GEO-EYE

RESEARCH ARTICLE

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

Salma Sultana[1], **M Inayathulla**[2]*

1 *Research Scholar, Civil Engineering Department, Bangalore University, Bangalore*
2 *Professor, Civil Engineering Department, Bangalore University, Bangalore*

OPEN ACCESS

Received: 18.08.2022

Accepted: 22.11.2022

Published: 04.12.2022

Citation: Sultana S, Inayathulla M. (2022). Precision Land Use and Land Cover Classification Using Google Earth Engine: Integrating Random Forest and Support Vector Machine Algorithms. Geo-Eye. 11(2): 9-14. https://doi.org/10.53989/bu.ge.v11i2.4

* **Corresponding author**.
inayathulla@bub.ernet.in

Funding: None

Competing Interests: None

Published By Bangalore University, Bengaluru, Karnataka

ISSN
Print: 2347-4246
Electronic: XXXX-XXXX

Abstract

Land Use and Land Cover (LULC) classification plays a pivotal role in understanding and managing environmental resources. This study presents a novel methodology utilizing sentinel satellite data in conjunction with two robust machine learning algorithms: Random Forest (RF) and Support Vector machine (SVM) on Google Earth Engine platform. Sentinel data, renowned for its high- resolution multispectral imagery, provides rich information for classification. Google Earth Engine provides access to vast geospatial datasets and computational resource, enabling effective analysis. RF and SVM, distinguished for their ability to handle complex dataset, are employed to optimize classification accuracy. A systematic work flow for preprocessing of Sentinel imagery is outlined, followed by implementation of RF and SVM algorithm 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. A compelling study demonstrates the utility of RF and SVM within GEE for precise LULC mapping, highlighting their pivotal role in supporting informed decision-making for environmental planning and conservation initiatives.

Keywords: GEE; SVM; RF; Sentinel Data; LULC

1 Introduction

Land Use and Land Cover (LULC) classification plays a pivotal role in understanding the dynamic interactions between human activities and the environment. LULC classification is fundamental for assessing changes in the Earth's surface, which is very important for environmental management. Understanding LULC dynamics provides insights into ecosystem health, habitat fragmentation, urbanization patterns, and climate change. Remote sensing techniques coupled with advanced classification algorithms are pivotal for accurate LULC mapping. LULC classifications serve as valuable inputs for land-use planning, agricultural management, water resource assessment, and policy formulation at local, regional, and global scales. Continuous monitoring and updating of LULC datasets are essential for tracking landscape transformations over time, supporting sustainable development goals and informed decision-making processes[1].

Sentinel-2 satellite imagery 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]. Pre-processing steps such as radiometric calibration, atmospheric correction, and geometric rectification are essential to ensure the quality and accuracy of Sentinel-2 imagery for LULC classification[3]. Continuous monitoring using Sentinel-2 imagery allows for temporal analysis of LULC changes over time, supporting land management decisions, environmental monitoring, and climate change studies. Sentinel-2 imagery can capture seasonal variations and changes in vegetation health, aiding in the identification of agricultural land use, forest cover, urban areas, water bodies, and other land cover types[4]. Integration of ancillary data such as digital elevation models, land use maps, and field surveys enhances the accuracy of LULC classification using Sentinel-2 imagery. Despite its advantages, challenges such as cloud cover, atmospheric interference, and spectral confusion in highly heterogeneous landscapes can affect the accuracy of LULC classification using Sentinel-2 imagery, requiring careful consideration and mitigation strategies. The open-access nature of Sentinel-2 data and the availability of cloud-based processing platforms like Google Earth Engine make LULC classification using Sentinel-2 imagery accessible and cost-effective for a wide range of applications.

Google Earth Engine (GEE) provides a powerful platform for conducting Land Use and Land Cover (LULC) analyses using a vast archive of satellite imagery. GEE offers access to an extensive collection of remote sensing datasets, including Sentinel-2, Landsat, MODIS, and more, covering a wide range of spatial and temporal resolutions suitable for LULC classification. GEE's JavaScript-based programming interface enables users to develop custom scripts for LULC classification, offering flexibility and customization options to suit specific research or application needs. GEE's collaboration features enable researchers and practitioners 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 at local, regional, and global scales[5].

RF (Random Forest) is an ensemble learning method that combines the predictions of multiple decision trees, making it robust against over fitting and capable of handling complex LULC classification tasks. SVM is a supervised learning algorithm that works well for classifying LULC categories by finding the optimal hyperplane that separates different classes in the feature space. GEE provides built-in implementations of RF and SVM algorithms, allowing users to easily apply these techniques to classify LULC categories using satellite imagery[6]. GEE's JavaScript API enables users to customize RF and SVM classification workflows, including parameter tuning and feature selection, to optimize classification accuracy for specific LULC mapping objectives. Leveraging advanced machine learning algorithms such as Random Forest (RF) and Support Vector Machines (SVM) within the Google Earth Engine (GEE) platform has become increasingly popular for accurate and efficient LULC analysis[7].

In this paper, we explore the application of RF and SVM in LULC classification on the GEE platform. We examine their capabilities, advantages, and limitations, as well as the methodologies and workflows involved. By harnessing the computational power and data resources provided by GEE, RF and SVM algorithms enable scalable and accurate LULC mapping, supporting a wide range of environmental monitoring, land management, and conservation applications.

2 Methodology

Study Area

Bangalore is the fourth largest Municipal Corporation in India. It is responsible for a population of 6.8 million in an area of 741 km2. Its boundaries have expanded more than 10 times over the last six decades. Bangalore lies in the southeast of the South Indian state of Karnataka. It is in the heart of the Mysore Plateau at an average elevation of 900 m (2,953 ft). It is located at 12°58′44″N, 77°35′30″E and covers 741 km^2 (286 sq mi). The majority of the city of Bangalore lies in the Bangalore Urban district of Karnataka and the surrounding rural areas are a part of the Bangalore Rural district (Figure 1).

Data

This study utilizes Sentinel-2 satellite imagery, which offers medium resolution multispectral data, for LULC classification. Sentinel-2 data provide valuable spectral information across various wavelengths, including visible, near-infrared, and short-wave infrared bands. The Sentinel -2 image consists of 13 spectral bands. The data were accessed from Copernicus Open Access Hub using GEE (https://earthengine.google.com/) platform[8]. The Sentinel data used is dated from 2023-01-12 to 2024-03-07

The Sentinel-2 data utilized in this study are acquired from the European Space Agency's (ESA) Sentinel missions, specifically Sentinel-2A and Sentinel-2B satellites. These satellites orbit the Earth in a sun-synchronous manner, capturing images with a revisit time of approximately 5 days, ensuring frequent and consistent monitoring of the study area. The Sentinel-2 imagery is pre-processed to ensure data quality and consistency. This includes radiometric calibration, atmo-

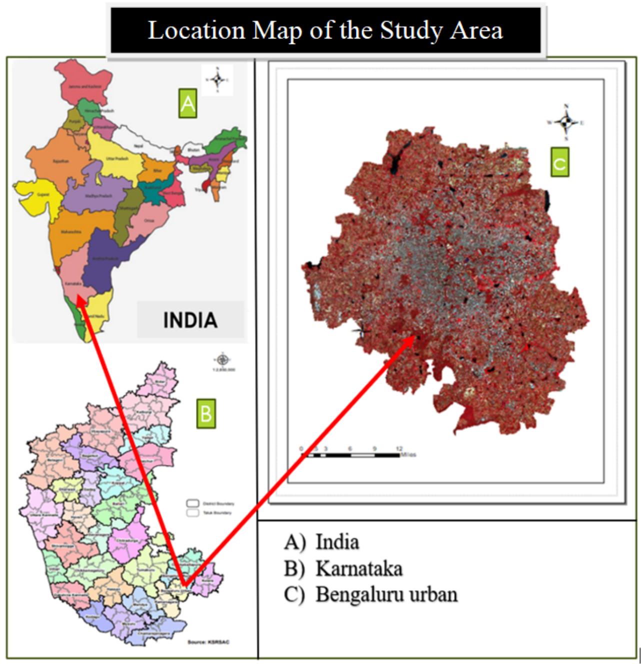

Fig. 1. Study Area

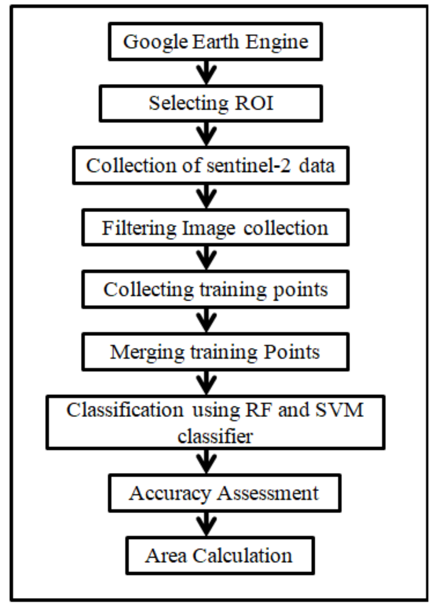

Fig. 2. Methodology for LULC classification using GEE platform

Table 1. Specification 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

spheric correction, and geometric rectification to remove distortions and artifacts from the images. Additionally, cloud masking techniques are applied to mitigate the impact of cloud cover on LULC classification accuracy. The spectral bands of Sentinel-2 image, coupled with its spatial resolution (ranging from 10 to 60 meters), enable the discrimination of various land cover features, including water bodies, urban, barren land and vegetation.

Algorithms

Support Vector Machine (SVM)

Support Vector Machines (SVM) are a powerful class of supervised learning algorithms used for classification and regression tasks. In the context of LULC classification[9]. SVM has emerged as a popular method due to its ability to handle high-dimensional data and nonlinear decision boundaries. SVM can efficiently handle datasets with a large number of features, making it suitable for LULC classification tasks that involve multispectral satellite imagery. In the case of LULC classification on the Google Earth Engine (GEE) platform, SVM algorithms are implemented to classify different land cover categories based on spectral signatures extracted from satellite imagery. One of the advantages of SVM is its versatility in handling both linear and nonlinear classification problems. By using 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 the data. SVM classifiers trained on GEE can be optimized using various techniques, such as parameter tuning and feature selection, to improve classification accuracy. Additionally, GEE's cloud-based infrastructure enables efficient processing of SVM algorithms, making it suitable for large-scale LULC classification tasks. Overall, SVM algorithms play a significant role in LULC classification by leveraging the rich spectral information provided by satellite imagery. Their ability to handle complex datasets and nonlinear decision boundaries makes SVM a valuable tool for accurately mapping land cover types and monitoring landscape changes over time.

Random Forest (RF)

Random Forest (RF) is a versatile and widely-used ensemble learning method employed in classification, regression, and other machine learning tasks. In the realm of LULC classification, RF has gained prominence due to its ability to handle high-dimensional data, complex decision boundaries, and mitigate overfitting. RF operates by constructing multiple decision trees during the training phase. Each tree is trained on a subset of the dataset and makes individual predictions. The final prediction is determined by aggregating the predictions of all the trees, commonly through a majority voting mechanism for classification tasks[(10)]. This ensemble approach enhances the robustness and accuracy of the classifier while reducing the risk of bias. In the context of LULC classification on platforms like the Google Earth Engine (GEE), RF algorithms are utilized to classify different land cover categories based on spectral signatures extracted from satellite imagery. RF provides measures of variable importance, which can help identify the most relevant features for classification. This insight is valuable for understanding the underlying relationships between spectral bands and land cover types, aiding in feature selection and model interpretation.

Accuracy Assessment

Accuracy assessment is a crucial step in evaluating the reliability and quality of land cover classifications derived from remote sensing data. Several methods can be employed for accuracy assessment, including error matrices, confusion matrices, kappa statistics, and user/producer accuracy. The Accuracy assessment was performed to evaluate the performance of the model. The composed points for water bodies, urban, barren land and vegetation have been scripted in JavaScript-based code and divided in 80% training and 20% testing datasets. A built-in method in GEE confusion matrix is a tabular representation used to evaluate the performance of a classification model. It provides a detailed breakdown of the predicted and actual classes, allowing for the calculation of various accuracy metrics. The overall accuracy (OA) and kappa coefficient (Kc) can be calculated using following equations[(11)].

$$OA = \left(\frac{Pc}{Pn}\right) * 100$$

Pc – number of pixels classified correctly

Pn - total number of pixels

$$\mathrm{k} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+}xx_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+}xx_{+i})}$$

Where,

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

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

x_{i+} - the marginal total of row i,

x_{+I} - the marginal total of column i, and

N = the total number of observation.

3 Results

The current research evaluates the performance of three machine learning techniques: SVM, and RF, for the LULC classification of Bangalore District, Karnataka, India. These algorithms were applied for the classification of four major. The LULC map of the Bangalore district of India produced by RF is shown in Figure 3 (a) and Figure 3 (b) shows LULC classification using SVM. The classification results revealed that 40.72 km^{2}is classified as water, 1534.353 km^2 as urban, 29.9 km^2 as barrenland, and 580.882 km^2 as vegetation. The results of the RF model for the LULC map of the current study are shown in Figure 4 (a) The SVM results of the LULC map of the study area illustrated that 41.37 km^2 is classified as water, 1369.986 km^2 as built up, 155.777km^2 as barrenland, and 618.723 km^2) as surface vegetation of the Bangalore. The results are shown in Figure 4 (b).

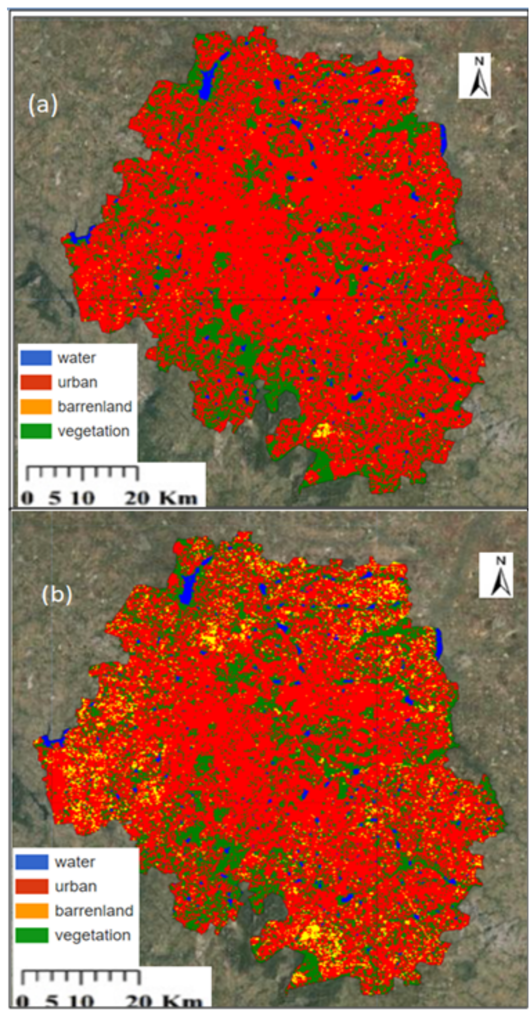

Fig. 3. 2(a) LULC classification using RF 2(b) LULC classification using SVM

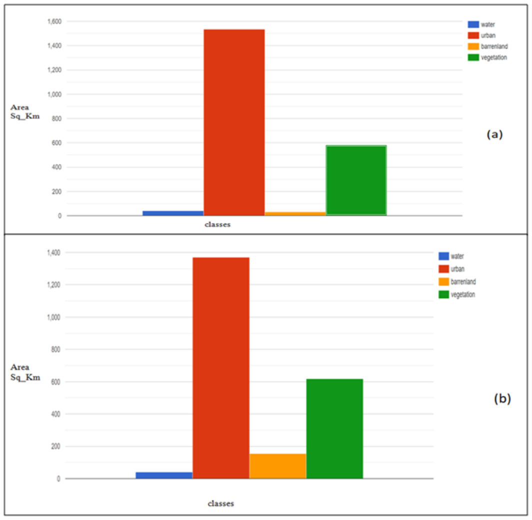

Fig. 4. (a) Area of LULC classes using RF (b) Area of LULC classes using SVM

Results Validation

In the validation of the Random Forest (RF) model for Land Use and Land Cover (LULC) classification, the Overall Accuracy (OA) was determined to be 89.74%, denoting the proportion of correctly classified pixels across all land cover categories. Additionally, the Kappa Coefficient (Kc) yielded a value of 0.87, indicating substantial agreement between observed and predicted classifications beyond what would be expected by chance alone. Furthermore, while the precision values for individual land cover categories were not explicitly provided, it can be inferred that the RF model demonstrated commendable precision in accurately identifying specific land cover classes among all instances classified as such.

Similarly, for the Support Vector Machine (SVM) model, the Overall Accuracy was found to be higher at 92.86%, showcasing a greater proportion of correctly classified pixels compared to the RF model. Moreover, the Kappa Coefficient (Kc) reached 0.89, indicating a high level of agreement between observed and predicted classifications, approaching near-perfect agreement. Additionally, the precision of the SVM model in LULC classification was notable, demonstrating the accuracy of identifying specific land cover categories among all instances classified as such.

These validation results underscore the effectiveness of both the RF and SVM models in accurately classifying LULC types, with the SVM model showing slightly higher performance. The inclusion of precision values would further enrich the assessment, providing insights into the accuracy of individual land cover classifications within the models' outputs.

4 Conclusion

Land Use and Land Cover (LULC) classification is pivotal for understanding the Earth's surface dynamics and supporting various environmental management initiatives. Leveraging Sentinel-2 satellite data and the Google Earth Engine (GEE) platform has revolutionized LULC mapping, offering access to high-resolution multispectral imagery and advanced processing capabilities.

In our study, we applied Random Forest (RF) and Support Vector Machine (SVM) models to classify LULC categories, utilizing the rich spectral information provided by Sentinel-2 data. Both RF and SVM algorithms demonstrated high accuracy in land cover classification, highlighting their effectiveness in extracting meaningful information from satellite imagery.

The availability of more training points and the use of robust classification algorithms like RF and SVM are crucial for improving accuracy in LULC mapping. By incorporating additional training data and optimizing model parameters, we can enhance the reliability and precision of land cover classifications, providing valuable insights into landscape dynamics and environmental changes.

Moving forward, continued advancements in remote sensing technology, machine learning algorithms, and geospatial data analysis techniques will further enhance the accuracy and applicability of LULC classification methodologies. By harnessing the power of Sentinel-2 data, GEE platform, and advanced classification techniques like RF and SVM, we can support informed decision-making processes, facilitate sustainable land management practices, and address pressing environmental challenges on a global scale.

References

1) Baig MF, Mustafa MRU, Baig I, Takaijudin H, Zeshan MT. Assessment of Land Use Land Cover Changes and Future Predictions Using CA-ANN Simulation for Selangor, Malaysia. MDPI Water. 2022. 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://doi.org/10.1080/24751839.2019.1694765.
3) Alshari AE, Gawali BW. Analysis of Machine Learning Techniques for Sentinel-2A Satellite Images. *J Electr Comput Eng*. 2022. Available from: https://www.hindawi.com/journals/jece/2022/9092299/.
4) Nguyen HTT, Doan TM, Tomppo E, McRoberts R. Land Use/Land Cover Mapping Using Multitemporal Sentinel-2 Imagery and Four Classification Methods-A Case Study from Dak Nong. *MDPI Remote Sens*. 2020;12(9). 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://doi.org/10.1080/15481603.2021.1947623.
6) Atef WI, & Ramadan HA, Abdel-Maguid. Modelling of land use land cover changes using machine learning and GIS techniques: a case study in El-Fayoum Governorate. *Monit Assess*. 2023;195. Available from: https://doi.org/10.1007/s10661-023-11224-7.
7) Leveraging Google Earth Engine (GEE) for determining land use and land cover changes around Tasik Chini Malaysia. IOP Publ. 2023. Available from: https://ui.adsabs.harvard.edu/abs/2023E&ES.1240a2017A/abstract.
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://doi.org/10.1016/j.rama.2023.10.007.
9) Kundu S, Kumar V. Supervised Classification- An Overview of Machine Learning in Agricultural Practices. *Bull Environ Pharmacol Life Sci*. 2022;(5). Available from: https://www.researchgate.net/publication/372658837_Supervised_Classification-An_Overview_of_Machine_Learning_in_Agricultural_Practices.
10) Donges N. Random Forest: A Complete Guide for Machine Learning. 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). *Int J Remote Sens Earth Sci*. 2022;19(2). Available from: https://jurnal.lapan.go.id/index.php/ijreses/article/view/3803.