GEO-EYE

ORIGINAL ARTICLE

OPEN ACCESS

Received: 11-07-2024

Accepted: 10-10-2025

Published: 20-10-2025

Citation: Sultana S, Inayathulla M. (2025). Precision Land Use and Land Cover Classification using Google Earth Engine: Integrating Random Forest and Support Vector Machine Algorithms. Geo-Eye. 14(1): 54-58. https://doi.org/10.53989/bu.ge.v14.i1.24.14

* **Corresponding author**.
salmas216@gmail.com

Funding: None

Competing Interests: None

Published By Bangalore University, Bengaluru, Karnataka

ISSN
Print: 2347-4246
Electronic: XXXX-XXXX

Precision Land Use and Land Cover Classification using Google Earth Engine: Integrating Random Forest and Support Vector Machine Algorithms

Salma Sultana[1]*, M Inayathulla[1]

1 *Department of Civil Engineering, Bangalore University, Jnana Bharathi, Bengaluru, Karnataka, India*

Abstract

Land Use and Land Cover (LULC) classification is crucial for understanding and managing environmental resources. This study introduces an innovative methodology that leverages Sentinel satellite data alongside two robust machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM), on the Google Earth Engine (GEE) platform. Renowned for its high-resolution multispectral imagery, Sentinel data offer rich information for classification. GEE provides access to extensive geospatial datasets and computational resources, enabling effective analysis. RF and SVM are known for their ability to handle complex datasets, optimizing classification accuracy. The study outlines a systematic workflow for preprocessing Sentinel imagery, followed by the implementation of RF and SVM algorithms, with a focus on accurately classifying vegetation, built-up areas, barren land, and water bodies. Evaluation metrics, including overall accuracy and kappa coefficient, demonstrate the efficacy of the proposed methodology. This compelling study highlights the utility of RF and SVM within GEE for precise LULC mapping, emphasizing their pivotal role in supporting informed decision-making for environmental planning and conservation initiatives.

Keywords: Google Earth Engine; Support Vector machine; Random Forest; Sentinel satellite data; Land Use and Land Cover

1 Introduction

LULC Classification plays an important role in understanding the dynamic interaction between the activities which are done by human being for the betterment of day to day life and the environment. It has become the fundamental source for accessing changes in the earth's surface. Understanding LULC classification provides insights into ecosystem health, habitat fragmentation, urbanization patterns, and climate change. Remote sensing techniques along with advanced classification algorithms are essential for accurate LULC mapping. These classifications provide a valuable input for land-use planning, agricultural management, water resource assessment, and policy formulation at local, regional, and global scales. Monitoring and continuously updating of LULC datasets are for tracking transformation in landscapes over time, thereby supporting to achieve sustainable development goals and informed decision making processes [1].

Sentinel-2 offers high-resolution multispectral data, enabling detailed and accurate classification of LULC categories. The spectral bands provided by Sentinel-2, including visible, near-infrared, and short-wave infrared, facilitate discrimination between various land cover types based on their unique spectral signatures [2]. The Sentinal-2 satellite image used need to undergo pre-processing steps such as radiometric calibration, geometric rectification, and atmospheric correction are essential to ensure the quality and accuracy of LULC classification [3]. Monitoring Sentinel-2 imagery continuously helps for temporal analysis of change in LULC over time, supporting making better decisions for land management, environmental monitoring, and climate change studies. Sentinal-2 satellite image has the ability to capture seasonal variations and changes in vegetation health, which is very essential for the identification of agricultural land use, forest covers, change in urban areas, water bodies and other type of land cover [4]. Accuracy of LULC classification can be enhanced through the integration of Sentinal-2 satellite image with digital elevation models (DEM), land use maps, and field surveys. The availability of cloud-based processing platform like GEE and open-access nature of sentinel-2 imagery make LULC classification cost-effective for a wide range of applications.

GEE is a powerful platform for conducting LULC analyses using a vast archive of satellite imagery. It offers access to a wide range collection of remote sensing datasets, such as Sentinal-2, Landsat, MODIs, and more, covering a wide range of spatial and temporal resolutions suitable for classification of LULC. JavaScript-based programming interface on GEE platform enables user to develop scripts for LULC classification, offering flexibility and customization options to suit specific research or application needs. Collaborative features present on GEE platform enable users to develop custom scripts to share code, datasets, and results, fostering collaboration and knowledge exchange in the field of LULC analysis. The scalability, accessibility and rich feature set of GEE make it a valuable tool for conducting LULC research, monitoring environmental changes, and supporting decision making processes [5].

RF is a method that combines the prediction of multiple decision trees, making it robust against over fitting and capable of handling complex tasks such as LULC classification. SVM comes under the category of supervised learning algorithm that works well for LULC classification by finding the optimal hyperplane that separates different classes in the feature space. GEE provides built in implementation of RF and SVM algorithms, allowing users to apply these techniques easily for LULC classification using satellite imagery [6]. JavaScript API on GEE enables users to customize RF and SVM classification workflows, such as parameter tuning and feature selection, thereby optimizing classification accuracy for LULC mapping objectives. Leveraging advanced machine learning algorithms such as RF and SVM within that GEE platform has become popular to perform the LULC analysis accurately and efficiently [7]. The study done in this paper explores the application of RF and SVM which is used for the classification of LULC on GEE platform.

2 Materials and Methods

Study Area

Bangalore is the fourth largest Municipal Corporation in India, responsible for managing a population of 6.8 million in an area of 741 km^2. Its boundaries have expanded tenfold over the last six decades. It lies in the southeast of south Indian state of Karnataka (Figure 1), Bangalore sits in the heart of Mysore Plateau at an average elevation of 900 meters. The city is located at 12°58′44″N latitude and 77°35′30″E longitude.

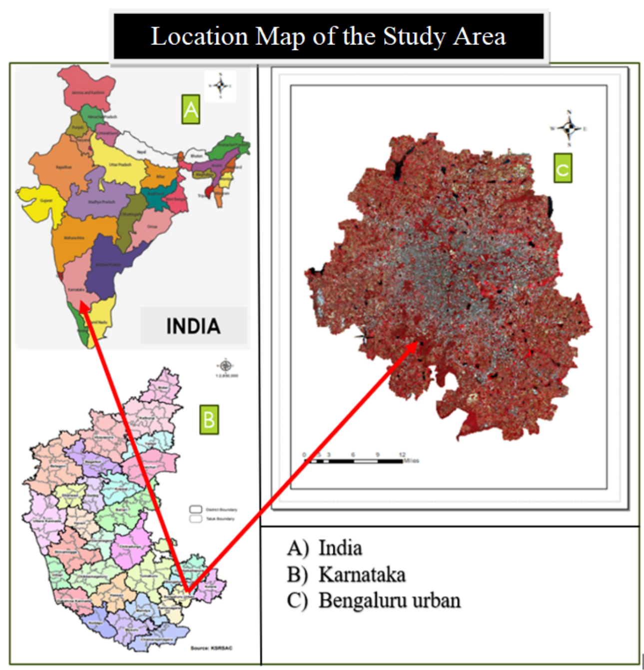

Fig. 1. Study Area

Data

This study uses Sentinel-2 satellite image, which provides medium resolution multispectral data, for classifying different LULC. Sentinal-2 satellite imagery provides spectral information across various wavelengths, including visible, near-infrared (NIR), and short-wave infrared bands (SWIR) bands. There are 13 spectral bands in Sentinel-2 image. The specifications of Sentinel-2 spectral bands are shown in

Table 1.

Table 1. Specifications of Sentinal-2 spectral bands

Bands	Description	Spatial resolution
2	Blue	10
3	Green	10
4	Red	10
5	Red Edge 1	20
6	Red Edge 2	20
7	Red Edge 3	20
8	NIR	10
11	SWIR	20
12	SWIR	20

NIR – Near- InfraRed reflectance
SWIR – Short wave InfraRed

The Sentinel-2 image used in the study of LULV classification is accessed directly from Copernicus Open Access Hub using GEE (https://earthengine.google.com/) platform[8]. The Sentinel-2 satellite image used in the study is dated form 2023-01-12 to 2024-03-07.

Methodology

Sentinel-2 satellite image used in the study are acquired from the European space Agency (ESA) Sentinel mission, specifically Sentinel-2A and Sentinel-2B satellites. These satellites orbit the Earth in a sun synchronous manner, capturing images with the revisit time of 5 days, ensuring frequent and consistent monitoring of the earth surface. Cloud masking techniques are applied to mitigate the impact of cloud cover. The spectral bands of Sentinel-2 image, coupled with its spatial resolution enable the discrimination of various land cover features, including water bodies, built-up area, barren land, vegetation. The step by step methodology followed in the present study on GEE platform is shown in Figure 2.

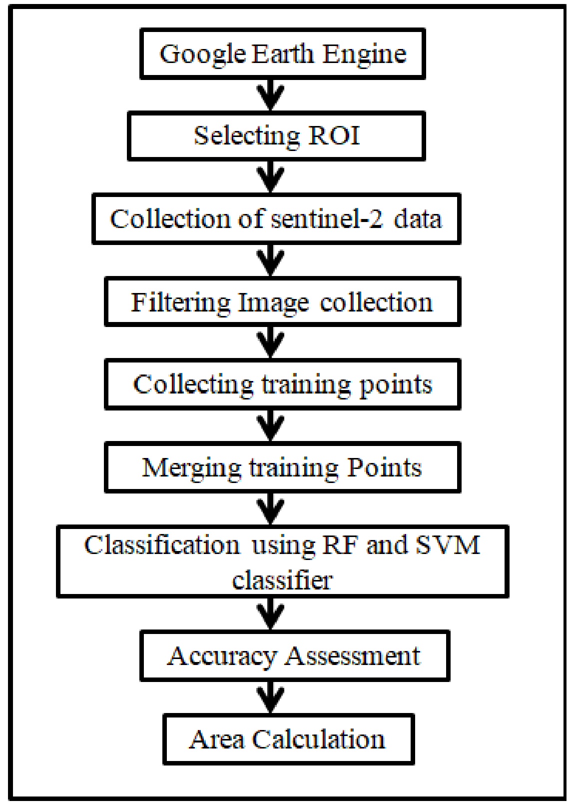

Fig. 2. Methodology for LULC classification using GEE platform

Algorithms

Support Vector Machine (SVM):

SVM is a supervised learning algorithm used for performing classification and regression tasks, in the context of LULC classification[9]. In the present days, SVM has emerged as a popular method, as it has the ability to handle high dimensional data and nonlinear decision boundaries, it can handle datasets with large no of features, which makes it suitable for LULC classification. SVM algorithms are implemented on GEE platform to classify different land cover categories based on the spectral signatures extracted from satellite image. The advantage of SVM is that it is versatile in handling both linear and non linear classification problems. With the help of Kernel functions SVM can map input features into higher dimensional spaces, allowing for more complex decision boundaries that can better capture the underlying structure of data. SVM classifiers trained in GEE can be optimized using parameter tuning and feature selection, thereby improving the accuracy of classification. The ability of SVM to handle complex datasets and non linear decision boundaries makes it a valuable tool for accurately mapping land cover type and monitoring landscape change over time.

Random Forest (RF):

RF is widely used ensemble learning method employed in regression, classification, and other machine learning task. The importance of RF in LULC classification has gained prominence due to its ability to handle high dimensional data, complex decision boundaries, and mitigate over fitting. RF works by constructing multiple decision trees during the training phase. Each tree is trained on a subset of the dataset and makes individual prediction, such that the final prediction is determined by aggregating the prediction of each tree, commonly through a majority voting mechanism for classification task[10]. This ensemble approach reduces the risk of bias, thereby enhancing the robustness and accuracy of the classifier. LULC classification on GEE platform uses RF algorithm to classify different land covers based on spectral signatures extracted from the satellite imagery.

Accuracy Assessment:

Accuracy assessment is very important step in evaluating the reliability and quality of LULC classification derived from satellite imagery. Accuracy can be checked by employing several methods, including error matrix, confusion matrices, kappa coefficient, and user accuracy. In the present study accuracy assessment is done to evaluate the performance of the models used. The training sets composed for water bodies, urban, barren land, and vegetation have been scripted in JavaScript and divided into 80% training and 20% testing datasets. The performance of classification model is evaluated using confusion matrix, it provides a detailed breakdown of predicted and actual classes, allowing for calculation of various accuracy matrices. The overall accuracy (OA) and kappa coefficient (K_C) is calculated using the following equation[11].

$$OA = \left(\frac{P_c}{P_n}\right) * 100$$

P_c - Number of pixels classified correctly

P_n - Total number of pixels

$$K_c = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+}\, x x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+}\, x x_{+i})}$$

Where,

r - Number of rows and columns in the error matrix,

x_{ii} - Number of observation in row i and column i,

x_{i+} - Marginal total of row i,

x_{+i} - Marginal total of column i, and

N = Total number of observation.

3 Results

The reseacrch dine eavluates the performance of the two machine learning techniques: SVM amd RF. In the present study four major types for land cover are classified water bodies, urban, abrren land and vegetation. The LULC classification of Bangalore district produced using RF anD SVM is shown in Figure 3. The results of RF model shown in Figure 4, revealed that 40.72 km^2 is calssified as water bosied, 153.353 km^2 as urban, 29.9 km^2 as barren land , and 580.882 km^2 as vegetation. SVm resukts shown in Figure 5, shows 41.37 km^2 is classified as water bodies, 1396.986 km^2 as urban, 155.11 km^2 as barren land, and 618.723 km^2 as surface vegetation.

Results Validation

After performing LULC classification the results obtained from RF and SVM algorithms were validated by determining overall accuracy (OA) and kappa coefficient (k_c). The OA for RF was determined to be 89.74%, representing the proportion of correctly classified pixels across all land cover categories. The K_c yielded a value of 0.87, indicating substantial agreement between observed and predicted classification.

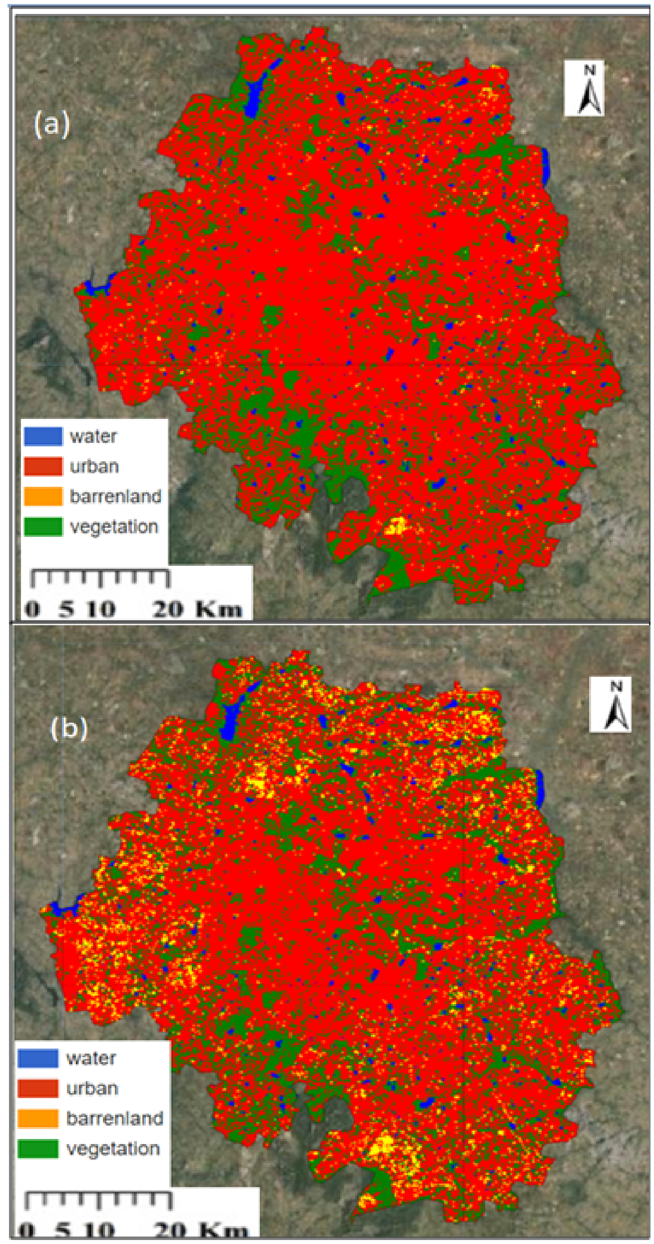

Fig. 3. (a) LULC classification using RF, (b) LULC classification using SVM

OA for SVM model was found to be higher at 92.86% compared to RF , which shows that the great proportion of pixels are correctly classified compared to RF model. The k_c yielded the value of .89, indicating a high level of agreement between observed and predicted classes. The study shows that the performance of SVM algorithm is better than RF.

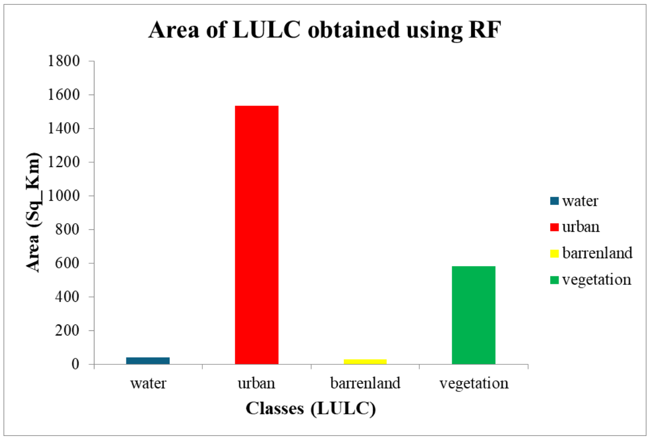

Fig. 4. Area of LULC classes using RF

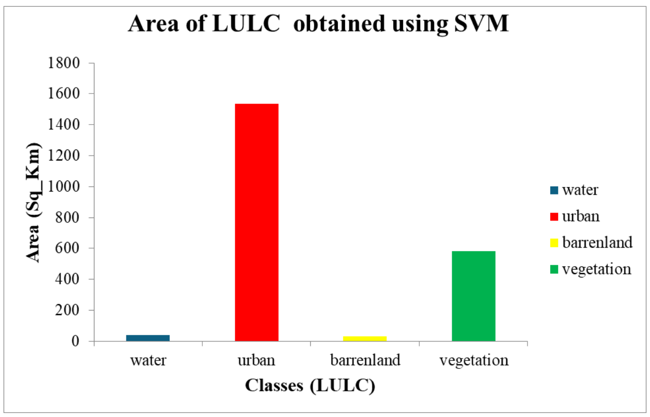

Fig. 5. Area of LULC classes using SVM

The results underscore the effectiveness of both RF and SVM models in accurately classifying LULC, where SVM showing slightly higher performance than RF. The inclusion of precision values further enriches the assessment, providing insight into the accuracy of individual classification of land cover within the model outputs.

4 Conclusion

In the study both RF and SVM algorithms demonstrated high accuracy in LULC classification, there by highlighting effectiveness in extracting meaningful information from satellite imagery. Leveraging Sentine-2 satellite image and GEE platform has revolutionized LULC mapping, by providing access to high resolution multispectral imagery and advanced processing capabilities. The accuracy of LULC mapping can be increased by providing more training points and the use of robust classification algorithms can enhance the reliability and precision of land cover classification.

Furthermore, continued advancement in remote sending technologies, machine learning algorithms, and geospatial techniques will further enhance the accuracy of LULC classification. By making use of Sentinel-2 satellite imagery, GEE platform, and advanced classification algorithms like RF and SVM, informed decisions can be made, thereby having sustainable land management practices, and also can address pressing environmental challenges on global scale.

References

1) Feras M, Baig MRU, Mustafa I, Baig H, Takaijudin MB, Zeshan T. Assessment of Land Use Land Cover Changes and Future Predictions Using CA-ANN Simulation for Selangor, Malaysia. *Assessment of Land Use Land Cover Changes and Future Predictions Using CA-ANN Simulation for Selangor, Malaysia*. 2022;14(3):402. Available from: https://doi.org/10.3390/w14030402.
2) Kobayashi N, Tani H, Wang X, Sonobe R. Crop classification using spectral indices derived from Sentinel-2A imagery. *Journal of Information and Telecommunication*. 2020;4(1):67–90. Available from: https://dx.doi.org/10.1080/24751839.2019.1694765.
3) Alshari EA, Gawali BW. Analysis of Machine Learning Techniques for Sentinel-2A Satellite Images. *Journal of Electrical and Computer Engineering*. 2022;2022:1–16. Available from: https://dx.doi.org/10.1155/2022/9092299.
4) Nguyen HTT, Doan TM, Tomppo E, McRoberts RE. Land Use/Land Cover Mapping Using Multitemporal Sentinel-2 Imagery and Four Classification Methods—A Case Study from Dak Nong, Vietnam. *Remote Sensing*. 2020;12(9):1367. Available from: https://doi.org/10.3390/rs12091367.
5) Shafizadeh-Moghadam H, Khazaei M, Alavipanah SK, Weng Q. Google Earth Engine for large-scale land use and land cover mapping: an object-based classification approach using spectral, textural and topographical factors. *GIScience & Remote Sensing*. 2021;58(6):914–928. Available from: https://dx.doi.org/10.1080/15481603.2021.1947623.
6) Atef I, Ahmed W, Abdel-Maguid RH. Modelling of land use land cover changes using machine learning and GIS techniques: a case study in El-Fayoum Governorate, Egypt. *Environmental Monitoring and Assessment*. 2023;195(6). Available from: https://dx.doi.org/10.1007/s10661-023-11224-7.
7) Akhir NSM, Salim PM, Yusoff ZM. Leveraging Google Earth Engine (GEE) for determining land use and land cover changes around Tasik Chini Malaysia. In: 9th International Conference on Geomatics and Geospatial Technology 2023 (GGT2023) ;vol. 1240 of IOP Conference Series: Earth and Environmental Science. IOP Publ. 2023;p. 012017. Available from: https://doi.org/10.1088/1755-1315/1240/1/012017.
8) Zhao Z, Islam F, Waseem LA, Tariq A, Nawaz M, Islam IU, et al. Comparison of Three Machine Learning Algorithms Using Google Earth Engine for Land Use Land Cover Classification. *Rangeland Ecology & Management*. 2024;92:129–137. Available from: https://dx.doi.org/10.1016/j.rama.2023.10.007.
9) Kundu S, Kumar V. Supervised Classification- An Overview of Machine Learning in Agricultural Practices. *Bulletin of Environment, Pharmacology and Life Sciences*. 2022;(5):398–406. Available from: https://bepls.com/spl(5)2022/70.pdf.
10) Donges N. Random Forest: A Complete Guide for Machine Learning. Mar. 08, 2024. Available from: https://builtin.com/data-science/random-forest-algorithm.
11) Aldiansyah S, Saputra RA. Comparison Of Machine Learning Algorithms For Land Use And Land Cover Analysis Using Google Earth Engine (Case Study: Wanggu Watershed). *International Journal of Remote Sensing and Earth Sciences (IJReSES)*. 2022;19(2). Available from: https://dx.doi.org/10.30536/j.ijreses.2022.v19.a3803.