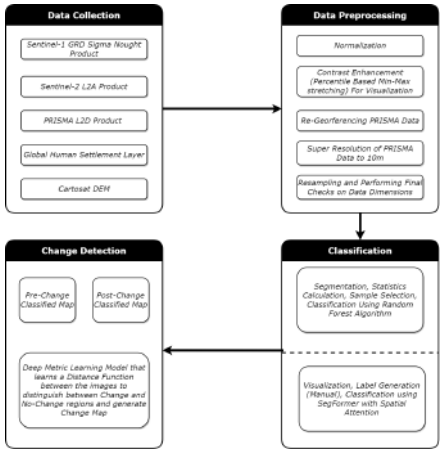


Figure 8. Overview of the Methodology.

4.1 Visual and Analytical Understanding of MSRS Data

4.1.1 Visual Analysis An initial visual inspection of the multi-sensor data was carried out to understand which data sources are useful for distinguishing different land cover classes. Based on expert interpretation using Google Earth imagery and Sentinel-2 composites, a classification schema was designed, as shown in Table 3. The utility of each data source was assessed and summarized in Table 4.

Table 3. Land Cover Classes and their Codes

Code	Class	Code	Class
0	Crop Land	9	Road
1	Fallow/Barren Land	10	Bridge
2	Plantation	11	Flyover
3	High Density Residential	12	Forest
4	High Rise Apartments	13	Dry Vegetation
5	Low Rise Apartments	15	Dry River Bed
6	Commercial Structures	16	Train Track
7	Grassland	17	Industrial
8	Water Body	18	Runway

Table 4. Data Features and Their Utility in Detecting Land Cover Classes

Data / Feature	Benefitted Classes
B2, B3, B4, B8 (Sentinel-2)	General differentiation across all classes, especially vegetation: Crop Land, Plantation, Forest, Grassland
NDVI	Crop Land, Plantation, Forest, Grassland, Dry Vegetation
NDWI	Water Body, Shallow Water, Wetlands
NDBI	High Density Residential, High Rise Apartments, Low Rise Apartments, Commercial, Industrial
DEM	Elevation-based classes: Forest, Plantation, Grassland, Train Track, Bridge, Flyover
Slope	Forest, Train Track, Bridge, Flyover (indirect terrain variation)
Aspect	Forest, Plantation, Grassland (sun exposure and vegetation patterns)
GHSL	Residential, Commercial, Industrial (urban classes); not useful for Road, Bridge, Flyover
VV (SAR)	Industrial, Runway, Road (detects hard surfaces, penetrates vegetation)
HIBI	Residential, Commercial, Industrial, Runway, Apartments (Low/High Rise)

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W2-2025-7-2025 | © Author(s) 2025. CC BY 4.0 License.

Figure 9. Classification chips corresponding to the defined schema.

4.1.2 SHAP-Based Analysis To quantitatively evaluate the importance of each input feature, SHAP (SHapley Additive exPlanations) analysis was performed using a Random Forest classifier. SHAP values provide insight into how individual features contribute to classification decisions. An overview of the SHAP process is shown in Figure 10.

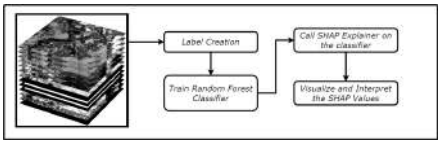


Figure 10. SHAP-based analysis overview.

4.2 Classification Approaches

4.2.1 Random Forest Classifier The Random Forest algorithm was employed for initial classification. The 2024 data stack included Sentinel-1, Sentinel-2, NDVI, NDWI, NDBI, HIBI, GHSL, and Cartosat-derived DEM, slope, and aspect. Object-based segmentation was carried out using the Felzenszwalb algorithm, followed by feature extraction and classification using Random Forest with 150 decision trees.

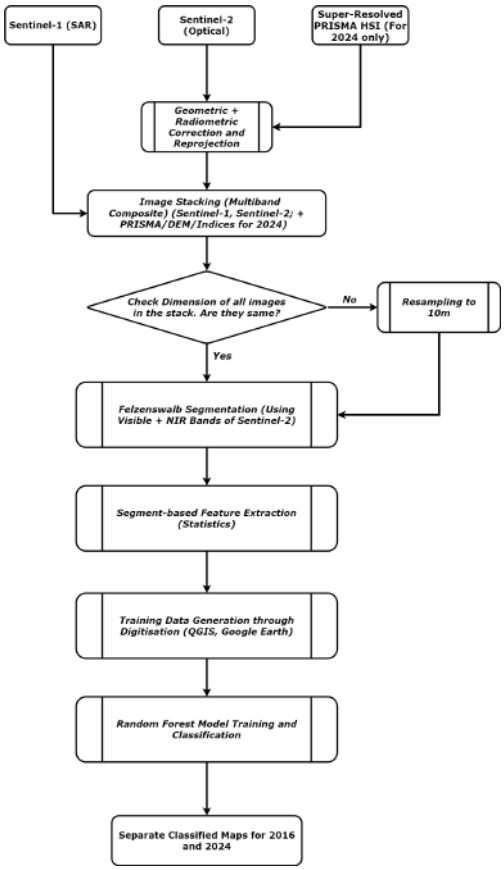


Figure 11. Random Forest classification workflow.

4.2.2 SegFormer with Spatial Attention Module To overcome the time and complexity of manual feature extraction, a deep learning model based on the SegFormer architecture was implemented. The model was modified to accept a 13-band input corresponding to the data layers. A Spatial Attention Module (SAM) was introduced to enhance focus on relevant regions in the imagery.

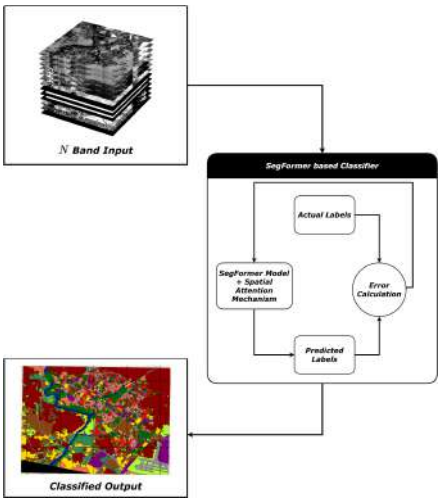


Figure 12. SegFormer classifier overview.

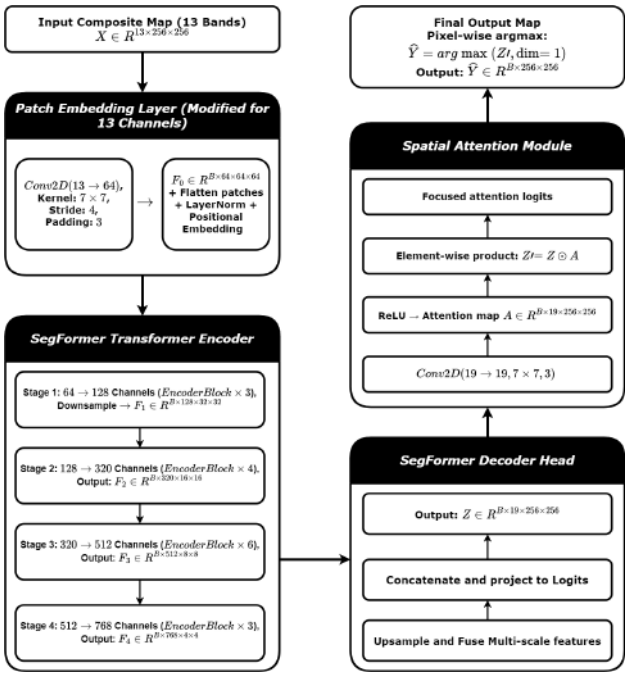


Figure 13. SegFormer Architecture with Spatial Attention.

Training Data and Patch Preparation Labeled training data was generated using Mean Shift segmentation and manual annotation. A total of 6614 segments were labeled. To support deep learning, patches of size 256×256 were extracted from the images and masks.

Training Details The model was trained using the AdamW optimizer with a learning rate of 0.0001 for 100 epochs. Augmentations included rotations, flips, and brightness/contrast adjustments. A Spatial Attention Module applied over the decoder logits helped improve focus on semantically relevant areas.

Comparison of Classifiers

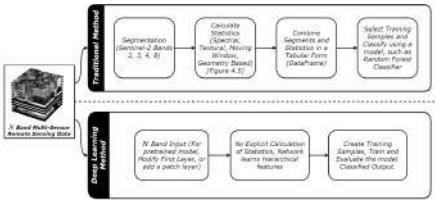


Figure 14. Comparison between Random Forest and SegFormer-based classifier.

Table 5. Comparison between Random Forest and SegFormer with Spatial Attention

Aspect	**Random Forest Classifier**	**SegFormer with Spatial Attention**
Model Type	Ensemble learning using decision trees	Transformer-based deep neural network
Input Requirements	Structured tabular input	Multi-channel image tensor
Feature Engineering	Requires manual computation from segments	Learns features end-to-end automatically
Spatial Context Usage	Indirect, via segment statistics	Direct modeling through attention mechanisms
Scalability	Efficient but preprocessing-intensive	Training is heavier, but more scalable
Interpretability	High; feature importance is measurable	Lower; complex architecture, less transparent
Remote Sensing Suitability	Effective for pixel- and object-based classification	Strong performance in high-resolution multi-sensor image classification

Table 6. Comparison between CNN-based and SegFormer-based Models

Aspect	**CNN-based Models**	**SegFormer with Spatial Attention**
Model Type	Convolutional neural network	Transformer-based segmentation model
Feature Extraction	Uses convolutional filters for local features	Learns both local and global features using attention
Spatial Context	Captures through receptive fields	Captures long-range dependencies via self-attention
Generalization	Good with strong augmentation	Stronger due to modeling global context
Performance	Strong, e.g., U-Net, DeepLabV3+	State-of-the-art for remote sensing classification

5. Results

This section presents the land cover classification results for the year 2024 using both Random Forest and the proposed SegFormer with Spatial Attention model. The performance of each method is evaluated using accuracy, precision, recall, F1-score, and confusion matrices. Additionally, SHAP analysis is used to interpret feature importance for multi-sensor inputs.

5.1 SHAP Analysis

SHAP (SHapley Additive exPlanations) analysis was used to evaluate the contribution of each feature to model predictions.

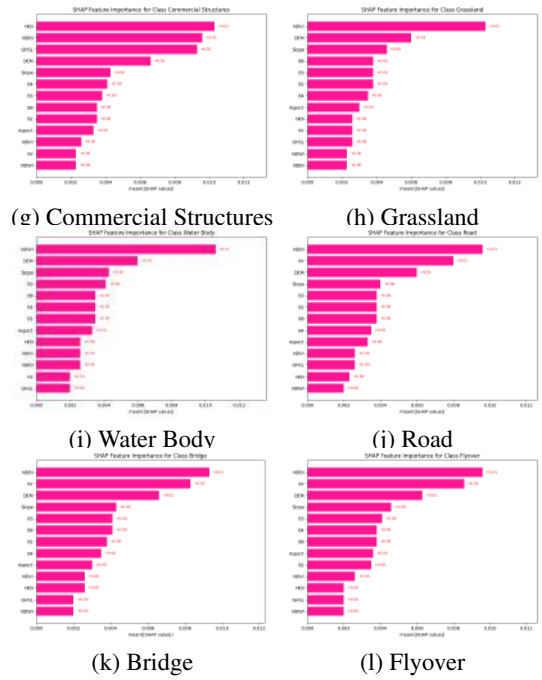
figures 15, 16, and 17 show the SHAP value plots for different classes.

(a) Crop Land (b) Fallow/Barren Land

(c) Plantation (d) High Density Residential

(e) High Rise Apartments (f) Low Rise Apartments

Figure 15. SHAP bar plots showing feature importance for each land cover class. (Part 1)

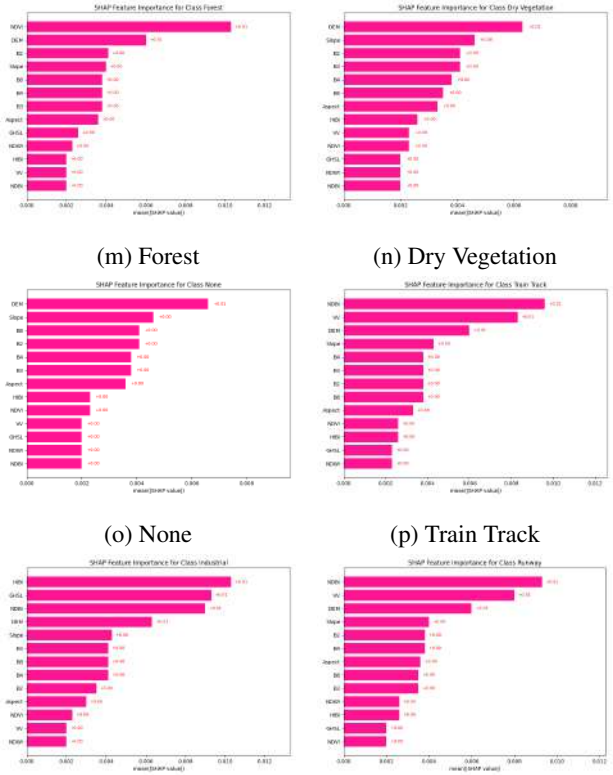


(g) Commercial Structures (h) Grassland

(i) Water Body (j) Road

(k) Bridge (l) Flyover

Figure 16. SHAP bar plots showing feature importance for each land cover class. (Part 2)



(m) Forest (n) Dry Vegetation

(o) None (p) Train Track

(q) Industrial (r) Runway

Figure 17. SHAP bar plots showing feature importance for each land cover class. (Part 3)

Insights.

- Sentinel-2 reflectance bands and NDVI are key for vegetation classes.
- Built-up indices (NDBI, HIBI) are critical for urban structures.
- DEM and slope help distinguish elevation-dependent infrastructure.
- VV SAR backscatter improves performance for impervious surfaces.
- GHSL contributes to urban density discrimination.

These results validate the selection of multi-sensor features and justify their class-specific contributions.

5.2 Classification using Random Forest

Figure 18 shows the 2024 land cover classification result from the Random Forest model. The model achieved an overall accuracy of **86.0%**, with the classification report and confusion matrix shown in figures 19 and 20.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W2-2025-7-2025 | © Author(s) 2025. CC BY 4.0 License.

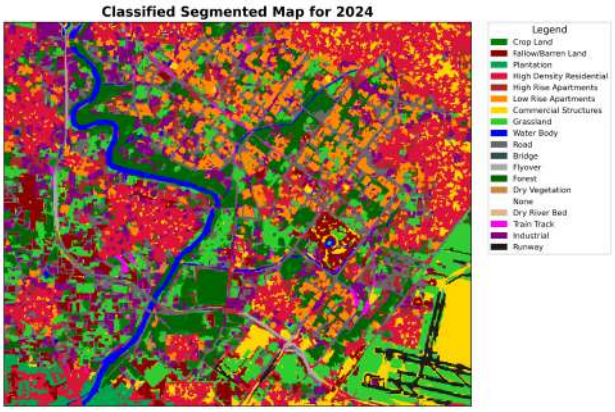


Figure 18. Land cover classification map (2024) using Random Forest.

	Precision	Recall	F1-Score	Support
0	1.0	1.0	1.0	1.0
1	1.0	1.0	1.0	2.0
2	0.0	0.0	0.0	0.0
3	0.9	0.85	0.88	13.0
4	0.71	0.57	0.63	7.0
5	0.86	1.0	0.92	6.0
6	0.95	0.99	0.97	75.0
7	1.0	1.0	1.0	3.0
8	0.88	0.89	0.89	9.0
9	0.83	0.91	0.87	11.0
10	1.0	0.75	0.86	4.0
11	1.0	0.5	0.67	6.0
12	1.0	0.89	0.94	9.0
13	1.0	0.67	0.8	3.0
16	0.8	0.75	0.77	8.0
17	0.84	0.94	0.89	16.0
18	1.0	1.0	1.0	2.0

Figure 19. Classification report for Random Forest (2024).

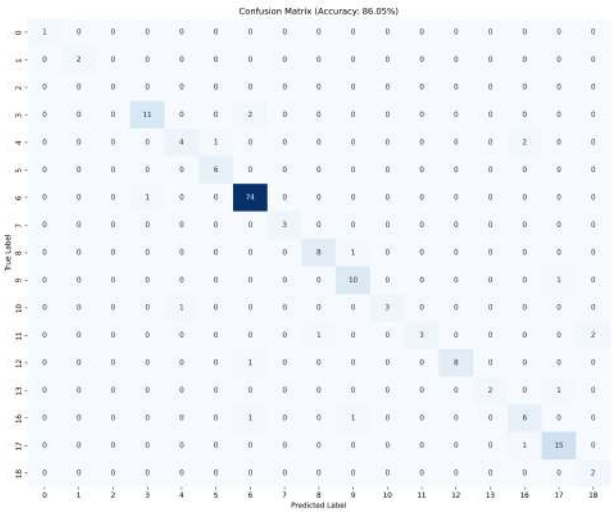


Figure 20. Confusion matrix for Random Forest (2024).

5.3 SegFormer with Spatial Attention

The SegFormer model with Spatial Attention achieved a peak accuracy of **88%** on the full 13-band dataset. It demonstrated strong performance even with a limited number of training samples.

5.3.1 Effect of Feature Subsets The model was tested with different input subsets to evaluate the effect of added data. The five configurations were:

- Optical only (4 bands)
- Optical + SAR (5 bands)
- Optical + SAR + DEM (8 bands)
- Optical + SAR + DEM + Indices (11 bands)
- All bands (13 total)

Classification outputs for each configuration are shown in Figures 22 and 21.

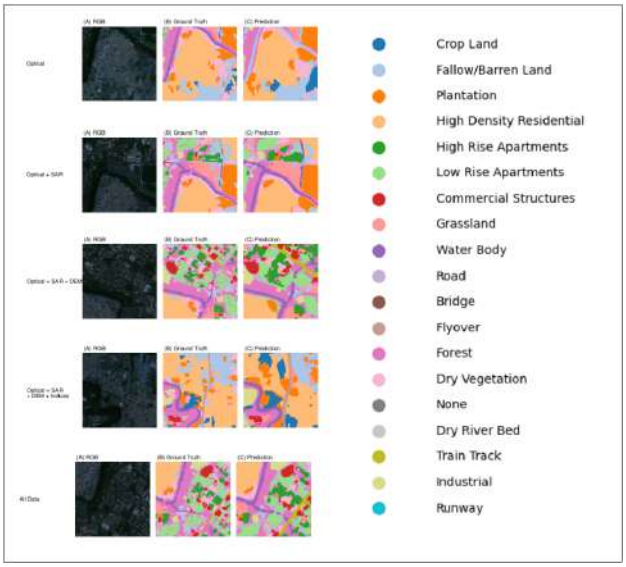


Figure 21. Qualitative output of SegFormer for different input subsets.

Optical	Optical + SAR	Optical + SAR+ DEM	Optical + SAR + DEM + Indices	All Features
0.723	0.792	0.654	0.702	0.879

Figure 22. Accuracy comparison across input subsets.

5.3.2 Sample Inference Results figure 23 shows per-pixel classification results for a selected image patch. It includes the input image, ground truth, predicted output, classification report, and confusion matrix.

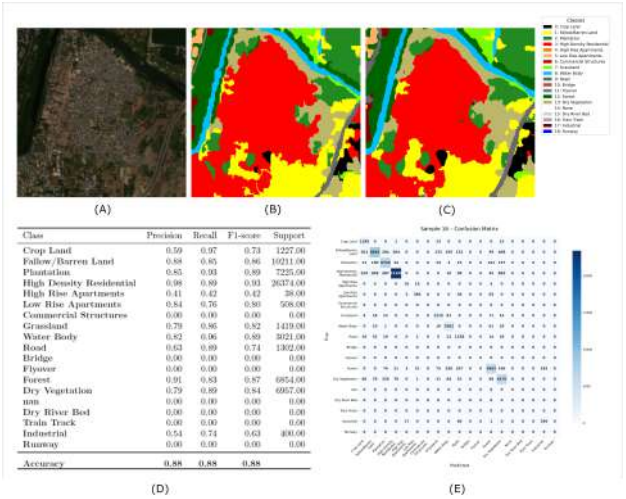


Figure 23. Classification results for sample 16.

5.3.3 Final Classified Map The full land cover map for the study area, generated using SegFormer with Spatial Attention and all 13 bands, is shown in Figure 24. The output demonstrates higher spatial consistency and class separation than the Random Forest output. It was produced at an accuracy of **88%**.

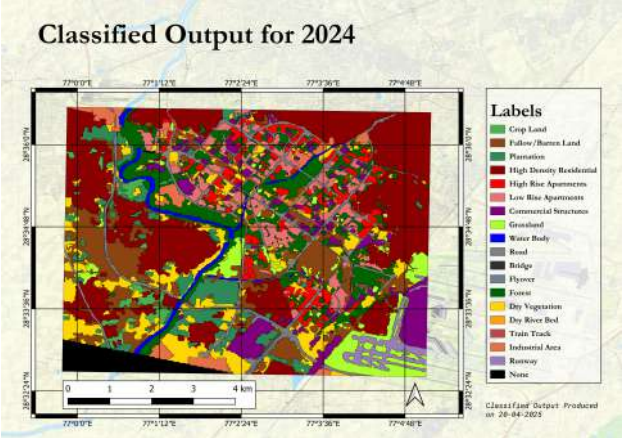


Figure 24. Final land cover map (2024) using SegFormer with Spatial Attention.

References

Anderson, P. R. J., 1971. Land-use classification schemes-used in selected recent geographic applications of remote sensing. Technical report, University of Florida.

Chen, H., He, Y., Zhang, L., Yang, W., Liu, Y., Gao, B., Zhang, Q., Lu, J., 2024. A Multi-Input Channel U-Net Landslide Detection Method Fusing SAR Multisource Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17.

Gaur, S., Das, N., Bhattacharjee, R., Ohri, A., Patra, D., 2023. A novel band selection architecture to propose a built-up index for hyperspectral sensor PRISMA. *Earth Science Informatics*, 16.

Han, H., Feng, Z., Du, W., Guo, S., Wang, P., Xu, T., 2024. Remote Sensing Image Classification Based on Multi-Spectral Cross-Sensor Super-Resolution Combined With Texture Features: A Case Study in the Liaohe Planting Area. *IEEE Access*, 12.

Huang, C., Davis, L. S., Townshend, J. R., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23.

Hütt, C., Waldhoff, G., Bareth, G., 2020. Fusion of sentinel-1 with official topographic and cadastral geodata for crop-type enriched LULC mapping using FOSS and open data. *ISPRS International Journal of Geo-Information*, 9.

Kaur, R., Pandey, P., 2022. A review on spectral indices for built-up area extraction using remote sensing technology. *Arabian Journal of Geosciences*, 15.

Kebede, T. A., Hailu, B. T., Suryabhagavan, K. V., 2022. Evaluation of spectral built-up indices for impervious surface extraction using Sentinel-2A MSI imageries: A case of Addis Ababa city, Ethiopia. *Environmental Challenges*, 8.

Parmegiani, N., Posolieri, M., 2002. DEM Data Processing for a Landscape Archaeology Analysis (Lake Sevan-Armenia). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIV.

Pelizari, P. A., Spröhnle, K., Geiß, C., Schoepfer, E., Plank, S., Taubenböck, H., 2018. Multi-sensor feature fusion for very high spatial resolution built-up area extraction in temporary settlements. *Remote Sensing of Environment*, 209, 793-807.

Richards, J. A., Jia, X., 2006. *Remote sensing digital image analysis: An introduction*.

Thurmond, A. K., Abdelsalam, M. G., Thurmond, J. B., 2006. Optical-radar-DEM remote sensing data integration for geological mapping in the Afar Depression, Ethiopia. *Journal of African Earth Sciences*, 44.

Uddin, M. P., Mamun, M. A., Hossain, M. A., 2021. Pca-based feature reduction for hyperspectral remote sensing image classification.

Valdivieso-Ros, C., Alonso-Sarría, F., Gomariz-Castillo, F., 2023. Impact of segmentation algorithms on multisensor LULC classification in a semiarid Mediterranean area. *Earth Science Informatics*, 16.

Wang, X., Xu, M., Xiong, X., Ning, C., 2020. Remote Sensing Scene Classification Using Heterogeneous Feature Extraction and Multi-Level Fusion. *IEEE Access*, 8.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., Luo, P., 2021. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 15.