Grenze International Journal of Engineering and Technology, January Issue

Geospatial Analysis of Land Cover Maps: A Case Study in Mumbai Metropolitan Region, India

Devanand K Bathe[1] and Sapna Prabhu[2]
[1-2]Department of Electronics Engineering, Fr CRCE, Mumbai, India
Email: devanandkb@somaiya.edu, sapna@fragnel.edu.in

***Abstract*— Land Use Land Cover (LULC) classification is crucial for effective resource management, urban planning and environmental monitoring. Remote sensing based satellite imagery in association with machine learning has great potential for LULC classification. This study suggests a hybrid machine learning model that combines the ensemble learning and feature fusion approach of Support Vector Machine (SVM), Random Forest (RF) and Gradient Tree Boosting (GTB) as base learner and bagging based RF as meta-classifier on Sentinel 1 - VH and VV polarization band satellite imagery for the year 2024. Experimental results show that the hybrid model outperforms standalone models in terms of overall accuracy. Additionally, feature importance scores are computed. The generated LULC map is supported with the ESA world cover map. This proposed study is highly beneficial to the policy makers.**

***Index Terms*— LULC, Machine Learning, Hybrid Model.**

I. INTRODUCTION

LULC dynamics play a critical role in understanding environmental change, urban growth, deforestation, agricultural expansion and climate variability. LULC classification refers to the process of categorizing earth's surface features into various categories. Land cover refers to what physically covers on the earth surface while land use refers to how these features are utilized. Traditional LULC mapping methods such as field surveys and manual interpretation are often limited, time consuming and expensive. Satellite Remote Sensing (RS) offers a powerful alternative by providing consistent, repeatable, and large-scale Earth observation data. Additionally, Machine Learning (ML) algorithms have further revolutionized the potential of satellite imagery in classification land cover and change detection. Techniques such as RF, GTB and SVM enable the extraction of high-level patterns and spatio-temporal insights with improved accuracy and scalability. From the related literature, it is apparent that the aforementioned methods are mainly utilized on optical data. Unavailability of such optical satellite images in the rainy season poses several challenges in LULC classification application. To address this issue, this research work utilizes Synthetic Aperture Radar (SAR) satellite imagery in association with machine learning to perform LULC classification of study regions. This integration addresses the growing need for robust, intelligent systems capable of monitoring rapid landscape changes due to urbanization, deforestation, natural disasters, and climate shifts.

Research contribution in this paper is as follows:

- The proposed hybrid model utilizes RF, SVM and GTB as base learners. Additionally meta-learner is implemented on stacked prediction features from aforementioned ML models and Sentinel-1 VH polarization band.

Grenze ID: 01.GIJET.12.1.157

- An integrated approach of ensemble learning through stacking and feature fusion is employed for LULC classification.
- Feature importance score is calculated for the meta-learner.

The rest of the paper is ordered as: Section II highlights the related work; Section III defines the proposed methods; Section IV represents experimental outcomes and related discussion, and lastly Section V mentions the conclusion.

II. LITERATURE REVIEW

LULC classification has evolved significantly due to evolution of technologies in Remote Sensing (RS) and the abundant availability of satellite imagery. In the related literature, a variety of methods for LULC classification methods are studied. From the related literature, it is evident that researchers are working on conventional approaches along with ML and Deep Learning (DL) based approaches for assessing LULC changes. For instance, Rashmi Saini and Suraj Singh [1] have classified seven distinct land cover classes: snow, forest, water bodies, grassland, agricultural land, fallow land, and built-up areas for Tehsil Karnaprayag, situated in the vicinity of the Greater Himalaya, by using multi-spectral sentinel-2 imagery and XGBoost and K-NN as ML classifiers. In spite of its notable performance, it has been constrained in terms of non-availability of data in cloud cover, ground truth validation and multi sensor or multi temporal analysis. To address the limitation of cloud cover, Kavita B et al. [2] studies the dynamics of LULC classification in the city Mumbai, Maharashtra state of India, using Sentinel-1 satellite imagery from Google Earth Engine (GEE) and ML. Five ML algorithms, including RF, SVM, GTB, K-NN, Classification and Regression Tree (CART) are experimented on SAR data for LULC classification for the years 2014 and 2024. Experimental results depict that RF performs better than other ML algorithms. Though promising results are observed, generalization is a major concern. In another study, Shubhra Swetanisha et al. [3] utilize a supervised classification approach, applying SVM to process multi-temporal satellite imagery. Genetic algorithm (GA) is integrated with SVM to enhance the feature selection process. It highlights the gap in the integration of more advanced AI techniques which could potentially improve the accuracy and efficiency of change detection processes. Nitu Wu et al. [4] utilized RF, SVM, K-NN, and Naive Bayes (NB) and performed studies on the semiarid grassland region of central Xilingol region in northern China for mapping semiarid grassland by comparing various RF, SVM, K-NN, XGBoost as ML algorithm for both pixel based and object based LULC classification. Here, though object based RF outperformed pixel based other methods; absence of multi-temporal fusion approaches could enhance class separability and temporal robustness in dynamic grassland ecosystems. Harsh Waghela et al. [5] used ML algorithms including the CART, RF, DT and SVM to perform the LULC classification. Experimental results indicated that CART and RF perform well but need to explore ensemble methods to improve spatial resolution and classification robustness. Rashmi Saini et al. [6] implemented ANN, SVM, K-NN, and RF for LULC classification of Nainital district, Uttarakhand state, India. Though RF outperformed other ML methods, to enhance spatial details and generalization, ensemble methods are needed. Sana Basheer et al. [7] has selected the study area as Charlottetown City in Canada. ArcGIS Pro and GEE platforms are used considering datasets obtained from Landsat, Sentinel. LULC classification is performed for three classifiers in ArcGIS Pro, considering RF/RT, SVM, maximum likelihood (ML) and four classifiers including minimum distance (MD), SVM, CART, RF/RT are used to develop LULC maps for the span 2017-2021. For a dynamic urban environment, ensemble methods are needed though RF methods outperforming other ML methods. Tapan Kumar Das et al. [8] utilizes various ML techniques, including RF and K-NN to classify satellite images and predict LULC changes. The research identifies gaps in the integration of advanced ML models for more accurate LULC prediction. Rajeshwari Harini Bikkasani et al. [9] have done analysis of long term changes for LULC using ML at Manipur state, India. Changes in LULC features aid to ensure sustainable development of the study region of state. Significant reduction of forest cover is seen over recent years. Ouchra et al. [10] have proposed ML algorithms for Morocco as case study. By using GEE and Landsat satellite data, satellite image classification is proposed an experimental study of six supervised ML algorithms, such as RF, CART, MD, SVM, DT and GTB, to classify various areas such as water, cultivated, built up, barren, sandy and forest on the Moroccan region to find out the best performing higher accuracy classifier. A Review on LULC change modeling for urban development has been done by Shrishti Gaur et al.[11] by giving a review of all existing LULC modeling techniques with novel approaches. The highlight of study is the utility of the hybridization of ML models combined with statistical models to complement their strengths. Sanjay Moulik [12] focuses on utilizing remote sensing data for land cover classification in India. The study employs various existing standalone ML algorithms to classify land cover types, aiming to enhance accuracy and efficiency of land use planning and resource management. The research identifies gaps in the application of advanced ML techniques for land cover classification in India and suggests integration of such methods to improve classification outcomes. In another research, Rehab Mehmoud et al [13] proposed the prediction model. LULCRV, consisting of 8 LULC classes with 3.0 m spatial multi-spectral high resolution satellite images of Egypt. Land changes in Egypt can be generalized using an ensemble learning approach. Dipthi Chennu et al. [14] have evaluated various machine learning classification techniques using GEE. Over the study area of Machilipatnam region utilizing two distinct years that are 2017 and 2022, researchers evaluated the LULC classification performance of the using GEE with datasets obtained from Sentinel, Landsat. In spite of notable performance of ML algorithms on optical data, it has been constrained in terms of non-availability of data in cloud cover. While many researchers used optical data, Kavita Bathe et al. [15] used RF on multi-temporal SAR data to perform LULC of Jagatsinghpur district of Odisha state from India. This work has utilized only standalone models. Ensemble learning can enhance the results. The study by Priyanka Gupta et al. [16] performed LULC mapping and change detection in the Imphal Valley using ML based hybrid approach combining Sentinel-1 SAR and Sentinel-2 optical imagery. The research gap lies in the limited exploration of temporal fusion of AI techniques for improved accuracy in LULC mapping and change detection. In another study, Luvkesh Attri et al. [17] conducted urban land cover classification using Polarimetric Interferometric SAR (PollanSAR) data from ALOS-2 and RADARSAT-2 satellites. The study applied ML algorithms to discriminate complex urban features with improved accuracy. Ensemble learning framework is suggested to enhance urban feature differentiation and generalization across diverse environments. The paper by Eman A. Alshari et al. [18] implements a hybrid classifier combining ANN and RF to perform LULC classification for the city Sanaa, Yemen using multispectral data from Sentinel-2A and Landsat-8. The authors highlight that there is a gap in simpler ML approaches that can deliver good accuracy in comparison to DL methods entailing high complexity, long training times and large parameter sets. From these papers, it is evident that Hybrid ML algorithms can provide an efficient balance of accuracy, interpretability and computational simplicity. They perform well with limited training data, integrate multisource features and require fewer resources for hybrid ML models. Thus they offer a practical and reliable alternative for large scale LULC classification and remote sensing applications. This is the motivation behind this research work.

III. PROPOSED METHOD

The research is conducted in Maharashtra state of India over Mumbai Metropolitan Region (MMR). It consist 9 municipal corporations, 9 municipal councils and more than 900 villages lie over 2834.57 km2 area. The MMR area's latitudinal extent is 18.60° N to 19.50° N, and the longitudinal expansion is 72.80° E to 73.25° E. This study area spans all parts of Mumbai City, Mumbai Suburban, a sub district area of Thane and Raigad which constitutes Thane and Navi Mumbai. The study area was depicted using a map obtained from the Administrative Boundary Database provided by the Survey of India [19], concentrating on the city of Mumbai and some of its neighboring sub-districts. Over the past decade, there have been major LULC changes in the area. This makes the study region an ideal choice to reflect diverse land cover types and the transformations that have taken place. Fig. 1 displays the research area's geographical location.

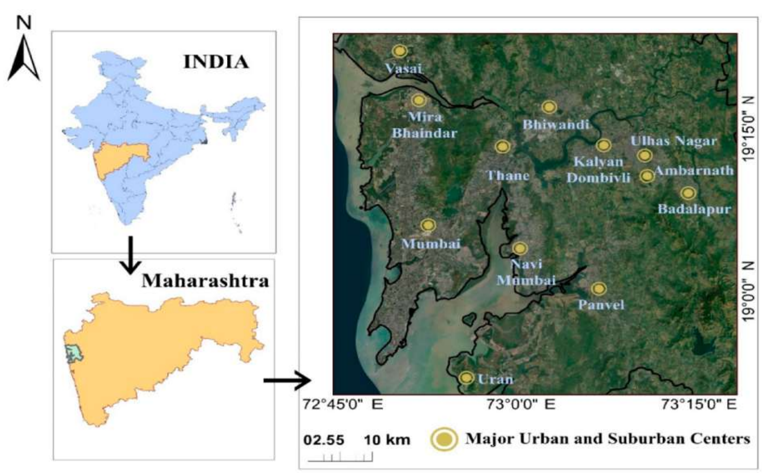

Figure 1. MMR's geographical location.

A. Methodology

The study engages Sentinel-1 SAR satellite imagery for LULC classification using GEE. Fig. 2 depicts a block diagram of a machine learning based proposed hybrid model proposed in this research work. The several stages are discussed in detail below.

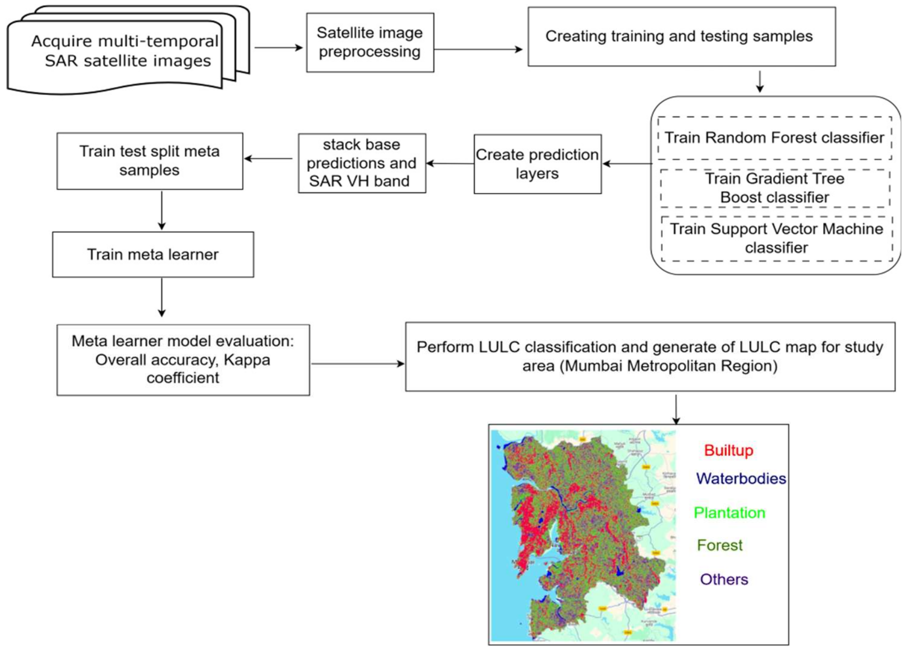

Figure 2. Block diagram of proposed methodology

Implementation Stages:

- Data Acquisition: In this study, Sentinel-1 data is utilized with VH and VV polarization satellite imagery and acquired satellite images from Google Earth Engine (GEE) [20] from January 2024 to December 2024 and clipped to the study area. Table I depicts acquisition of satellite images with configuration details.

TABLE I: CONFIGURATION DETAILS

Cloud platform	Satellite sensor	Product type	Operational modes	Polarizations	Orbit Properties pass	Acquisition dates
GEE	Sentinel-1	Level-1 GRD	Inter-ferometric wide swath mode (IW)	VH ,VV	Ascending, Descending	01/01/2024 to 31/12/2024

- Data Processing: SAR satellite images are complex in nature and are greatly affected by speckle noise. Therefore, the Refined Lee filter [21][22] with kernel size of 7x7 is utilized for image despeckling. Apart from this the median is calculated for all images to get a clean, stable image. The final single stacked polarization image is used.
- Creation of training and testing data sample: The ground truth survey of the study area conducted from the period January 2024 to December 2024 and latitude and longitude of various ground truth points are collected for land cover classes such as Water bodies, Plantation, Built up, Forest and Others. These records are imported to GEE and the training data points are created by using the marker function in GEE. 50 samples are considered for each land cover class making a total data set of 300 samples. These data samples are then split into training (70%) and validation samples (30%) to improve the testing and overall accuracy of the models being trained.
- Train base learners: Machine learning provides state of art algorithms. Each of these algorithms have its own pros and cons. Ensemble learning methods combine various supervised machine learning algorithms and build a strong classifier. In the present study, three ML algorithms, RF, GTB and SVM are utilized as base learners with hyper parameter values given in Table-II. These base learners are trained on training data.

TABLE II: HYPER-PARAMETERS

Sr. No.	ML algorithm	Hyperparameter values
1	SVM	Kernel type: Cost:10, RBF, Gamma:0.5
2	RF	Number of decision tree: 60, Bag fraction:0.5, minimum number of leaf population:1
3	GTB	Number of trees:10, Sampling rate:0.7, Shrinkage: 0.005

- Create prediction layer and Stacking: The aforementioned base learners are implemented on test data and prediction layer is created. The prediction layers from three base learners namely SVM, RF and GTB are stacked with Sentinel-1 polarization band satellite image.
- Train Test split: The images obtained are split up to 70% for training and 30% for testing.
- Train meta-learner: In this study, the stacking ensemble learning method utilizes bagging and boosting for creating meta-learner. Initially RF is experimented as a meta-learner and next GTB is utilized as a meta-learner.
- Evaluate meta-learner: Finally RF is used as an algorithm for meta-learner and evaluated based on the parameters such as Overall Accuracy (OA) and Kappa coefficient.
- Generate study area's LULC map: The meta-learner has classified the stacked satellite images and generated the LULC map of a study area. Built-up, Water-bodies, Forest, Plantation and Others are 5 land cover classes of generated LULC maps.

IV. RESULTS AND DISCUSSIONS

In this section, the implementation of the proposed hybrid model is assessed on the MMR study area. Various performance evaluation metrics used for the comparison include OA, PA, UA and Kappa coefficient. These were selected based on its high ability to extensively measure the performance of the proposed hybrid model. The state-of-the-art models are used for comparison purposes, and a summary of performance shown in Table III. Experimental results show that the proposed model outperformed the other ML models. Fig. 3 depicts the performance evaluation of the Kappa coefficient on VH and VV polarization band. Further, validation of the results is done by comparing it with the ESA world cover map as shown in Fig. 4. While most of the studies in the literature have utilized ML algorithms and generated the LULC map, this study has computed the feature scores of ML features which help to build trust on the proposed model. Fig. 5 depicts the feature importance score. It shows that random forest has played a vital role in terms of feature importance score.

TABLE III: PERFORMANCE EVALUATION OF DIFFERENT MODELS ON BENCHMARK DATA SETS

Algorithm	Overall Accuracy (VH)	Overall Accuracy (VV)
RF	0.9528	0.9537
SVM	0.5471	0.5601
GTB	0.75	0.7962
RF + GTB + Meta learner	0.92	0.9523
RF + SVM + GTB + Meta learner	0.9333	0.9649

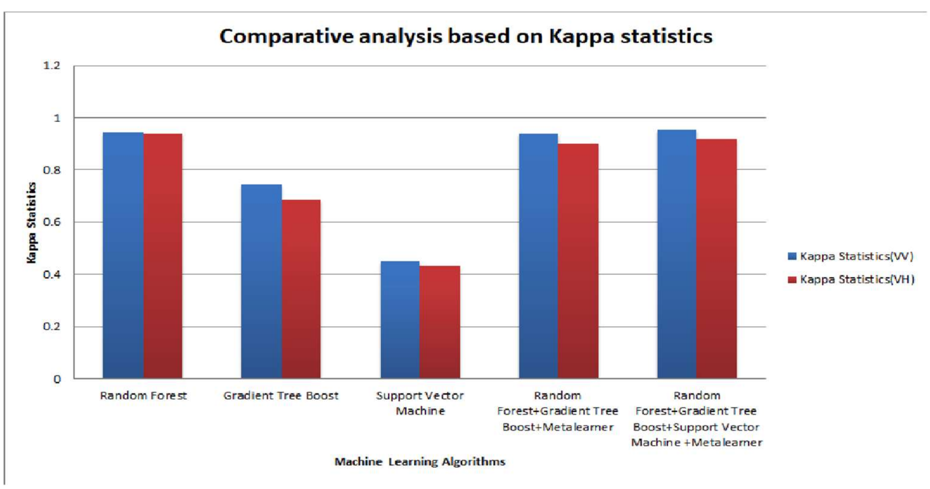

Figure 3. Performance evaluation of Kappa coefficient

2024 Hybrid model-VV **ESA world cover map**

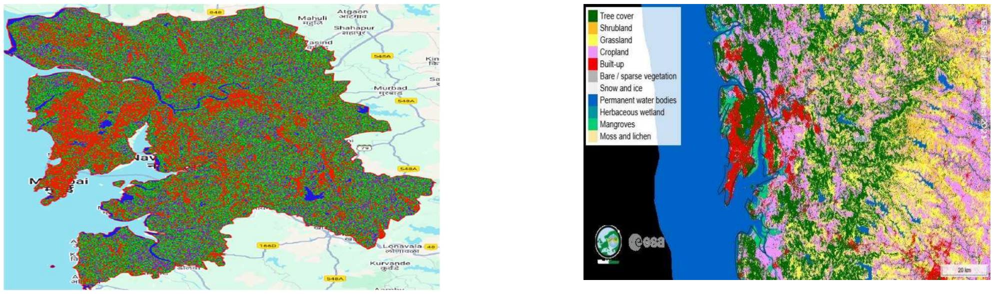

Figure 4. Validation of results with ESA world cover map

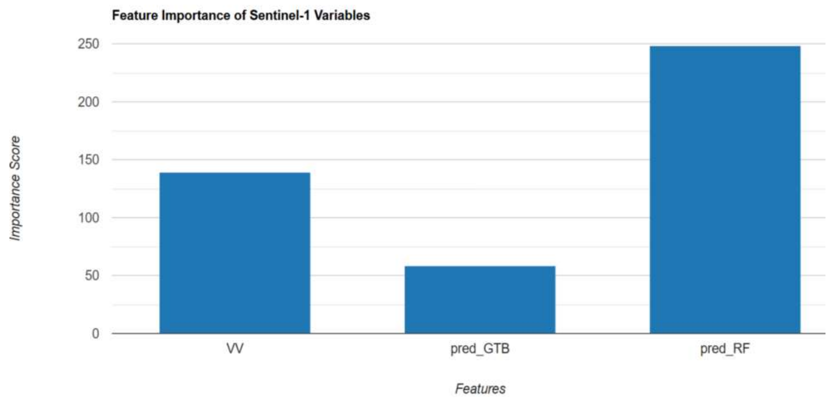

Figure 5. Feature importance score of various features for meta-learner

The proposed model has attained a remarkable performance study area. The results obtained show that VV polarization has given better results in terms of OA and kappa coefficient as compared to VH polarization for all experimented algorithms except CART. The proposed hybrid model has outperformed with remarkable accuracy of 98.61% and kappa coefficient of 0.9823. The hybrid model of RF, GTB and meta-classifier is also experimented and the results show that it has shown the least performance in contrast to the proposed model that utilizes RF, SVM, GTB and Meta learner. RF and GTB will be the meta-learners used. From the experimental results it is clear that hybrid models having base learner models combine strength of diverse ML algorithms, while meta classifier handles nonlinear blending leading to higher accuracy. Both the meta classifiers have shown similar performance and efficiently generated study areas LULC map.

V. Conclusion

In this study, a hybrid-ML model is implemented on MMR. The study deals with implementation of the proposed hybrid machine learning method on Sentinel-1 SAR data for LULC classification. The hybrid-ML model utilizes the potential of ensemble learning and feature fusion approach. The study has utilized RF, SVM and GTB as base learners and bagging based RF as meta-classifier on Sentinel 1- VV and VH polarization band satellite imagery for the year 2024. The experimental results exhibit that the proposed method with VV polarization outperforms the traditional single ML algorithms and it has efficiently generated the LULC map of the MMR region.

References

[1] Rashmi Saini, Suraj Singh, “Land Use Land Cover Classification Using Machine Learning and Remote Sensing Data: A Case Study of Karnaprayag, Uttarakhand, India”, 2024 IEEE First International Conference on Electronics, Communication and Signal Processing (ICECSP 2024).

[2] Kavita B, V. Mehta and A. Mehta, "Land Use Land Cover Classification and Change Dynamics: A Case Study in Mumbai, Maharashtra, India," 2024 IEEE India Geoscience and Remote Sensing Symposium (InGARSS), Goa, India, 2024, pp. 1-4, doi: 10.1109/InGARSS61818.2024.10984209.

[3] Shubhra Setanisha, Amiya Ranjan Panda and Dayal Kumar Behera, “Change Detection using Machine Learning Models: A Case Study on the Puri District of Odisha, India”, 2021 19th OITS International Conference on Information Technology (OCIT).

[4] Wu, N., Crusiol, L.G.T.; Liu, G., Wuyun, D. Han, G., “Comparing Machine Learning Algorithms for Pixel/Object-Based Classifications of Semi-Arid Grassland in Northern China Using Multisource Medium Resolution Imageries”, MDPI International Journal on Remote Sensing, 2023, 15, 750. https://doi.org/10.3390/rs15030750

[5] Harsh Waghela, Saurin Patel, Pooja Sudesan, Soham Raorane, Rohan Borgalli, “Land Use Land Cover Classification using Machine Learning”, IEEE International Conference on Automation,Computing and Renewable Systems (ICACRS 2022).

[6] Rashmi Saini, Shivam Rawat, “Land Use Land Cover Classification In Remote Sensing Using Machine Learning Techniques”, 1st IEEE International conference on Innovations in High Speed Communication and Signal Processing (IEEE-IHCSP) 4-5 March, 2023.

[7] Sana Basheer, Xiuquan Wang , Aitazaz A. Farooque, Rana Ali Nawaz, Kai Liu, Toyin Adekanmbi,Suqi Liu, “S. Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques”, MDPI International Journal on Remote Sensing 2022, 14, 4978.

[8] Tapan Kumar Das, Dilip Kumar Barik, K.V.G. Raj Kumar, “Land-Use Land-Cover Prediction from Satellite Images using Machine Learning Techniques”, 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), 26-27 May 2022. .

[9] Rajeswari Harini Bikkasani, Dhanya M, Veena S V, “Analysis of Long-term Changes for Land Use and Land Cover using Machine Learning : A case study”, IEEE Proceedings of the Second International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2023)

[10] Hafsa Ouchra, Abdessamad Belangour, Allae Erraissi, “Machine Learning Algorithms for Satellite Image Classification Using Google Earth Engine and Landsat Satellite Data: Morocco Case Study”, IEEE Access, 2023.

[11] Srishti Gaur, Rajendra Singh, “A Comprehensive Review on Land Use/Land Cover (LULC) ChangeModeling for Urban Development: Current Status and Future Prospects”, MDPI International Journal on Sustainability 2023, 15, 903.

[12] Sanjay Moulik, “Land Cover Classification: A study from an Indian Perspective”, 2023 9th International Conference on Smart Computing and Communications (ICSCC).

[13] Rehab Mehmoud, Mohamed Hassanin, Haytham AI Feel, Rasha M. Badri, “Machine Learning Based Land Use and Land Cover Mapping Using Multi-spectral Satellite Imagery: A Case Study in Egypt”, Sustainability 2023, 15, 9467. https://doi.org/10.3390/su15129467.

[14] Dipthi Chennu, Suneetha Manne, Tharun Sai P., “Machine Learning Techniques of Land Use and Land Cover Classification Using Google Earth Engine: A Comparative Study”, 2023 Innovations in Power and Advanced Computing Technologies (i-PACT)

[15] Bathe, K.D., Patil, N.S. “Assessment of land use-land cover dynamics and its future projection through Google Earth Engine, machine learning and QGIS-MOLUSCE: A case study in Jagatsinghpur district, Odisha, India” J Earth Syst Sci 133, 111 (2024).

[16] Priyanka Gupta, Arun Kumar Shukla, and Derris Praise Shukla, “ML-Based Hybrid SAR and Optical Image LULC Mapping and Change Analysis With Variations in the Air Quality of the Imphal Valley, North-East India”, Earth and Space Science, 11, e2023EA003176 (2024).

[17] Luvkesh Attri, Shashi Kumar, Sandeep Maithani, “Machine learning based urban land cover classification using PolInSAR data: a study with ALOS-2 and RADARSAT-2 datasets”, Discover Geoscience (2024).

[18] Eman A. Alshari, Mohammed B Abdulkareem and Bharti W. Gawali, “Classification of Land Use/Land Cover using Artificial Intelligence”, Frontier Artificial Intelligence, 06 January 2023Sec. Machine Learning and Artificial Intelligence, Volume 5 - 2022 | https://doi.org/10.3389/frai.2022.964279.

[19] Bhanage Vinayak, Han Soo Lee, Shirishkumar, Gedem, “Prediction of Land Use and Land Cover Changes in Mumbai City, India, Using Remote Sensing Data and a Multilayer Perceptron Neural Network-Based Markov Chain Model”, MDPI International Journal on Sustainability 2021, 13, 471.

[20] Meisam Amani, Arsalan Ghorbanian, Seyed Ali Ahmadi, Mohammad Kakooei, Armin Moghimi, Mohammad Mirmazloumi, “Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review” IEEE Journal of selected topics in applied Earth Observations and Remote Sensing, Vol. 13, 2020.

[21] Lee J 1980 Digital image enhancement and noise Bltering by use of local statistics; IEEE Trans. Pattern Anal. Mach. Intell. PAMI-2 165–168.

[22] Yommy A S, Liu R and Wu A S 2015 SAR image despeckling using reBned Lee Blter; 7th International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, pp. 260–265