Satellite image Land Cover Classification (LULC) monitoring Urban

1st Md Saib Hossain
dept. of Computer Science and Engineering
United International University, Dhaka.
mhossain221450@bscse.uiu.ac.bd
ID: 011221450

2nd Md.Maynul Hossen
dept. of Computer Science and Engineering
United International University Dhaka
mhossen222095@bscse.uiu.ac.bd
ID: 011222095

3rd Eila Afrin
dept. of Computer Science and Engineering
United International University Dhaka
eafrin221149@bscse.uiu.ac.bd
ID: 011221149

4th Saba Jasi
dept. of Computer Science and Engineering
United International University Dhaka
sjasi221005@bscse.uiu.ac.bd
ID: 011221005

***Abstract*—The uncontrolled growth of cities in developing countries is a major concern with respect to environmental sustainability. This study aims to assess the effects of urban expansion by examining the trends of Land Use and Land Cover (LULC) change in three important cities of Bangladesh: Dhaka, Khulna, and Rajshahi.The Google Earth Engine cloud platform is capable of processing multi-year time series and, as an example, we used it to extract Sentinel-2 satellite images covering the March 2017 to March 2024 period.For this case study, we carried out supervised classification using three powerful machine learning algorithms: Random Forest, Support Vector Machine and XGBoost. The landscape was classified into four primary categories: vegetation, water bodies, built up areas, and barren land. All the models were analyzed in a comparative way, and it was verified that the Random Forest classifier trained the best model achieving overall accuracy of 99.9% with a Kappa coefficient of 0.999. The analysis done reveals patterns of urbanization constantly replacing conveniently located natural and agricultural lands, leading to a decline in greenery and water bodies. The LULC maps along with change statistics prepared for this study provide a rich understanding that is beneficial for developing policies for sustainable urban development, resource management, and planning.**

***Index Terms*—land use land cover (LULC), remote sensing, machine learning, google earth engine, sentinel-2, urbanization, Bangladesh.**

I. Introduction

The rapid growth of cities is quite evident in the developing world, and this acceleration is mainly observed in the 21st century. Although this urban expansion serves as a driver for economic development, it is often accompanied by unplanned expansion that leads to significant environmental challenges such as the permanent destruction of agricultural land, ecosystem degradation, and reduced livability in cities [1]. Thus, effective urban policy depends on having geo-spatial information that is precise, scalable, current, and updated [2].

In terms of monitoring the Earth's surface, satellite remote sensing is very crucial. The high-resolution satellite missions, such as those from the European Space Agency, offer consistent multispectral data streams suitable for time-series analysis [3].The scope of remote sensing technology is superseded by cloud based processing systems like Google Earth Engine (GEE) which provides the public with extensive libraries of satellite images and promotes large-scale environmental assessments [4]. The use of GEE in conjunction with sophisticated artificial intelligence models has made it possible to classify complex terrains with great precision, which enables tracking of subtle but important changes over time in Land Use and Land Cover (LULC) [5].

Bangladesh is one of the most densely populated countries in the world which epitomizes the effects of rapid urbanization. The key metropolitan areas are Dhaka, Khulna and Rajshahi which are undergoing rapid expansion towards the south, depleting natural resources and overburdening infrastructure systems [6], [7]. Other studies have documented the critical agricultural and wetland regions being transformed into urban centers throughout the nation [8]. Still, there remains a literature gap on a comparative multi-city analysis focused on the latest high-resolution satellite images and an exhaustive comparison of top machine learning classifiers.

This research fills the gap by providing an integrated LULC analysis of Dhaka, Khulna, and Rajshahi for the years 2017–2024. Specifically, we aim to analyze Sentinel-2 images in conjunction with the Google Earth Engine (GEE) platform to quantify and evaluate the impacts of urban sprawl vis-a-vis environmental degradation. In addition, we propose a novel hybrid approach to classification that utilizes spectral indices (NDVI, MNDWI, NDBI) along with the original spectral bands to improve classification accuracy. This study applies Random Forest, SVM, and XGBoost classifiers and evaluates their accuracy for Bangladesh's sustainable urban development while also enhancing the methodology used for LULC mapping in developed, and densely populated metropolitan areas.

II. Literature Review/Related Works

A. LULC Monitoring and Urbanization in Bangladeshs

Satellite remote sensing is key in monitoring the Land Use and Land Cover (LULC) shifts and forms the bedrock of contemporary urban and environmental studies. The literature shows clear evolution in approaches from primitive classification algorithms to sophisticated ones that leverage advanced cloud computing and artificial intelligence powered with machine learning. Satellite imaging in urban policy planning and the understanding of the complex socio-environmental dynamics of cities along with the space imagery in Bangladesh has been well documented [1], [9]. A substantial portion of scholarly work in Bangladesh has relied on multitemporal satellite data, especially from the Landsat series, to study urbanization. For example, Ahmed and Ahmed [7] simulated the growth of Dhaka and found that the urban footprint increased by over 300% in thirty years predominantly at the expense of cultivated land and wetlands. Comparable works in Khulna [10], Mymensingh [8], and Rajshahi [11] showed the same phenomenon of rapid built-up expansion driven by population growth and depletion of vegetation and water bodies. These research works lay an essential historical foundation for current change detection analysis.

B. Advancements in Remote Sensing Methodologies

Recent technological advancements have methodically transformed the field, particularly with the Sentinel-2 mission's higher spatial resolution and shorter revisit times, which improves LULC analysis. Moreover, the emergence of cloud computing platforms such as Google Earth Engine (GEE) has revolutionized the processing of large-scale satellite image time series (SITS) data [4]. This has been particularly very useful in recent LULC studies in Bangladesh where GEE's capabilities, along with advanced machine learning classifiers, were seamlessly applied over large geospatial regions [6], [12].It has been demonstrated that using spectral indices like NDVI and NDBI as input features increases classification accuracy by improving the separability of different LULC classes [11].

C. Machine Learning in LULC Classification

It is well-known that traditional classification techniques are greatly outperformed by machine learning algorithms, especially Random Forest (RF) and Support Vector Machines (SVM). These latter algorithms perform very well on the high-dimensional and complex data generated by multispectral imagery [4], [13]. Talukdar et al. [12] performed a study on LULC classification using different machine learning techniques in South Asia and found that RF performed best of all. It has been demonstrated that incorporating some indices such as NDVI and NDBI as input features increases classification accuracy due to better separability of different LULC classes.The applications of GEE together with machine learning techniques were created to generate LULC maps with better precision for a wide variety of regions [4], [6].

D. Research Gap

Despite the abundant literature, there appears to be a lack of gap in the cross-comparative analysis of LULC changes over time in multiple metropolitan areas of Bangladesh with a recent, high-resolution dataset after 2017. Most studies tend to focus on one city or work with older datasets of lower resolution. Also, sparse literature exists that focuses on benchmarking classifiers, including RF, SVM, and XGBoost, for the diverse urban settings of Bangladesh. This is precisely why we intend to conduct this research, to provide a robust in methodology multi-city study using the full potential of sentinel two data, GEE, and modern machine learning algorithms.

III. Dataset Description

A. Data Collection

- **Satellite Data Source:** Sentinel-2 L1C images-from Google Earth Engine (GEE). These images contain multi-spectral data, which is then classified into vegetation, water bodies, and urban areas.
- **Selected Bands:** The study utilized six spectral bands: **B3 (Green)**, **B4 (Red)**, **B8 (Near Infrared)** and **B11 (Shortwave Infrared)**. These bands were selected based on their effectiveness in detecting vegetation health, water content, and built-up features.
- **Study Area:** Spatial subsets of the three administrative districts—**Dhaka**, **Khulna**, and **Rajshahi**—were extracted using GIS shape files. These areas represent different land use patterns and urbanization levels.
- **Temporal Coverage:** Monthly composite images were generated from **March 2017 to March 2024**. The composites were averaged to reduce noise and enhance the temporal understanding of land cover changes.

B. Data Pre-processing

The raw satellite images underwent several pre-processing steps:

- cloud and shadow masking using QA60 bands and Sentinel-2 cloud probability layers.
- clipping to district boundaries using administrative shapefiles.
- resampling and band stacking, and stacking for uniform resolution and multi-band analysis.
- Computation of vegetation and land cover indices such as **NDVI (Normalized Difference Vegetation Index)**, **NDWI (Normalized Difference Water Index)**, and built-up indices for enhanced feature extraction.

IV. Methodology and Workflow

To capture the unique land patterns of each study area, Dhaka, Khulna, and Rajshahi, a district-specific classification process was carried out. The following steps summarize the methodology.

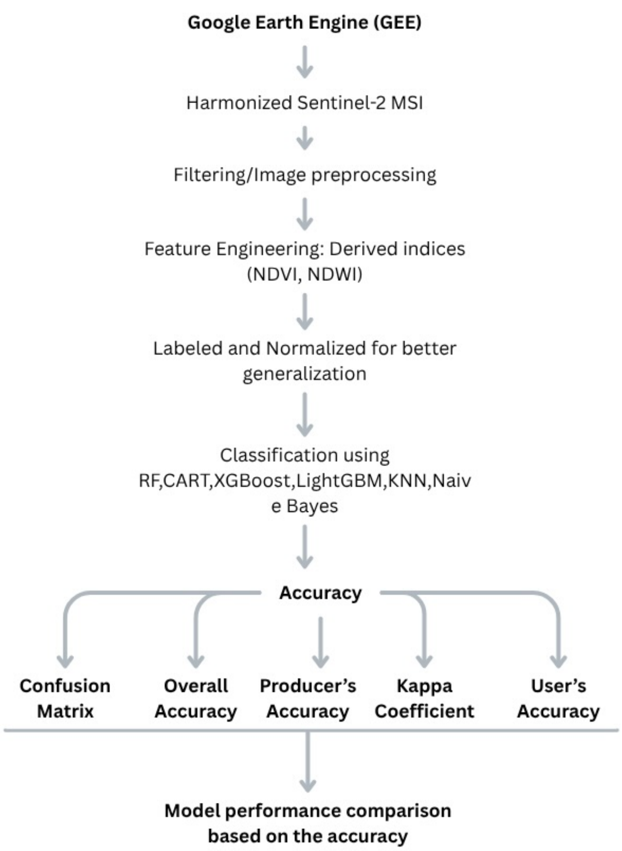

Fig. 1: Methodology for Land Use/ Land Cover classification using Google Earth Engine platform.

A. Data Labeling and Feature Selection

Manual annotation was employed to generate labeled data for four land cover classes: **vegetation**, **water**, **built-up**, and **barren land**. The labels were cross-verified using high-resolution Google Earth imagery and local knowledge, ensuring improved annotation accuracy.

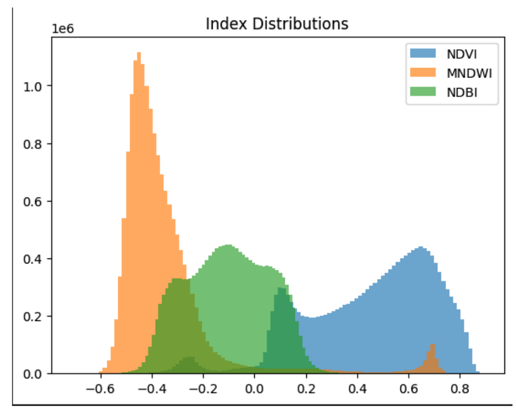

Fig. 2: labeling Dhaka 2024 dataset

Spectral bands **B3**, **B4**, **B8**, **B11**) were combined with three widely used indices to enhance classification quality:

- **NDVI (Normalized Difference Vegetation Index):** Reflects vegetation density and health, calculated as:

$$\text{NDVI} = \frac{B8 - B4}{B8 + B4}$$

- **NDWI (Normalized Difference Water Index):** Highlights surface water features, computed as:

$$\text{NDWI} = \frac{B3 - B8}{B3 + B8}$$

- **NDBI (Normalized Difference Built-up Index):** Used to identify built-up areas, defined by:

$$\text{NDBI} = \frac{B11 - B8}{B11 + B8}$$

B. Model Training and Classification

Six different machine learning classifiers were trained independently for each district:

- Random Forest (RF)
- CART (Classification and Regression Tree)
- XGBoost
- LightGBM
- K-Nearest Neighbors (KNN)
- Naïve Bayes

C. Evaluation and Performance Metrics

Model performance was assessed using the following metrics:

- **Overall Accuracy (OA):** Proportion of correctly classified pixels across all classes.
- **Kappa Coefficient:** Statistical measure of agreement between prediction and ground truth, adjusted for chance agreement:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the observed agreement and p_e is the expected agreement by chance.

- **User's Accuracy (UA):** Probability that a pixel classified into a given class actually represents that class on the ground:

$$\text{UA}_i = \frac{\text{Correctly classified pixels in class } i}{\text{Total pixels classified as class } i}$$

- **Producer's Accuracy (PA):** Probability that a reference pixel has been correctly classified:

$$\text{PA}_i = \frac{\text{Correctly classified pixels in class } i}{\text{Total reference pixels of class } i}$$

Every district conducted individual assessments and neural network construction, disentailing a distinguished ecological pattern of the district. All expressions include GPR, Random Forest D and different Merle algorithms AD. Each area was undergone preliminary attempts to evaluate its features as well as final output based on those features.

D. Time-Series Analysis

A long-term time-series analysis was conducted for **Dhaka** and its surrounding six districts—**Gazipur**, **Narayanganj**, **Manikganj**, **Munshiganj**, **Tangail**, and **Narsingdi**—covering the period from **2017 to 2024**. The analysis utilized monthly composites of Sentinel-2 imagery to monitor temporal variations in land cover classes.

The key observed trends include:

- **Urban Expansion:** A consistent increase in built-up areas, particularly in Dhaka, Gazipur, and Narayanganj, reflecting significant urban growth.
- **Vegetation Loss:** A gradual decline in vegetative cover, especially along the urban fringes.
- **Water Body Variation:** Seasonal fluctuations in surface water extent, with noticeable reductions during the dry season.
- **Rural Stability:** Relative land use stability in rural districts like Tangail and Manikganj, although slight infrastructure growth was noted.

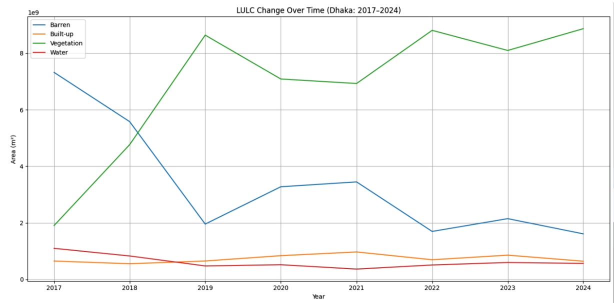

Fig. 3: Temporal trend of Land Use/Land Cover changes in Dhaka and its surrounding six districts from 2017 to 2024 based on monthly Sentinel-2 composites.

This visual representation highlights the changes in LULC classes overistol state in both seasonal andovicus and fortified advances variations and human made changes during different seasons and time intervals. LULC changes due to human efforts such as agriculture predis letter in name of anthropogenic cultivation are depicted heewith figuraments and each, what and in the years marked.

V. Experiments and Results

1st Dataset: Multiclass land cover classification over Dhaka (Vegetation, Water, Built-up, Barren).

Experiment Setup: Six classifiers applied using the same training and validation strategy.

Best Performing Models: Random Forest and CART achieved highest OA (0.999) and perfect UA/PA in most classes.

XGBoost and KNN also performed excellently with OA 0.99 and strong UA/PA balance.

Weakest Model: Naive Bayes, with significant drops in accuracy, especially for Built-up (PA = 0.206).

Observation: Tree-based ensemble models (Random Forest, XGBoost) are more effective in high-dimensional, imbalanced classification for urban data.

Model	Overall Accuracy (OA)	Kappa Coefficient (Kc)	UA - Vegetation	UA - Water	UA - Built-up	UA - Barren	PA - Vegetation	PA - Water	PA - Built-up	PA - Barren
Random Forest	0.999	0.999	1	0.999	0.999	0.999	1	0.999	0.998	0.999
CART	0.999	0.998	1	0.998	0.998	0.998	1	0.998	0.998	0.998
LightGBM	0.994	0.988	0.999	0.979	0.983	0.986	0.999	0.985	0.982	0.986
KNN	0.995	0.991	0.998	0.996	0.989	0.989	0.998	0.995	0.989	0.989
Naive Bayes	0.814	0.651	0.922	0.76	0.581	0.596	0.943	0.93	0.206	0.712
XGBoost	0.994	0.989	0.999	0.985	0.984	0.987	0.999	0.987	0.983	0.987

Fig. 4: Different Model Evaluation for Dhaka 2024 dataset

The Random Forest and CART models achieved close to flawless results in classification, displaying strong performance on all LULC classes in Dhaka. Performance metrics for vegetation and built-up areas were robust for both LightGBM and XGBoost. KNN is a reliable algorithm, but his classification accuracy was lower for water and barren areas. Naive Bayes was computationally fast, but PA for built-up classification was 0.206, which demonstrates the problem of many dimensions for spatial data.

These results support the assumption that ensemble-based models (RF, XGBoost, LightGBM) are effective for LULC complex classification tasks, particularly in densely populated areas like Dhaka.

Khulna's Experiments and Results Aim: We set out to classify land cover types in Khulna, Bangladesh, utilizing machine learning models based on multispectral satellite data along with assessing the accuracy benchmarks of different models. Recommended Models (High Performance): Random Forest Achieved Accuracy: 0.998 CART Achieved Accuracy: 0.997 XGBoost Achieved Accuracy: 0.993 Poor Performance Model (Low Precision): Naive Bayes Achieved Accuracy: 0.812

Model	OA	Kc	UA_Vegetation	UA_Water	UA_Builtup	UA_Barren	PA_Vegetation	PA_Water	PA_Builtup	PA_Barren
Random Forest	0.998	0.998	0.999	0.998	0.998	0.998	0.998	0.997	0.997	0.997
CART	0.997	0.996	0.998	0.997	0.997	0.996	0.998	0.996	0.996	0.995
LightGBM	0.993	0.987	0.998	0.978	0.981	0.985	0.997	0.984	0.98	0.984
KNN	0.994	0.99	0.997	0.995	0.988	0.988	0.997	0.994	0.987	0.987
Naive Bayes	0.812	0.648	0.92	0.759	0.579	0.595	0.941	0.928	0.205	0.71
XGBoost	0.993	0.988	0.998	0.984	0.983	0.986	0.998	0.986	0.982	0.985

Fig. 5: Different Model Evaluation for Khulna 2024 dataset

The same classification pipeline was applied to Rajshahi, and preliminary evaluations confirmed consistent model performance across districts. However, due to space limitations, this paper highlights the Dhaka and Khulna district only. Comparative multi-district results are reserved for a future extended version of this study.

VI. Future Work

In the future, we want to take our project a step further from the detection of greenery and built-up area.We would like to identify different types of pollution, such as air and water pollution, from satellite images.For this, we will utilize more advanced image processing and machine learning algorithms to decipher patterns in the satellite data.We also want to track how land cover and pollution changes over time.This will aid in understanding long-term environmental trends and allow for better planning and decision making.We could also integrate other data, such as weather information or ground sensors, to help tune our results.Our long term goal is to make sustainable development.

Conclusion

This project is a great example of using satellite image data and image processing technique.From observing land cover from above, we can get vital information on urban growth, deforestation, and degradation of the environment. Our approach is quite affordable and scalable for monitoring vast spaces without conducting ground surveys.

The results of this research can be used to guide better planning and environmental management. In the future, our project could have the potential to grow into monitoring also,which would great for sustainable development as well as for natural activities

References

[1] "A Review on The Impact of Satellite Imagery in Urban Policy Planning," referenced in project documentation.

[2] "Urban Remote Sensing with Spatial Big Data: A Review and Future Prospects," referenced in project documentation.
W

[3] Z. Zhang, et al., "Deep learning for land use and land cover classification based on Sentinel-2 and Landsat-8 satellite images," Applied Sciences.

[4] M. A. M. Ali, et al., "Multi-temporal Land Use/Land Cover classification using machine learning algorithms on Google Earth Engine," Remote Sensing.

[5] "Deep Learning for Satellite Image Time Series Analysis: A Review," referenced in project documentation.

[6] A. B. Azam, M. Nasrin, and M. Z. Mredul, "Monitoring Urban Growth and Its Impact on Arable Land Consumption in the Lower Turag Basin of Dhaka: A Google Earth Engine Based Study," Land, vol. 13, no. 2, p. 112, Feb. 2024.

[7] B. Ahmed and R. Ahmed, "Modeling urban land cover growth dynamics using multi-temporal satellite images: A case study of Dhaka, Bangladesh," ISPRS International Journal of Geo-Information, vol. 1, no. 1, pp. 3-31, 2012.

[8] M. M. Rahman and M. A. M. Ali, "Monitoring land use/land cover changes using remote sensing and GIS techniques: A case study of the Mymensingh district, Bangladesh," Journal of the Indian Society of Remote Sensing, 2021.

[9] "Understanding the Urban Environment from Satellite Images with New Classification Method Focusing on Formality and Informality," referenced in project documentation.

[10] K. Islam, M. Jashimuddin, B. Nath, and T. K. Nath, "Land use classification and change detection by using multi-temporal remote sensing imagery: A case study of the City of Khulna, Bangladesh," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 42, pp. 391-397, 2018.

[11] M. S. Islam, M. A. Rahman, and M. A. Islam, "Assessing the LULC changes and its impact on LST in Rajshahi, Bangladesh using GEE," Journal of Geographic Information System, vol. 13, no. 5, pp. 585-605, 2021.

[12] S. Talukdar, P. Singha, S. Mahato, S. Shahfahad, J. Pal, Y. Liou, and A. Rahman, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," Remote Sensing, vol. 12, no. 7, p. 1135, Apr. 2020.

[13] L. Mountrakis, J. Im, and C. Ogole, ”Support vector machines in remote sensing: A review,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, no. 3, pp. 247-259, May 2011.