BD OPEN LULC MAP: HIGH-RESOLUTION LAND USE LAND COVER MAPPING & BENCHMARKING FOR URBAN DEVELOPMENT IN DHAKA, BANGLADESH

Mir Sazzat Hossain   *Rakibul Hasan Rajib*   *Md Akil Raihan Iftee*   *Ovi Paul*
Abu Bakar Siddik Nayem   *Anis Sarker*   *Md. Ashraful Amin*   *Amin Ahsan Ali*
*AKM Mahbubur Rahman**

Center for Computational & Data Sciences, Independent University, Bangladesh

ABSTRACT

Land Use Land Cover (LULC) mapping using deep learning significantly enhances the reliability of LULC classification, aiding in understanding geography, socioeconomic conditions, poverty levels, and urban sprawl. However, the scarcity of annotated satellite data, especially in South/East Asian developing countries, poses a major challenge due to limited funding, diverse infrastructures, and dense populations. In this work, we introduce the **BD Open LULC Map (BOLM)**, providing pixel-wise LULC annotations across eleven classes (e.g., Farmland, Water, Forest, Urban Structure, Rural Built-Up) for Dhaka metropolitan city and its surroundings using high-resolution Bing satellite imagery (2.22 m/pixel). BOLM spans 4,392 km^2 (891 million pixels), with ground truth validated through a three-stage process involving GIS experts. We benchmark LULC segmentation using DeepLab V3+ across five major classes and compare performance on Bing and Sentinel-2A imagery. BOLM aims to support reliable deep models and domain adaptation tasks, addressing critical LULC dataset gaps in South/East Asia.

Index Terms— Land Use Land Cover, Satellite Imagery, Deep Learning, BD Open LULC Map, Segmentation

1. INTRODUCTION

Land use/land cover (LULC) changes have gained significant attention due to rapid global ecosystem transformations [1]. Developing countries in South and East Asia are experiencing substantial LULC changes. Grasslands, woodlands, bushlands, and other vegetation covers are being extensively converted into agricultural and settlement areas to accommodate growing populations. Land cover refers to the physical materials covering the earth's surface, such as forests, mountains, deserts, and water, while land use describes how humans utilize land for socio-economic activities like farming and urban development. Monitoring LULC changes over time helps understand geographical, socio-economic conditions, poverty levels, and urban sprawl. Recently, deep learning methods
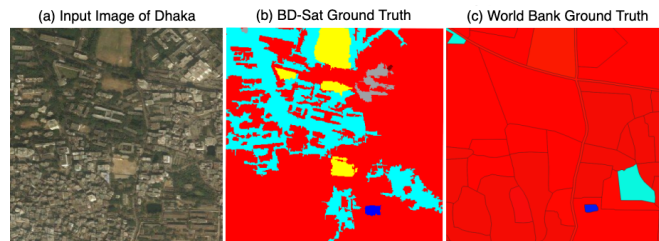

Fig. 1: Comparison between World Bank and BOLM(ours) Ground Truth (Here red color represents various kinds of built-up, cyan represents forest/tree cover, yellow represents bare land/meadow and blue represents water)

have effectively extracted LULC information from satellite images [2]. In this paper, we introduce the BD Open LULC Map (BOLM) to address the lack of LULC datasets for developing countries.

To effectively apply deep learning methods for LULC in developing countries with similar topographies (Bangladesh, Thailand, Indonesia, Malaysia, India, Sri Lanka), three major challenges must be addressed. First, there is a scarcity of LULC-annotated data. Large volumes of satellite images with pixel-level LULC annotations are essential for training deep models. Annotated datasets from developed countries [3, 4, 5] often produce erroneous results when models are applied to developing countries due to differences in building materials, land characteristics, and cultural contexts. Furthermore, UAV data collection, like in OpenEarthMap [6], is costly and limited in developing countries. Thus, high-resolution (HR) satellite LULC annotations are critically needed, especially for Bangladesh.

Second, as shown in Table 1, most datasets from this region feature low-resolution images (10m, 15m, 30m), which are inadequate for densely populated, unstructured urban areas. Low-resolution data and patch-based annotations induce errors in deep learning models. For example, datasets from India [7] lack sufficient resolution and detail for accurate pixel-level LULC mapping. The World Bank's urban mapping of Dhaka [11] lacks the necessary details for precise annotations.

---

*Corresponding author. Email: akmmrahman@iub.edu.bd

Table 1: Summary of the LULC satellite datasets from developing countries

Dataset Name	Region	Source	Acquisition Time	GSD (Per Pixel)	Area (Sq KM)
IndiaSat[7], 2021	India	Landsat-7, Landsat-8 and Sentinel-2	2016-2019	10, 15 m	162
OpenEarthMap[6], 2023	Africa, Asia, Australia, Europe, USA South America, South Asia (Dhaka, Cox's Bazar, Nepal Only)	UAV	2020	0.25m –0.5m	799
DynamicEarthNet[8], 2022	Africa, Australia, Europe, USA South America, China, South Asia (India), East Asia (Thailand)	Sentinel 1 & 2	2020-2021	3m - 10m	707
LandCoverNet[9], 2020	Africa, Asia, Australia, Europe,North America and South America	Sentinel-2	2018	10 m	12,905
Mauro et al.[10], 2020	Hanoi, Vietnam	Landsat-5,7,8	1989-2019	30 m	3,050
World bank [11], 2021	Dhaka	QuickBird, Pleiades, Landsat and Sentinel-2	2006,2017	5m-30m	1,034
BOLM(Ours)	**Dhaka**	**Bing**	**2019**	**2.22 m**	**4,392**

Third, many deep learning methods rely solely on RGB channels achieving good performance compared to traditional approaches [12]. However, HR RGB data is not frequently updated in developing countries [13]. Conversely, satellite data (e.g., Sentinel-2/3, Landsat) are regularly updated and offer multi-spectral channels, which can be leveraged using indices like NDVI, NDWI, and RVI [14]. Despite their lower resolution, these channels can enhance LULC mapping when combined with DL models. The key contributions of this paper are:

- Introduction of the BD Open LULC Map (BOLM), a high-resolution, pixel-by-pixel annotated dataset covering 4,392 sq km of Dhaka's urban and rural areas, using Bing satellite imagery (2.22 m/pixel).

- Benchmarking of state-of-the-art deep learning algorithms (DeepLabV3+ [15]) on both low-resolution (Sentinel-2A) and high-resolution (Bing) data.

- Comprehensive analysis of model performance using different Sentinel-2A channels and indices (NDVI, NDWI, RVI), and comparison with HR data results.

- Conclusive recommendations on the applicability of BOLM dataset for developing countries in South and East Asia, for both Sentinel and Bing/Google imagery.

2. EXISTING DATASETS AND LIMITATIONS

Several high-quality LULC datasets exist, mainly from developed countries. The Potsdam and Vaihingen datasets [3] cover German regions with five classes: building, low-vegetation, tree, car, and background. The Gaofen (GID) dataset [5] spans China with classes like forest, built-up, farmland, meadow, and water. MiniFrance [4] provide urban and rural scene annotations for China and France. FloodNet [16] includes ten LULC classes related to floods in Texas and Louisiana.

In developing countries, India's IndiaSat [7] use Sentinel and Landsat imagery for land cover classification. Africa's Mpologoma Catchment dataset [17] (Uganda) and Libokemkem district dataset [2] (Ethiopia) cover land changes over time. LandCoverNet [9] offers global LULC data, while a Hanoi study [10] tracks urbanization from 1989–2019. High-quality, pixel-wise annotated LULC data is scarce in developing regions, and existing datasets (Table 1) of Developing Countries have several limitations:

Most datasets use low-resolution imagery (10m–30m), making it hard to distinguish small urban features like houses, roads, and vegetation. Unlike structured cities, developing regions lack clear zoning, complicating land cover classification.

Deep learning models trained on one region often fail in others due to geographic and cultural variations [18]. Differences in building materials, landscapes, and urban structures cause domain shifts, requiring high-resolution ($\sim 2m$) datasets with precise pixel-wise annotations.

LULC datasets from Bangladesh have coarse annotations, leading to misclassified pixels and degraded model performance. For instance, the World Bank's urban mapping of Dhaka merges key categories (e.g., construction sites, airports, industrial areas), covers only urban regions, and is not publicly available. Figure 1 shows how finer annotations improve classification accuracy.

3. DATASET CREATION

The study focuses on the Dhaka division, centered around Dhaka city, covering a 117 km × 124 km region from N24°23'04", E89°57'26" (Upper Left Corner) to N23°22'13", E91°00'26" (Lower Right).

3.1. Sources of Satellite Data

This study utilizes two image sources: Bing imagery (2.22m/pixel) for ground truth annotation and high-resolution segmentation, and Sentinel-2A data for experiments using various band combinations and index images as inputs.

Bing RGB: Bing provides high-resolution aerial imagery (2.22m/pixel) at 17 zoom levels [19]. Acquired on April 20, 2019, the dataset has a 48906×47256 resolution in TIFF (RGB) format. Figure 2 shows the full image, with the blue rectangle marking the annotated area (Dhaka and surroundings) and the red rectangle highlighting the test region within Dhaka city.
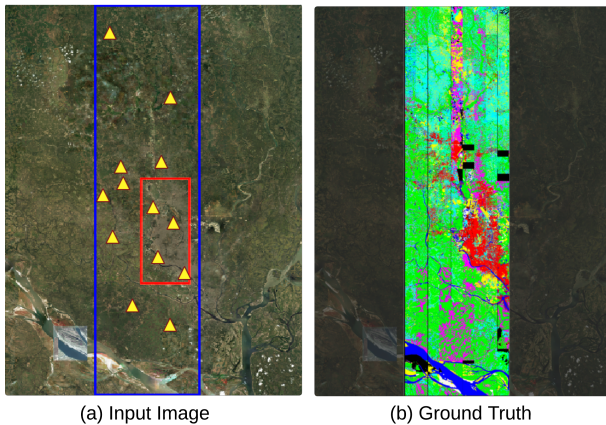

(a) Input Image   (b) Ground Truth

Fig. 2: (a) is the input image from the bing used and (b) is the corresponding ground truth.

Sentinel-2A: Sentinel-2A data, acquired on April 11, 2019, at 04:38:29 UTC (Tile ID: "L1C_T45QZG_A016175 _20190411T043829"), includes 12 spectral bands with resolutions from 10m to 60m per pixel. The red (band 2), green (band 3), blue (band 4), and NIR (band 8) have 10m resolution, while Coastal Aerosol (band 1), Water Vapor (band 9), and SWIR Cirrus (band 10) have 60m. NIR and SWIR bands are at 20m resolution.

SWIR combinations help analyze moisture and distinguish minerals. False Color Infrared (FCI) (NIR, red, green) aids vegetation detection, while Atmospheric Penetration (ATM) (two SWIR, one NIR) enhances classification clarity [20]. NDVI assesses vegetation health, RVI estimates vegetation water content, and NDWI detects water stress in plants, with wetter vegetation yielding higher values.

3.2. Ground Truth Annotation Process

The Bing image (2.22m/pixel) covers 4,392 square kilometers in Dhaka, totaling 891 million pixels. This area was annotated into eleven LULC classes (see Table 4 in the supplementary material).

The large Bing image (48,906 × 47,256 pixels) was divided into 1,089 sub-images (1,500 × 1,500 pixels). The study focused on folders 8 to 19, each containing 33 images. A

Fig. 3: Various Regions on the Dhaka Division

total of 24 annotators (two per folder) used eCognition software and a rule-based approach to classify polygons. Multi-resolution segmentation was applied, merging adjacent polygons of the same class before exporting the final ground truth. Figures 8 and 9 illustrate the process (see supplementary material).

To ensure accuracy, training sessions covered eCognition usage, rule-set creation, and segmentation. Experts clarified class distinctions, and a tutorial video was provided for reference (footnote 1).

3.3. Annotation Validation and Data Reliability

To ensure annotation reliability, we followed a three-stage validation process. In Stage 1, two annotator groups independently labeled the dataset, with a pixel-wise agreement matrix highlighting in Table 4 (details in the supplementary material) disagreements in all classes except Brick and Road. The highest agreement was observed for the Forest class (95.24%), while the Road class had the lowest (88.18%). Stage 2 focused on refining disagreement regions (200 × 200 pixels) through manual corrections, significantly reducing inconsistencies. Finally, Stage 3 involved GIS experts validating the dataset, achieving over 99% agreement for most classes. The expert-labeled pixels were adopted as the final ground truth, ensuring high-quality annotations. All details are in the supplementary material.

3.4. In Situ Visual Assessment

To enhance annotation reliability, an on-site validation was conducted in the Dhaka Division. The team visited 13 diverse locations, covering urban, rural, agricultural, industrial, and natural landscapes. Using a high-resolution camera, Gaia GPS, and field notes, they compared dataset annotations with real-world observations.

Land cover characteristics—including infrastructure, vegetation, water bodies, and farmland—were documented with photographs and GPS coordinates for precise alignment. Figure 2 shows the visited locations, while Figure 3 presents sample images illustrating LULC variations.

---

[1] Annotation tutorial video: `https://www.youtube.com/watch?v=psQwvRDxuTo`

(a) Area Percentage of 11 Class
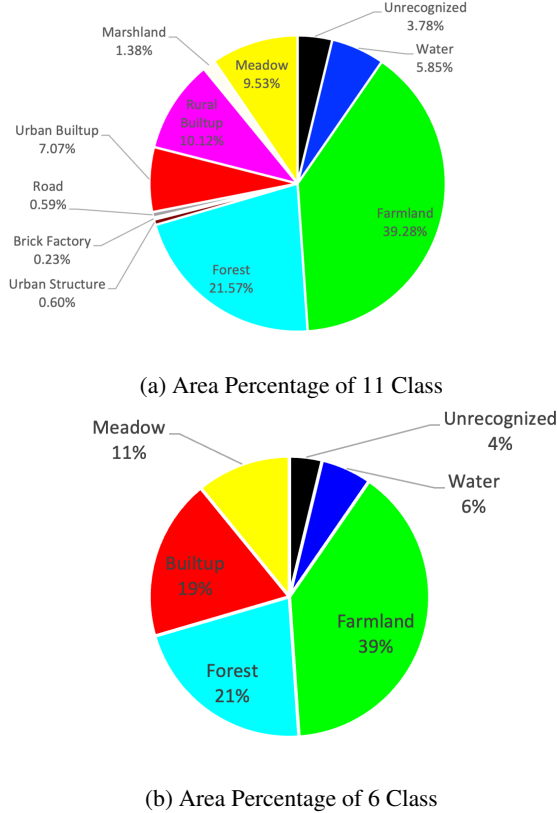

(b) Area Percentage of 6 Class

Fig. 4: Area Percentage of the 11 class and 6 class dataset.

3.5. Conversion from 11 Classes to 6 Classes

Due to class imbalance (Figure 4a), we merged eleven classes into six broader categories (Table 4) for better balance in experiments. Farmland, Water, and Forest remained unchanged, while Urban Structure, Rural Built-up, Urban Built-up, and Brick Factory were merged into Built-up. Meadow and Marshland were combined into Meadow. The final class distribution (Figure 4b) is: Farmland (39%), Water (6%), Forest (21%), Built-up (19%), Meadow (11%), and Unrecognized (3.78%).

4. IMAGE PROCESSING FOR SATELLITE DATA

Sentinel-2A Index and Combination Images: Sentinel-2A provides 12 spectral bands, which we combine for segmentation tasks. Since different channels capture diverse land textures, we utilize multiple combinations from Sentinel-2A data. Index images are crucial for LULC classification. The RGB representations of these index and combination images are shown in Figure 5. Given the varying resolutions of Sentinel-2A channels, we upsampled them to 10m/pixel using bilinear interpolation. The 3-channel index and combination images, listed in Table 2, were generated using QGIS.

Bing RGB Image: The Bing image, captured at a 17-zoom

Table 2: Sentinel-2 Index and Combination RGB Channel Formation

Image Type	Channel-1	Channel-2	Channel-3
RGB	B4(Red)	B3(Green)	B2(Red)
ATM	B12(SWIR)	B11(SWIR)	B8(NIR)
FCI	B8(NIR)	B4(Red)	B3(G)
SWI	B12(SWIR)	B8(NIR)	B4(R)
NDVI	NDVI	NDVI	NDVI
NDWI	NDWI	NDWI	NDWI
RVI	RVI	RVI	RVI

(a)RGB (b)Atmospheric penetration (c)False Color Infrared (d)Short Wave Infrared
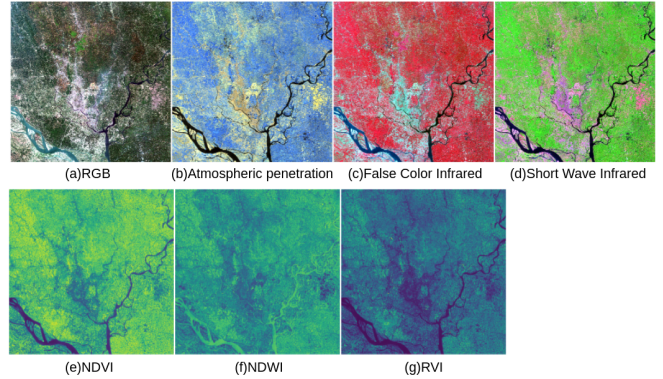
(e)NDVI (f)NDWI (g)RVI

Fig. 5: Combination and Index images generated from Sentinel-2 multispectral bands

level (2.22m/pixel), serves as a high-quality ground truth for accurate LULC segmentation. Its high resolution enables precise annotation and improves model performance. However, the large data size increases processing and training time. Despite this challenge, its clarity benefits both manual annotation and deep learning-based segmentation

5. EXPERIMENTAL SECTION

Implementation details: For benchmarking, we utilized DeepLabV3+ [15], an advanced version of DeepLabV3 with an encoder-decoder structure and atrous convolution for improved semantic segmentation. The input and ground truth images are split into 513×513 patches for DeepLab V3+ training. Some ground truth regions are difficult to label and remain unrecognized, with unrecognized pixels assigned black in both input and ground truth. Training uses a sliding window with 75% overlap in both directions, while testing uses non-overlapping patches. The Dhaka city region (red rectangle) served as the test set, while surrounding areas (blue rectangle) were used for training showed in figure 2 . We computed IoU and F1 scores from confusion matrices based on model predictions and ground truth for test image pixels. Accuracy was not used due to potential bias from imbalanced LULC class samples. The experimental settings are as follows:

LULC with Index Images: Three-channel input images are created using NDVI, NDWI, and RVI. Training images are generated with a 513×513 sliding window, overlapping 75% in both directions. The batch size is 8, and training runs for 25 epochs. Testing uses non-overlapping patches.

LULC with Channel-Combined Images: The same sliding window approach is used for training, with a batch size of 8 and 25 epochs, using DeepLabV3+. Testing is performed without overlapping.

LULC with Full Bing RGB: The full Bing image is used, following the same training procedure as above, with DeepLabV3+ and no overlap during testing.

6. RESULTS

6.1. Index images

Among three index images, NDWI achieves the highest IoU (0.20) and F1 score (0.32) for forests and performs best for water (IoU: 0.24, F1: 0.51). NDVI outperforms for built-up and farmland, while RVI has a slightly better F1 score for meadow. On average, NDVI provides the best segmentation for five classes (Avg IoU: 0.30, Avg F1: 0.45). Built-up, farmland are segmented with high precision using index images.

Table 3: Scores for 5 classes of all the image types.

	Metric	Forest	Built-Up	Water	Farmland	Meadow	Avg
NDVI	IoU	0.17	**0.52**	**0.68**	**0.35**	0.51	**0.30**
	F1	0.17	**0.68**	0.27	**0.51**	0.50	**0.45**
NDWI	IoU	**0.20**	0.38	0.24	0.34	0.28	0.28
	F1	**0.32**	0.55	**0.38**	0.51	0.50	0.43
RVI	IoU	0.13	0.50	0.17	0.31	0.36	0.29
	F1	0.22	0.66	0.29	0.47	**0.52**	0.43
ATM	IoU	0.17	**0.56**	0.29	**0.48**	0.19	0.33
	F1	0.29	**0.72**	0.45	**0.64**	0.31	0.48
FCI	IoU	0.20	0.53	0.30	0.46	0.17	0.33
	F1	0.33	0.69	0.45	0.62	0.29	0.47
SWI	IoU	**0.22**	0.56	**0.32**	0.46	0.22	**0.35**
	F1	**0.35**	0.71	**0.48**	0.62	0.37	**0.50**
SentinelRGB	IoU	0.18	0.54	0.30	0.46	**0.25**	0.34
	F1	0.30	0.70	0.46	0.63	**0.40**	0.49
BingRGB	IoU	0.33	0.58	0.48	0.57	0.26	0.44
	F1	0.49	0.73	0.65	0.72	0.41	0.60

For the farmland class, NDVI, RVI both achieved an F1 score of 0.79, but NDVI had a higher IoU (0.35). As NDWI detects water/moisture content, its superior water detection (IoU: 0.24) was expected. NDVI performed best for built-up areas (IoU: 0.52, F1: 0.68), while RVI was best for meadow (IoU: 0.52, F1: 0.68). NDVI and RVI had the same average F1 score, but NDVI had the highest average IoU (0.30).

6.2. Channel Combination Images

We analyze the results of combined images for Sentinel-2A from Table 3. SWI outperforms ATM in Forest segmentation with an IoU of 0.22. For Built-up, both ATM and SWI achieve

an IoU of 0.56, but SWI attains higher accuracy. Water regions are better detected with SWI, while ATM performs best for Farmland IoU. Meadow segmentation achieves the highest IoU (0.22) with SWI. On average, SWI attains the highest IoU (0.35) and F1 score (0.50) across all classes.

SentinelRGB exhibits slightly lower performance than SWI. Built-up and Farmland are accurately segmented across all compositions, with ATM achieving the best IoU (0.56) and F1 (0.72) for Built-up and IoU (0.48), F1 (0.64) for Farmland. Forest and Water segmentation is superior with SWI, while Meadow achieves IoU (0.25) and F1 (0.40) with SentinelRGB. Overall, SWI achieves the best average IoU and F1 score among four compositions and three index images derived from Sentinel-2A data.

SWI scores the highest in Forest (0.22 IoU). Built-up IoU is identical for ATM and SWI. SWI leads in Water segmentation, ATM in Farmland, and RVI in Meadow. Combination images generally outperform index images due to richer information across three channels. Though SentinelRGB is widely used for LULC, it does not surpass most combination images. Index images, being single-channel, contain less information, making them less effective than combination images.

6.3. Bing RGB Image

We describe the results from BingRGB, where the experiment includes training and testing using the full RGB image captured by the Bing satellite. The model with a full Bing RGB image outperforms the rest by significant margins. The average IoU is 0.44, and the average F1 score is 0.60. Not only in averages, but BingRGB also segments all five LULC classes more accurately compared to other models.

7. CONCLUSION

This paper presents high-resolution pixel-wise Land Use Land Cover (LULC) ground truth annotations for the Dhaka metropolitan region, covering 4,392 square kilometers. The BOLM dataset includes eleven LULC classes across urban and rural areas. Baseline experiments using Sentinel 2A and Bing satellite imagery for five major LULC classes show that Bing's high-resolution imagery (2.22m/pixel) consistently outperforms Sentinel 2A. However, Bing is limited to RGB channels and lacks free monthly availability in developing regions. Sentinel 2A, with multi-channel data including SWIR, remains valuable when high-resolution imagery is unavailable. Our findings highlight the trade-off between precision and accessibility, suggesting Bing/Google/ESA imagery for urban monitoring and Sentinel 2A for periodic applications like crop monitoring.

8. LICENSING AND ETHICAL CONSIDERATIONS

This dataset has been made available online here under the Creative Common (CC) license: `https://doi.org/10.7910/DVN/LLR3RR`. Custom codes were developed to process the dataset, including converting RGB images into index format for the model. To improve performance and reduce confusion, unknown label regions are projected onto the input image. If accepted, the code link will be provided.

9. REFERENCES

[1] Gebeyehu Abebe, Dodge Getachew, and Alelgn Ewunetu, "Analysing land use/land cover changes and its dynamics using remote sensing and gis in gubalafito district, northeastern ethiopia," *SN Applied Sciences*, vol. 4, no. 1, pp. 1–15, 2022.

[2] Fikirte Demissie, Kumelachew Yeshitila, Mengistie Kindu, and Thomas Schneider, "Land use/land cover changes and their causes in libokemkem district of south gonder, ethiopia," *Remote Sensing Applications: Society and Environment*, vol. 8, pp. 224–230, 2017.

[3] Haifeng Li, Kaijian Qiu, Li Chen, Xiaoming Mei, Liang Hong, and Chao Tao, "Scattnet: Semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 5, pp. 905–909, 2021.

[4] Javiera Castillo Navarro, Bertrand Le Saux, Alexandre Boulch, Nicolas Audebert, and Sébastien Lefèvre, "Minifrance," 2020.

[5] Xin-Yi Tong, Gui-Song Xia, Qikai Lu, Huanfeng Shen, Shengyang Li, Shucheng You, and Liangpei Zhang, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sensing of Environment*, vol. 237, pp. 111322, 2020.

[6] Junshi Xia, Naoto Yokoya, Bruno Adriano, and Clifford Broni-Bediako, "Openearthmap: A benchmark dataset for global high-resolution land cover mapping," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 6254–6264.

[7] Chahat Bansal, Hari Om Ahlawat, Mayank Jain, Om Prakash, Shivani A Mehta, Deepanshu Singh, Harshavardhansushil Baheti, Suyash Singh, and Aaditeshwar Seth, "Indiasat: A pixel-level dataset for landcover classification on three satellite systems - landsat-7, landsat-8, and sentinel-2," in *Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies*, New York, NY, USA, 2021, COMPASS '21, p. 147–155, Association for Computing Machinery.

[8] Aysim Toker, Lukas Kondmann, Mark Weber, Marvin Eisenberger, Andrés Camero, Jingliang Hu, Ariadna Pregel Hoderlein, Çağlar Şenaras, Timothy Davis, Daniel Cremers, et al., "Dynamicearthnet: Daily multispectral satellite dataset for semantic change segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21158–21167.

[9] Hamed Alemohammad and Kevin Booth, "Landcovernet: A global benchmark land cover classification training dataset," *arXiv preprint arXiv:2012.03111*, 2020.

[10] Giovanni Mauro, "Rural–urban transition of hanoi (vietnam): Using landsat imagery to map its recent periurbanization," *ISPRS International Journal of Geo-Information*, vol. 9, no. 11, 2020.

[11] "Dhaka (bangladesh) - land use / land cover maps (esa eo4sd-urban) - eo4sd-urban dhaka operations report," October 2021.

[12] Selçuk Reis, "Analyzing land use/land cover changes using remote sensing and gis in rize, north-east turkey," *Sensors*, vol. 8, no. 10, pp. 6188–6202, 2008.

[13] "Bing maps," Apr. 2022, Page Version ID: 1084073274.

[14] Maie Elgammal, Rafat Ali, and Rasha Samra, "Ndvi threshold classification for detecting vegetation cover in damietta governorate," *Journal of American Science*, vol. 10, pp. 108–113, 2014.

[15] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," 2018.

[16] Maryam Rahnemoonfar, Tashnim Chowdhury, Argho Sarkar, Debvrat Varshney, Masoud Yari, and Robin Roberson Murphy, "Floodnet: A high resolution aerial imagery dataset for post flood scene understanding," *IEEE Access*, vol. 9, pp. 89644–89654, 2021.

[17] Jackson Bunyangha, Mwanjalolo JG Majaliwa, Agnes W Muthumbi, Nathan N Gichuki, and Anthony Egeru, "Past and future land use/land cover changes from multi-temporal landsat imagery in mpologoma catchment, eastern uganda," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 24, no. 3, pp. 675–685, 2021.

[18] A. K. M. Mahbubur Rahman, Moinul Zaber, Qianwei Cheng, Abu Bakar Siddik Nayem, Anis Sarker, Ovi Paul, and Ryosuke Shibasaki, "Applying state-of-the-art deep-learning methods to classify urban cities of the developing world," *Sensors*, vol. 21, no. 22, 2021.

[19] Rbrundritt, "Bing maps tile system - bing maps," .

[20] J. H. Everitt, C. Yang, R. S. Fletcher, M. R. Davis, and D. L. Drawe, "Using aerial color-infrared photography and QuickBird satellite imagery for mapping wetland vegetation," *Geocarto International*, vol. 19, no. 4, pp. 15–22, 2004.

Supplementary Materials

9.1. Analysis of Normalized Confusion Matrix

The normalized confusion matrices in Figures 11a–12 illustrate the performance of DeeplabV3+ on different datasets. BingRGB achieves the highest accuracy across all classes, with an average IoU of 0.44 and F1 of 0.60. The model struggles with inter-class confusions, particularly between Forest and Built-up areas, due to their proximity in urban settings. Water is often misclassified as Farmland due to similar colors in urban environments, and Meadow remains the most challenging class.

SentinelRGB achieves high accuracy for Built-up (84%) but performs poorly in Forest (25%) and Farmland (53%) due to lower resolution. Sentinel-2 ATM improves Farmland and Water classification but struggles with Forest and Built-up, likely due to its spectral properties emphasizing atmospheric penetration. SWI achieves overall better performance than RGB but still struggles with Meadow misclassification due to resolution differences. FCI enhances Forest, Built-up, and Water segmentation but performs worst for Meadow, misclassifying it as Built-up due to spectral similarities in dry land areas.

Overall, combination images provide richer information, outperforming single-channel index images, while BingRGB delivers the highest segmentation accuracy.

9.2. Annotation Validation and Data Reliability

To ensure the reliability of the annotation, we have followed a three-stage validation process.

In stage 1, two groups of annotators independently annotated the dataset at the pixel level. A pixel-wise agreement matrix was calculated to compare their annotations (Table 4). Disagreements were observed in all classes except the Brick and Road class. Notably, the Water class showed confusion with Farmland (3.4%) and Marshland (5.66%), as floating weeds in ponds and lakes often caused misclassification. The Farmland class also exhibited overlap with Roads, Rural Built-up areas, Marshland, and Meadow, particularly when crops were young or harvested.
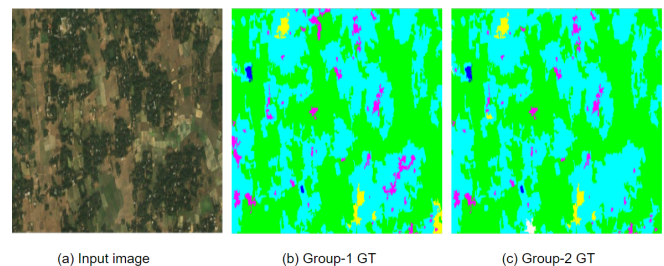

(a) Input image   (b) Group-1 GT   (c) Group-2 GT

Fig. 6: Disagreement in annotation between Group 1 and Group 2 members. Some of the water areas have been missed by Group 2

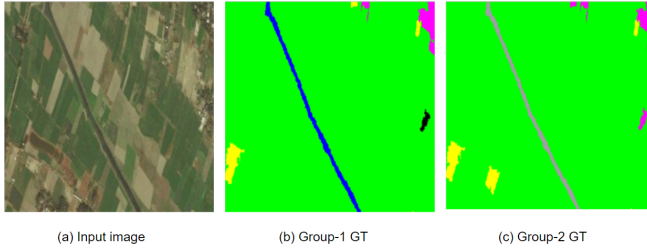| (a) Input image | (b) Group-1 GT | (c) Group-2 GT |

Fig. 7: Disagreement in annotation between Group 1 and Group 2 members. One of the group 1 member annotated the long thin region as water (blue), whereas one group 2 member marked it as road. In Bangladesh, there are a lot of places where roads or rivers pass through vast farmland areas.

For the Forest class, minor confusion occurred mainly with Rural Built-up areas, as trees often surround houses in rural regions. This resulted in a 3.93% disagreement. The highest agreement was observed for the Forest class (95.24%), while the Road class had the lowest agreement (88.18%). Figures 6 and 7 illustrate these disagreement issues. Many discrepancies were resolved in the second stage, with expert verification finalizing the annotations.

In stage 2, disagreement regions from stage 1 were isolated for detailed examination. Special attention was given to these regions (200 × 200 pixels), where annotator groups 1 and 2 disagreed. Each case was reviewed in discussions with the respective annotators. Using Photoshop on an iMac (24-inch screen), annotators manually refined the overlapping class boundaries at the pixel level.

Tables 3 and 4 of the supplementary material show the revised agreement percentages. After stage 2, the disagreement percentage was significantly reduced, with the highest remaining disagreement at 4.2%.

In stage 3, we sought validation from GIS experts with extensive experience in LULC annotations. Initially, the agreed-upon portions of the dataset—where both annotator groups reached a consensus—were selected. To further verify accuracy, 10% of each class's agreed portion was given to the expert group for evaluation. After expert annotation, comparisons showed that all classes except the Urban Structure class had above 99% agreement, with Urban Structure at 98.585%. Given this high level of agreement, the entire agreed portion from groups 1 and 2 was accepted without further verification. Table 5 of the supplementary material details the agreement percentages between the annotator groups and experts.

Next, the areas of disagreement between group 1 and group 2 were fully annotated by the experts. Table 6 of the supplementary material presents the agreement/disagreement statistics between annotator disagreements and expert annotations. The expert-labeled pixels from Table 6 were adopted as final ground truth annotations. This stage marked the completion of the dataset, with the distribution of pixels across eleven classes illustrated in Figure 4a.
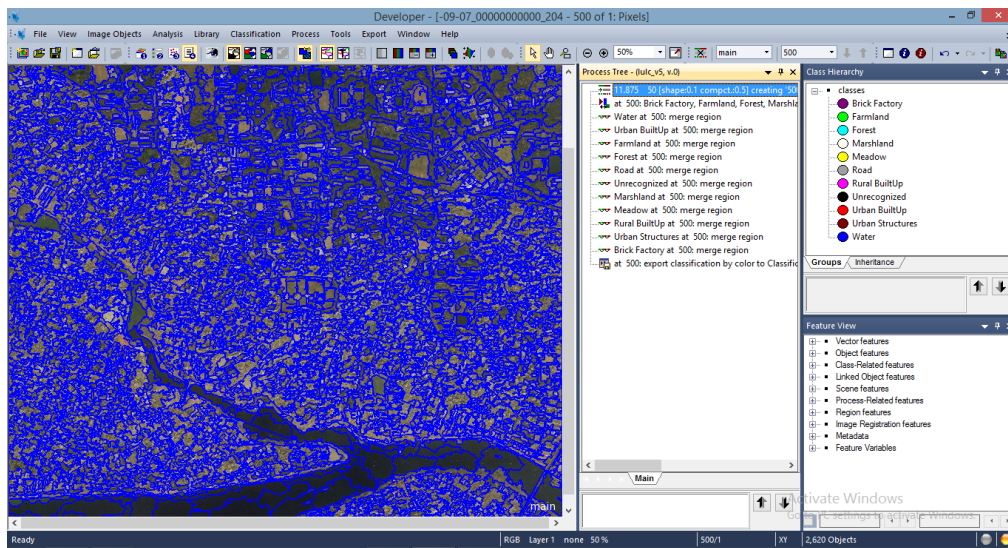
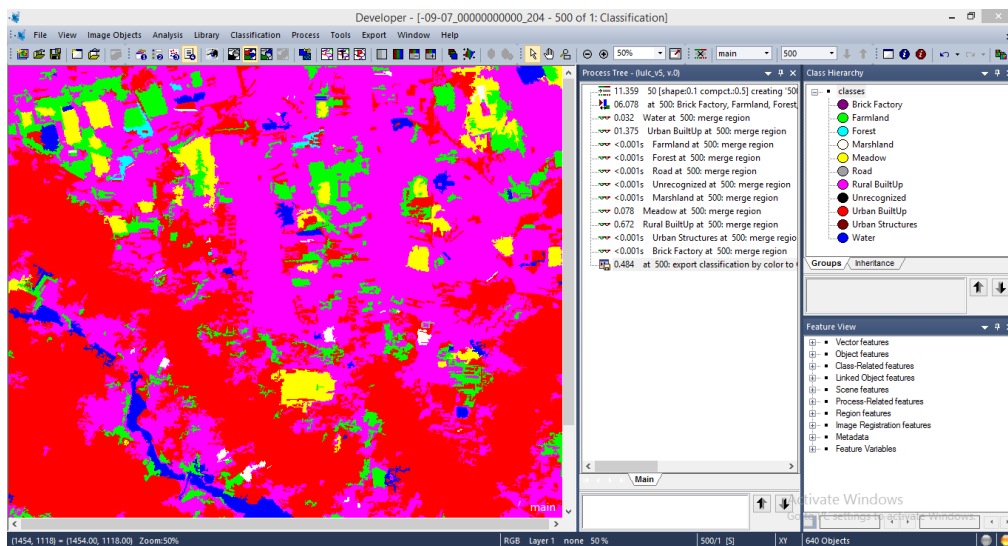**Fig. 8**: Segmenting image into small polygons
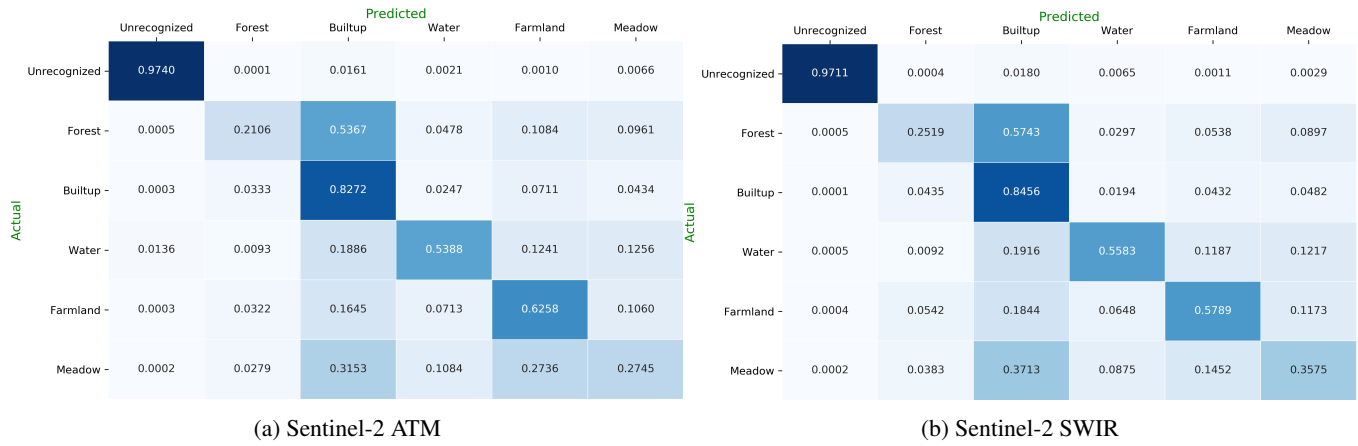

Fig. 9: Exporting after fully labeling the ground truth

	Predicted					
	Unrecognized	Forest	Builtup	Water	Farmland	Meadow
Unrecognized	0.9740	0.0001	0.0161	0.0021	0.0010	0.0066
Forest	0.0005	0.2106	0.5367	0.0478	0.1084	0.0961
Builtup	0.0003	0.0333	0.8272	0.0247	0.0711	0.0434
Water	0.0136	0.0093	0.1886	0.5388	0.1241	0.1256
Farmland	0.0003	0.0322	0.1645	0.0713	0.6258	0.1060
Meadow	0.0002	0.0279	0.3153	0.1084	0.2736	0.2745

(a) Sentinel-2 ATM

	Predicted					
	Unrecognized	Forest	Builtup	Water	Farmland	Meadow
Unrecognized	0.9711	0.0004	0.0180	0.0065	0.0011	0.0029
Forest	0.0005	0.2519	0.5743	0.0297	0.0538	0.0897
Builtup	0.0001	0.0435	0.8456	0.0194	0.0432	0.0482
Water	0.0005	0.0092	0.1916	0.5583	0.1187	0.1217
Farmland	0.0004	0.0542	0.1844	0.0648	0.5789	0.1173
Meadow	0.0002	0.0383	0.3713	0.0875	0.1452	0.3575

(b) Sentinel-2 SWIR

Fig. 10: Normalized confusion matrix of Sentinel-2 ATM and SWIR combination images.

	Predicted					
	Unrecognized	Forest	Builtup	Water	Farmland	Meadow
Unrecognized	0.9946	0.0002	0.0045	0.0004	0.0001	0.0003
Forest	0.0002	0.5554	0.2940	0.0069	0.1112	0.0323
Builtup	0.0001	0.1307	0.7144	0.0052	0.0822	0.0674
Water	0.0004	0.0730	0.0544	0.5652	0.2601	0.0469
Farmland	0.0001	0.0576	0.0600	0.0128	0.7882	0.0813
Meadow	0.0001	0.0758	0.2089	0.0271	0.3308	0.3574

(a) Bing RGB

	Predicted					
	Unrecognized	Forest	Builtup	Water	Farmland	Meadow
Unrecognized	0.9685	0.0012	0.0179	0.0088	0.0009	0.0027
Forest	0.0004	0.2510	0.6032	0.0272	0.0353	0.0829
Builtup	0.0001	0.0450	0.8444	0.0176	0.0348	0.0581
Water	0.0001	0.0189	0.2066	0.5089	0.1139	0.1517
Farmland	0.0003	0.0719	0.1993	0.0665	0.5277	0.1344
Meadow	0.0002	0.0482	0.3692	0.0796	0.1240	0.3789

(b) Sentinel-2 RGB

Fig. 11: Normalized confusion matrix of Bing RGB and Sentinel-2 RGB images.

Table 4: Stage-3: Group-1 and 2 disagreements vs. Expert annotation in number of pixels

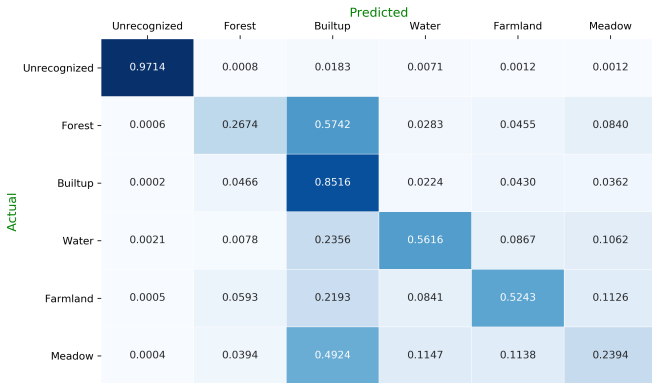| Total Pixels Selected for Validation (100% of 1 and 2 Disagree)			Expert										
			Unrecognized	Water	Farmland	Forest	Urban Structure	Brick Factory	Road	Urban Builtup	Rural Builtup	Marshland	Meadow
244,452	Both Group 1 and Group 2 Disagree	Unrecognized	243,437	37	49	244	24	0	0	122	244	49	244
1,058,694		Water	106	1,049,695	0	0	212	529	0	1,588	5,293	1,059	212
9,250,035		Farmland	925	1,850	9,172,335	32,375	9,250	13,875	0	0	3,700	11,100	4,625
1,901,067		Forest	380	190	6,844	1,878,254	3,802	951	2,852	0	4,753	2,471	570
169,327		Urban Structure	254	0	847	169	165,906	85	339	508	339	847	34
47,042		Brick Factory	0	235	0	0	24	46,680	0	56	47	0	0
185,121		Road	19	0	0	0	185	0	183,992	926	0	0	0
1,532,498		Urban Builtup	613	0	0	0	6,130	4,597	10,727	1,510,276	153	0	0
2,495,021		Rural Builtup	3,493	2,246	3,743	1,248	0	0	0	749	2,483,544	0	0
322,169		Marshland	32	1,289	967	644	0	0	0	0	0	319,108	129
3,149,345		Meadow	1,575	1,249	7,495	3,123	0	0	0	0	3,123	11,868	6,216,219

Fig. 12: Normalized confusion matrix of Sentinel-2 FCI image