PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol 12(2): 1-10
https://doi.org/10.14719/pst.6260

HORIZON
e-Publishing Group
HePG

REVIEW ARTICLE

Assessment of land use and land cover mapping using object-based classification techniques for the eastern districts of Tamil Nadu

Karthikkumar A[1], Kumaraperumal R[1]*, Pazhanivelan S[2], Muthumanickam D[1], Ragunath KP[2], Selvakumar S[2] & Kamaleshkanna S[1]

[1]Department of Remote Sensing and GIS, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

[2]Centre for Water and Geospatial Studies, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

*Email: kumaraperumal.r@tnau.ac.in

OPEN ACCESS

ARTICLE HISTORY
Received: 22 November 2024
Accepted: 07 January 2025
Available online
Version 1.0 : 13 April 2025
Version 2.0 : 23 April 2025

Check for updates

Additional information
Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE
Karthikkumar A, Kumaraperumal R, Pazhanivelan S, Muthumanickam D, Ragunath KP, Selvakumar S, Kamaleshkanna S. Assessment of land use and land cover mapping using object-based classification techniques for the eastern districts of Tamil Nadu. Plant Science Today. 2025; 12(2): 1-10. https://doi.org/10.14719/pst.6260

Abstract

LULC (Land use and land cover) mapping is crucial for understanding environmental monitoring, supporting sustainable development and managing natural resources. This study evaluated the accuracy of object-based LULC classification using Sentinel-2 data and machine learning classifiers in the Ariyalur, Perambalur and Mayiladuthurai districts of Tamil Nadu during the kharif season of 2023. OBIA (Object-based image analysis) clusters pixels based on their spectral and spatial characteristics, utilizing segmentation to generate masks that effectively represent the image content. The OBIA methodology involves multiresolution segmentation using eCognition software to delineate homogeneous image objects based on spectral, spatial and contextual characteristics. Several widely used machine learning algorithms, including Random forest (RF), Support vector machine (SVM), Decision Tree (DT), Naive bayes (NB) and k-nearest Neighbor (k-NN), were evaluated to improve classification accuracy. The classification results varied across the districts, with the RF algorithm consistently demonstrating high performance. The Perambalur and Mayiladuthurai RF achieved an overall accuracy of 88 %, with a kappa coefficient of 0.76 and 83 % and a kappa coefficient of 0.66. In Ariyalur, the DT model was used, with an accuracy of 85 % and a kappa coefficient of 0.70. The NB and k-NN classifiers achieved lower accuracies in all districts. In contrast, the RF algorithm was the most reliable for LULC classification in these areas, highlighting its strength and efficiency in accurately identifying complex land cover patterns.

Keywords

eCognition; land use; land cover; sentinel-2 data; segmentation

Introduction

Remote sensing is the most critical land use and land cover (LULC) mapping application, enabling efficient and large-scale analysis (1,2). Land use land cover maps are highly valuable for analysis. This classification provides a foundation for change detection studies, enabling the analysis of land cover changes over time. By comparing LULC maps from different periods, researchers can detect patterns such as deforestation, urban expansion and agricultural intensification. These insights are essential for understanding landscape transformation dynamics driven by natural processes and human activities (3). OBIA effectively identifies small objects and complex land cover types, reducing misclassification rates associated with spectral confusion (4). Machine learning techniques, such as neural networks, fuzzy systems, inheritable algorithms, intellectual agents and SVM, have recently been developed to improve accuracy (5-7). Machine learning algorithms (MLAs) are essential for accurately classifying land use land cover (LULC) using remote sensing data. Machine learning algorithms (MLAs) are necessary for accurately classifying LULC using remote sensing data. Numerous studies have shown that LULC classification using medium and low-resolution satellite images suffers from spectral and geographical limitations, reducing classification accuracy (8). To overcome these challenges, ML methods have gained importance in achieving high-precision LULC classification. However, the accuracy of results depends significantly on the choice of ML model, training data and input parameters, as different ML methods yield varying precision levels (9).

Several studies have used machine learning algorithms, including RF, SVM and Classification and Regression Trees (CART), to generate accurate LULC maps. RF achieves the highest overall accuracy of 97.02 % in the uMngeni catchment area, outperforming SVM and artificial neural networks (ANN) (10). Like the Kabul LULC map from google earth engine (GEE), RF demonstrated superior performance with an accuracy of 93.99 % and 94.42 % for Landsat-8 and Sentinel-2 imagery, respectively (11). The combination of Sentinel-2 data with OBIA and ML significantly enhances the accuracy of LULC classification. Some studies have demonstrated this improvement, with research on small farmlands achieving up to 95 % accuracy using kNN and other classifiers (12). Similarly, OBIA and ML techniques used in tropical forest mapping in Brazil reported an overall accuracy of 96 % (13).

Materials and Methods

The study area includes the districts of Ariyalur, Perambalur and Mayiladuthurai (Fig. 1), which are in the South Indian state of Tamil Nadu. The Perambalur District is located at approximately 11.2366° N latitude and 78.9070° E longitude, followed by the Ariyalur District at 11.2286° N latitude and 79.0370° E longitude and the Mayiladuthurai District at 11.1965° N latitude and 79.4298° E longitude. The climates in these districts are variable but are predominantly hot and dry for approximately eight months of the year. The challenging topography refers to areas with complex landforms such as mountains, steep slopes and varied vegetation, which make accurate land cover classification difficult. These regions often require advanced techniques to address issues like mixed pixels, overlapping spectral signatures and rapid changes in terrain. The climate contributes to the rich biodiversity and diverse landscape observed in the LULC of these areas. To minimize the complexity of the landscape, information in the study area has been divided into major categories, such as growing crops (rice, cotton, maize and cashew), forests, urban land, water bodies and mountains. Sentinel-2 (Level 1C) provides data in 13 bands across the visible, near-infrared and shortwave infrared spectra (ESA, 2023) (14). The multispectral Sentinel-2 data, with less than 5 % cloud cover from June to October 2023, was accessed via the Copernicus Data Space Browser and utilized in the current study. The Sentinel-2 satellite data, including 13 spectral bands (green 560 nm), Band 4 (red 665 nm) and Band 8 (NIR 842 nm), were utilized in this study. The satellite data were downloaded from the Copernicus Open Access Hub and optimized through preprocessing techniques. The preprocessing steps included (i) compositing bands to generate an RGB image (Fig. 2), (ii) mosaicking to create a new raster image by combining multiple downloaded scenes and (iii) extraction with the mask to obtain a Sentinel-2 image of the three districts with reduced cloud cover. The mission-wide swath, acceptable spatial resolution (10 m), multispectral capabilities (13 spectral bands) and frequent revisit interval (10 days at the equator with one satellite or 5

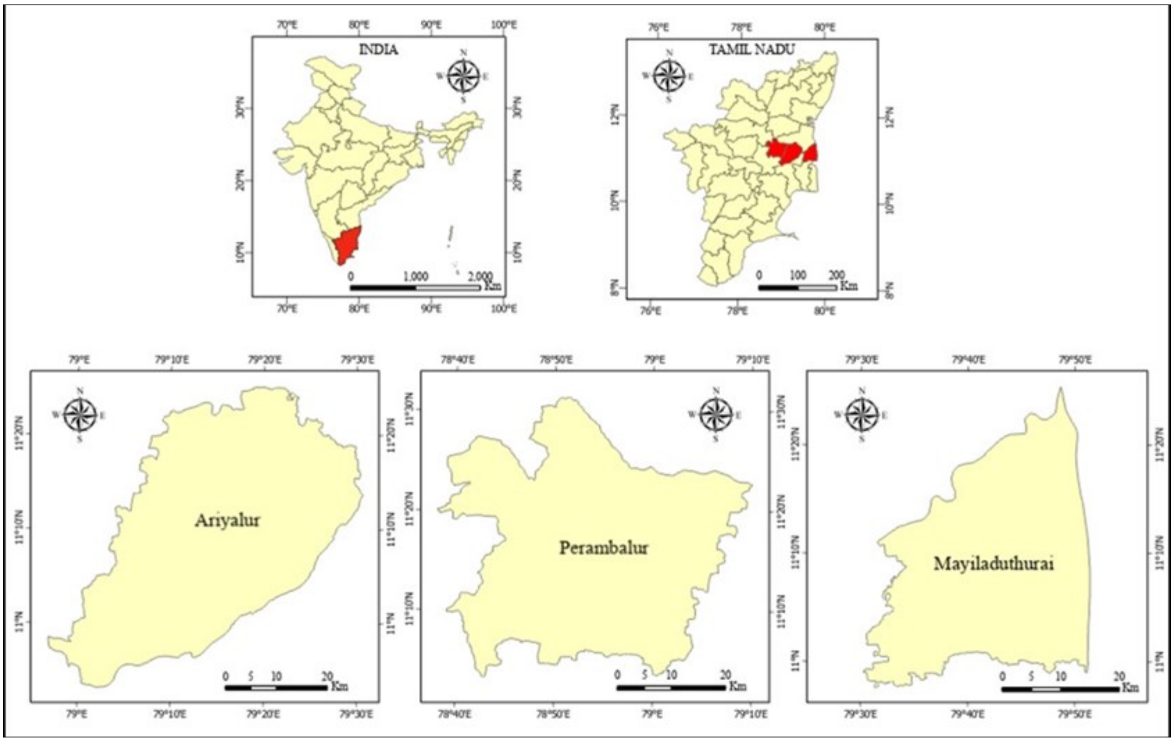

Fig. 1. Study area map.

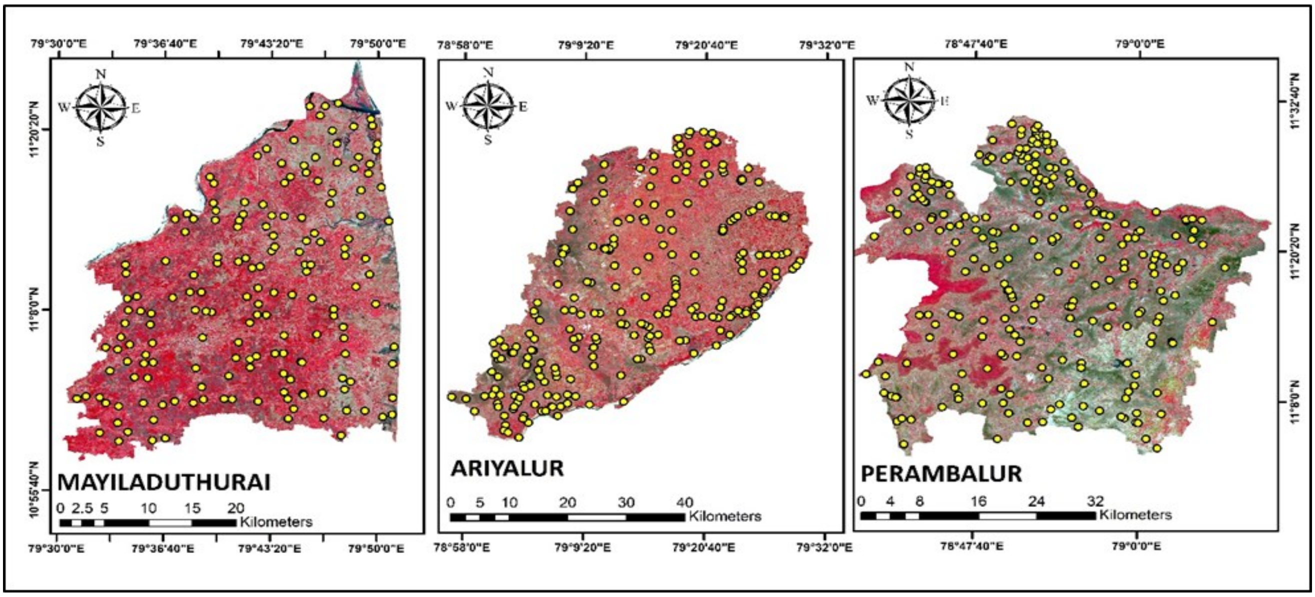

Fig. 2. Sentinel-2 False-colour composite data for the study area.

days with two satellites) enable a range of applications, including vegetation monitoring during growing seasons, forest management, land cover change detection and natural disaster response (15).

Machine learning algorithm models support vector machine (SVM)

The SVM is a supervised learning method widely used for LULC classification and regression analysis. SVM also finds applications in various domains, such as face recognition, text categorization and image classification and is grounded in the theoretical framework of structural risk minimization (16). The SVM classifier constructs a hyperplane, or a set of hyperplanes, in a high- or infinite-dimensional space, depending on the number of features, to effectively classify spatial data, even when training data are limited (17). SVM employs an iterative approach to construct the optimal hyperplane that separates patterns in the training dataset. It is achieved by selecting extreme data points (support vectors) crucial in defining the decision boundary. Key parameters, such as the cost parameter, kernel function and gamma, influence the identification of support vectors. Using nonlinear kernel functions, SVM can efficiently handle nonlinear classification tasks (18). However, the SVM complex training mechanism heavily depends on the dataset size, making it less suitable for large datasets. Despite this limitation, SVM remains highly effective for complex datasets due to its kernel-based approach. The performance of SVM relies significantly on the choice of kernel functions and the fine-tuning of parameters, such as the regularization parameter. Achieving high accuracy with SVM requires careful selection of model parameters and kernel functions to avoid overfitting or underfitting (19).

Random forest (RF)

The RF algorithm, widely recognized in supervised learning, constructs a forest of multiple decision ntrees, making it applicable to regression and classification tasks. It is based on the tree algorithm, which generates and combines multiple trees to produce the desired output. RF employs ensemble methods, such as bagging, to improve predictive performance (20). The algorithm divides the initial dataset into in-bag samples (approximately two-thirds of the data) and out-of-bag (OOB) samples (the remaining one-third). The in-bag samples are used to train individual decision trees, while the OOB samples are utilized to estimate the OOB error. The core principle of RF is that an ensemble of multiple bootstrap-aggregated classifiers achieves higher accuracy and robustness than a single classifier (21). RF aggregates predictions from each tree and the final output is determined based on the majority vote of these predictions. By averaging values from randomly selected data subsets, RF enhances predictive accuracy and reduces overfitting (22). RF is particularly crucial for LULC classification due to its high accuracy and ability to handle complex datasets. Studies have demonstrated that RF can significantly improve classification outcomes by utilizing various data inputs, such as spectral indices and training sample sizes. For instance, increasing the size of the training samples notably enhances overall accuracy (OA) and reduces uncertainty in LULC mapping (23).

K-nearest neighbor (KNN)

The kNN machine learning algorithm for classification and regression. It predicts the outcome by finding the 'k' closest data points (neighbours) to the input data and using them to make decisions. For classification, KNN assigns the class most common among the neighbours, while for regression, it predicts the average value of the neighbours. The basic theory behind kNN is to identify a group of k samples in the calibration dataset that are closest to an unknown sample, typically based on distance functions. The label (class) of the unknown sample is then determined by calculating the average of the response variable (i.e., the class attributes of the kNN) (24). As a result, a low *k* value produces an intricate decision boundary, while a higher *k* value generates a more generalized boundary. Additionally, as the number of training samples increases, the k-NN classification will require more computational resources since it does not rely on a pre-trained model (25). The kNN algorithm is a robust and effective LULC method, especially when processing high-dimensional remote sensing data. The performance of kNN can be improved by incorporating multilabel classification, enabling more detailed and realistic land use modelling. The choice of the 'k' value in kNN significantly affects its performance. The traditional kNN uses a constant 'k' value, which may only be optimal for some spatial units, potentially reducing the prediction accuracy (26).

Decision tree (DT)

The DT classification method is an algorithm with a tree-like structure, where each node represents a decision based on data attributes (27).In a DT, data are split into subsets, with each internal node representing a feature, each branch corresponding to a decision rule and each leaf node signifying an outcome or prediction. DTs are computationally efficient and do not rely on statistical assumptions about data distribution, making them well-suited for analyzing remote sensing data. They can process data at various measurement scales and are widely accessible through numerous software implementations. DTs are valuable tools for establishing thresholds and developing decision rules for land cover classification in remote sensing. As non-parametric models, they are suitable for diverse landscapes and their simplicity ensures straightforward interpretation (28,29). DTs are particularly useful for selecting the most influential variables and identifying thresholds for different land cover classes, further enhancing their applicability (30). While DTs are commonly used in classification tasks for accuracy, they are prone to overfitting and may perform poorly when handling continuous values. DTs are formulated using *if-then-else* rules to define classification criteria (31).

Naive bayes (NB)

Bayesian methods treat parameters as random variables with a known prior distribution. They are based on the Bayes classifier, which assumes that a particular feature's presence (or absence) is independent of any other feature's presence (or absence). The classifier estimates the probabilities of different attribute value occurrences for various classes in a training set and these probabilities are then used to classify recall patterns. (NB) simplifies computational complexity by reducing calculations to basic multiplications of probabilities, making it particularly effective for datasets with many attributes. During classification, it uses training data to calculate the highest probability for a given input and assigns it the corresponding class label (32). NB is relatively tolerant of noise, which makes it suitable for LULC classification, even when the data contain attribute noise caused by cloud cover or class noise resulting from location and labelling errors in crowd-sourced datasets (33). However, Bayesian methods, including NB, require a complete probability model, which can be a limitation when the evidence is incomplete or insufficiently provided by the data (34).

Ground truth data

Extensive crop exploration was conducted in June and July 2023 to understand the cropping systems during the *kharif* season. Sampling points were collected for each crop grown in the region, including cotton, rice, maize, coconut and cashew, as well as for other categories such as forestry, urban areas, water bodies and barren land. These sampling points are distributed throughout the entire study area. Nine hundred ground truth locations (Ariyalur 275, Perambalur 353, Mailaduthurai 271) (Fig. 2) were identified and documented for classification and verified through visual interpretation using satellite imagery from Google Earth Pro.

Software used

ArcGIS10.1 and QGIS 3.38.0 were used to handle spatial datasets and perform geographical information system (GIS) operations. eCognition 9.1 software was employed for object-based image classification and Excel was used to derive statistical analyses.

Object-based segmentation method

eCognition software utilizes a segmentation method called multiresolution segmentation, which is a bottom-up, region -based merging algorithm. The process starts by considering each pixel as a separate object. These objects are then iteratively merged into more significant segments based on a local homogeneity criterion that measures the similarity between neighbouring objects. At each step, the pair of objects with the smallest increase in the homogeneity criterion is merged. This merging continues until the rise in homogeneity surpasses a user-defined scale parameter, which acts as the threshold for stopping the segmentation. The scale parameter is a critical component in multiresolution segmentation. A formal method for determining the optimal scale factor, using the Estimation of Scale Parameter (ESP) tool, was proposed in prior studies (35). For this study, the ESP tool identified optimal values of 40 for Scale, 0.3 for Shape and 0.6 for compactness, as given in Fig. 3. Using these parameters, the multiresolution segmentation algorithm in eCognition Developer 9.1 was applied to divide the images into spectrally homogeneous objects (36).

Object feature extraction

After the feature creation process is segmented, image segments generate image objects. In this study, several spectral features were extracted. These included the mean of three bands, the false colour composite bands and shape texture features. Rule-based techniques and algorithm-based classification have enhanced land cover mapping, while feature space optimization has led to higher accuracy in agricultural applications (37). Each image object extracts the spatial and spectral feature information such as mean, ratios, indices, size, shape and texture of the object. Object-related feature values were developed using spectral and spatial indexes (38). LULC feature extraction in object-based methods involves extracting spatial and structural features from image objects, which are then classified using one-dimensional classifiers. This process requires the manual design of features, unlike Convolutional Neural Networks (CNNs), which automatically learn spatial features during training (4).

The sample selection process involves identifying representative objects for training and classification. These objects can be randomly selected manually, using expert knowledge, or automatically through random sampling. The selected objects are then used to train machine learning classifiers, ensuring accurate land cover classification. The trained models are applied to classify each object into its

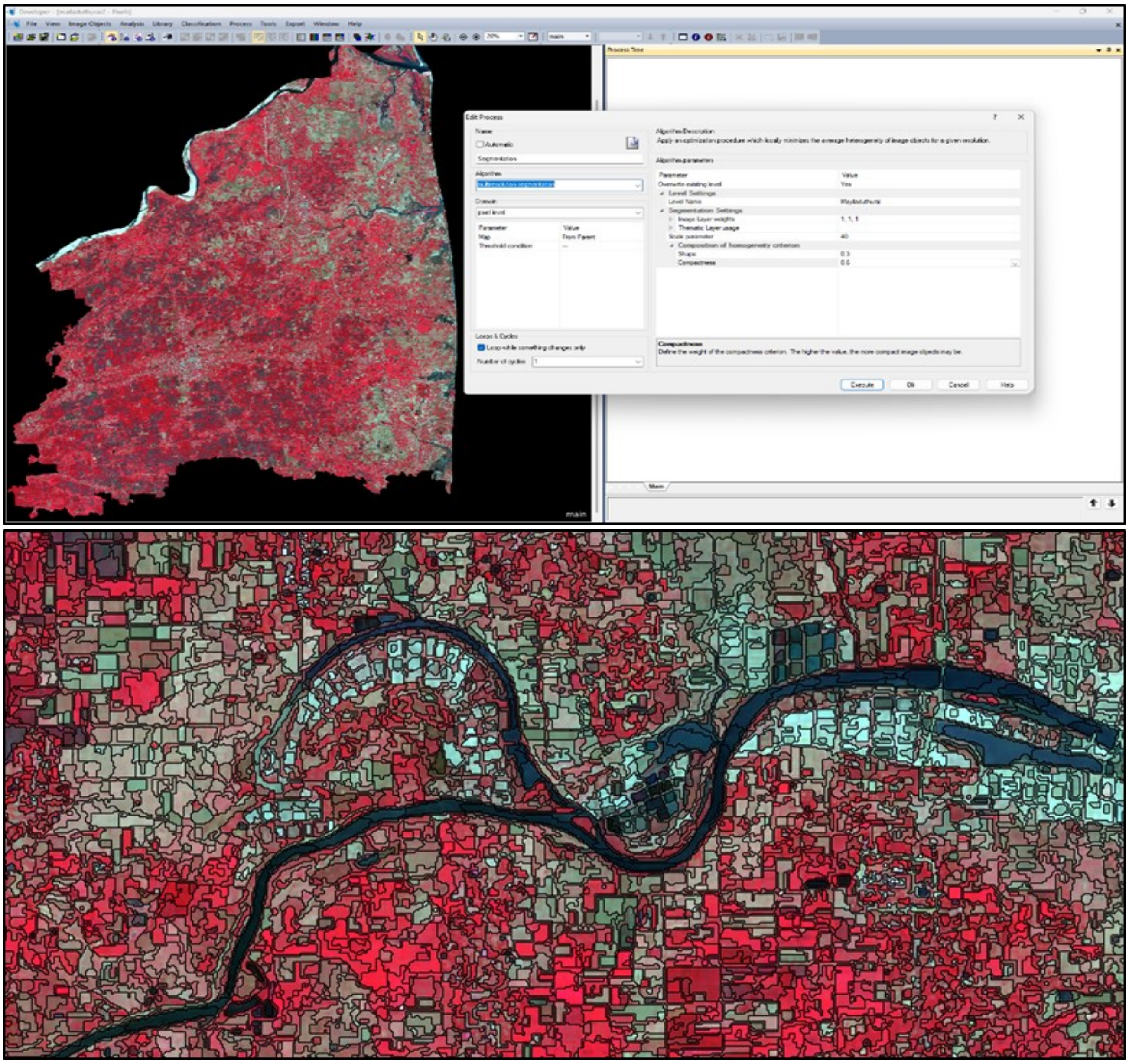

Fig. 3. eCognation segmentation process.

respective land cover class. Define the classes for the class hierarchy and create a new hierarchy that contains the target classes. Colour codes are assigned to each class to make visualization easier. RF, SVM, DT, kNN and NB. Machine learning algorithms classify objects based on their extracted features. Each segment (object) is labelled with a class, resulting in a classified map. Finally, the classified objects were compared with ground truth data using confusion matrices, kappa statistics and other accuracy metrics.

Accuracy assessment and validation

Accuracy assessment or validation is a crucial step in remote sensing data processing, as it evaluates the informational value of the generated data for end users (39). In LULC cover mapping, accuracy assessment is essential not only to gauge the quality and reliability of the map but also to understand thematic uncertainty and its potential impact on users (40). An error matrix and kappa statistics were employed to evaluate classification accuracy. The error matrix is constructed based on matched and mismatched pixels and it serves as the basis for calculating metrics such as the kappa coefficient, producer accuracy, user accuracy and overall accuracy (41)

Results

Different machine learning algorithm performances

The LULC map highlights significant classes such as rice, maize, cashew, settlements, barren land and water bodies. The classification results were compared based on the corresponding accuracy assessments, including overall accuracy and the kappa coefficient (Table 1). Various machine learning models were evaluated in the Ariyalur district (Fig. 4) to assess their effectiveness in LULC classification. The DT model demonstrated the highest performance, achieving an overall accuracy of 85 % and a kappa coefficient 0.7. It was closely followed by the RF model, which recorded an accuracy of 84 % and a kappa of 0.68. The SVM model showed moderate performance, attaining an overall accuracy of 78 % with a kappa coefficient of 0.56. The NB model yielded an accuracy of 73 % and a kappa coefficient of 0.46.

In contrast, the k-NN model had the lowest

Table 1. Individual classification results of efficient performing algorithm

Model	Overall accuracy (%)	kappa Coefficient
Ariyalur		
Model	**Overall accuracy (%)**	**kappa Coefficient**
Random Forest	84	0.68
Decision Tree	85	0.7
Support Vector Machine	78	0.56
Naive Bayes	73	0.46
k-Nearest Neighbour (k-NN)	71	0.42
Perambalur		
Model	**Overall accuracy (%)**	**kappa Coefficient**
Random Forest	88	0.76
Support Vector Machine	83	0.66
Decision Tree	82	0.64
k-Nearest Neighbor (k-NN)	74	0.48
Naive Bayes	72	0.44
Mayiladuthurai		
Model	**Overall accuracy (%)**	**kappa Coefficient**
Random Forest	83	0.66
Decision Tree	80	0.6
k-Nearest Neighbor (k-NN)	75	0.5
Support Vector Machine	74	0.48
Naive Bayes	73	0.46

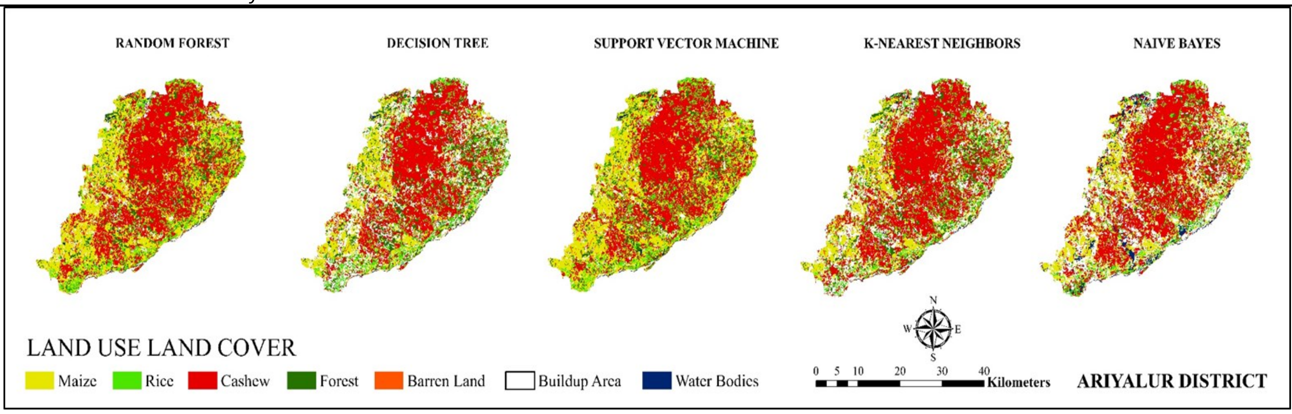

Fig. 4. LULC map for Ariyalur district.

performance, with an accuracy of 71 % and a kappa coefficient of 0.42. In the Perambalur district (Fig. 5), machine learning models were evaluated for their LULC classification performance. The RF model achieved the highest overall accuracy at 88 % with a kappa coefficient of 0.76, making it the most effective model in this analysis. The SVM model followed, with an accuracy of 83 % and a kappa coefficient of 0.66. The DT model had an overall accuracy of 82 % and a kappa coefficient of 0.64. Moreover, the k-NN model had an accuracy of 74 % and a kappa coefficient of 0.48 and the NB model had the lowest performance, with an accuracy of 72 % and a kappa coefficient of 0.44.

In the Mayiladuthurai district (Fig. 6), the performance of various machine learning models for LULC classification was assessed. The RF model achieved the highest overall accuracy, reaching 83 % with a kappa coefficient of 0.66. The DT model followed, with an accuracy of 80 % and a kappa coefficient of 0.6. The k-NN model recorded an accuracy of 75 % with a kappa of 0.5. The SVM model showed a slightly lower performance, with an accuracy of 74 % and a kappa coefficient of 0.48.

In contrast, the NB model had the lowest performance, with an accuracy of 73 % and a kappa of 0.46. Across the districts of Perambalur, Ariyalur and Mayiladuthurai, machine learning models demonstrated varying effectiveness in LULC classification. The RF model consistently achieved high accuracy, performing best in Perambalur (88 %) and Mayiladuthurai (83 %), whereas the DT and SVM models achieved competitive results in Ariyalur. Overall, RF emerged as the most effective model, followed closely by DT and SVM,

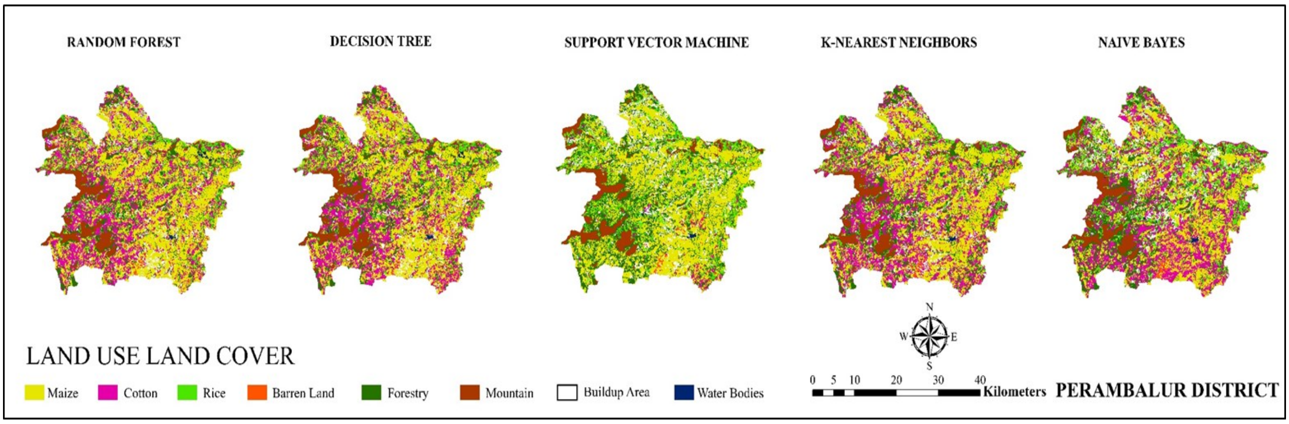

Fig. 5. LULC map for Perambalur district.

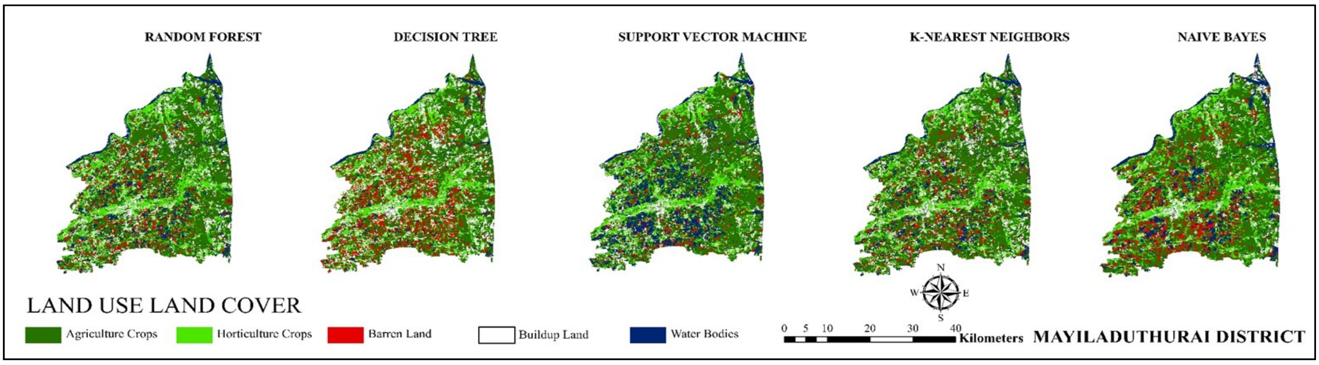

Fig. 6. LULC map for Mayiladuthurai district.

depending on the district. However, lower performances were generally observed with the NB and k-NN classifiers across all districts.

This study employed the confusion matrix approach to validate and evaluate the classification accuracy of various machine learning models. The metrics, including overall accuracy and the kappa coefficient, were calculated using the following equations (42). The accuracy of different object-based supervised classification methods was assessed using an error matrix with validation samples. An error matrix is generated based on the agreement and disagreement of classified pixels and it serves as the basis for calculating the kappa coefficient, user's accuracy, producer's accuracy and overall accuracy (41). Across the districts of Perambalur, Ariyalur and Mayiladuthurai, machine learning models demonstrated varying effectiveness in LULC classification. The RF model consistently achieved high accuracy, performing best in Perambalur (88 %) and Mayiladuthurai (83 %), while the DT and SVM models delivered competitive results in Ariyalur *(Supplementary Table 1A, B, C).* Overall, RF emerged as the most effective model, followed closely by DT and SVM, depending on the district. However, all districts generally observed lower performances with the NB and k-NN classifiers.

Discussion

This study utilized Sentinel-2 (S2) data from the 2023 Kharif season to perform LULC classification for Ariyalur, Perambalur and Mayiladuthurai districts in Tamil Nadu. OBIA, combined with segmentation techniques, was applied to group pixels into meaningful objects based on shape, size, texture and spectral properties. The optimal segmentation scale and careful feature selection enhanced classification accuracy while minimizing computational costs. The results revealed that the RF model consistently achieved the highest classification accuracy across all three districts, with a peak performance of 88 % in Perambalur. The DT and SVM models delivered competitive results, particularly in Ariyalur. At the same time, the k-NN and NB classifiers demonstrated relatively lower accuracies and kappa coefficients across the regions.

The study highlights the effectiveness of integrating OBIA with ML techniques, notably the RF algorithm, for LULC mapping. Similar findings have been reported in other studies, such as those using Sentinel-2 data for LULC classification in Parana, Brazil (43). The ability of RF to incorporate spatial context through OBIA significantly enhances classification results compared to traditional pixel-based methods (44). This superior performance can be attributed to RF's robustness against noise, its ability to handle high-dimensional and nonlinear data and its ensemble nature, which minimizes overfitting. Other studies further validate the advantages of OBIA combined with ML. For example, integrating textural data into object-based classification significantly improved land use classification accuracy using QuickBird imagery in Karbala, Iraq (45). Similarly, RF achieved 90.3 % accuracy (kappa 0.87), outperforming SVM for land cover classification in the Novosibirsk region (46). On the other hand, image segmentation and classification were performed using the SNIC and RF algorithms, supported by an efficient parallel processing workflow (47). The LULC map achieved an overall accuracy of 91.7 %, with the highest accuracy for water and wetlands and the lowest for rangelands.

Several studies emphasize the importance of model selection and parameter tuning for achieving optimal accuracy. The object-based RF algorithm achieved the highest overall accuracy (89.74 %) for land use classification in Wulian County, China (27). Conversely, a NB classifier using geotagged Flickr images achieved an accuracy of 87.94 %, indicating its potential for automated feature extraction (48).

Despite RF dominance, SVM has also shown competitive performance in certain studies. For instance, a study in El Fayoum Governorate, Egypt, reported that SVM achieved a kappa coefficient of 0.916, outperforming RF (49). Similarly, research in Kerala, India, indicated that RF attained an accuracy of 89.5 %, while SVM and regression tree models delivered slightly lower accuracies (50). These variations underscore the influence of regional, temporal and dataset-specific factors on ML model performance. The quality of training samples remains a critical factor influencing LULC classification accuracy. For instance, atmospheric variations, surface conditions and illumination changes can introduce classification errors, as highlighted in studies (51). The various classifiers exhibit differing performance outcomes depending on climatic and geographic conditions (52). Furthermore, Different classifiers exhibit varying performance across diverse climatic zones and geographical locations (53).

Parameter tuning complexity also affects the consistency and reproducibility of ML results. Feature selection improves classification accuracy by incorporating spectral, spatial and textural information (54). The high-precision LULC mapping depends on an optimized model setup, high-quality training samples and appropriate input parameters (55).

For algorithm improvement, future studies could integrate GLCM (Gray-Level Co-occurrence Matrix-derived textural features) (e.g., contrast, correlation, homogeneity, entropy) with spectral data can significantly enhance object-based image analysis, particularly for complex land cover types. Incorporating auxiliary datasets such as vegetation indices, soil details and elevation, combined with feature fusion techniques, can improve classification accuracy. Ensuring high-quality training samples, optimizing segmentation and utilizing multi-temporal data will enhance the consistency and reliability of LULC studies.

Conclusion

The study highlights the effectiveness of various machine learning models in LULC classification across the districts of Ariyalur, Perambalur and Mayiladuthurai, with the RF model consistently achieving the highest overall accuracy. The RF and DT models performed well in most cases. However, challenges such as accurately distinguishing mixed land cover classes and handling complex terrains were evident, particularly for models like NB and k-NN, which showed lower accuracy and kappa values. LULC classification includes challenges in accurately distinguishing mixed land cover classes, especially in heterogeneous landscapes. Classifier performance can vary significantly across regions due to topography and data quality differences.

Additionally, models like NB and k-NN tend to perform poorly in complex terrains and computational demands increase with high-resolution imagery. LULC studies are essential for understanding topography and supporting Sustainable Development Goals (SDGs), serving as the foundational layer for numerous earth-based applications. While most studies highlight the efficiency of object-based LULC classifiers, this study focused on assessing their performance in delineating the topography of the Eastern districts of Tamil Nadu, characterized by complex terrain. The classifiers demonstrated their utility in accurately segregating LULC at regional and sub-regional levels. The thematic maps generated from this study can support precise policy decisions and promote the judicious use of resources in these districts.

Acknowledgements

The authors thank the professor, head, staff and colleagues of the Department of Remote Sensing and GIS and the Centre for Water and Geospatial Studies, Tamil Nadu Agricultural University, Coimbatore.

Authors' contributions

KA conducted the material preparation, data collection and original draft writing. KR contributed to the conceptualization. SP handled the software work. MD generated the map. RKP was responsible for collecting ground truth data. SS and KS participated in sequence alignment, manuscript editing and visualization. All authors have read and approved the final version of the manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

References

1. Vizzari M, Hilal M, Sigura M, Antognelli S, Joly D. Urban-rural-natural gradient analysis with CORINE data: An application to the metropolitan France. Landsc Urban Plan. 2018;171:18–29. https://doi.org/10.1016/j.landurbplan.2017.11.005
2. Shalaby A, Tateishi R. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. Appl Geogr. 2007;27(1):28–41. https://doi.org/10.1016/j.apgeog.2006.09.004
3. Carranza-García M, García-Gutiérrez J, Riquelme JC. A framework for evaluating land use and land cover classification using convolutional neural networks. Remote Sens. 2019;11(3):274. https://doi.org/10.3390/rs11030274
4. Selvaraj R, Amali DGB, Bessie G. Assessment of object-based classification for mapping land use and land cover using google earth. Glob NEST J. 2023;25(7):131–38.
5. Yang X. Parameterizing support vector machines for land cover classification. Photogramm Eng Remote Sens. 2011;77(1):27–37. https://doi.org/10.14358/PERS.77.1.27
6. Kulkarni AD, Lowe B. Random forest algorithm for land cover classification. Int J Recent Innov Trends Comp Comm. 2016;4(3):58 - 63
7. Breiman L. Random forests. Mach Learn. 2001;45:5–32. https://doi.org/10.1023/a:1010933404324
8. Pal S, Talukdar S. Assessing the role of hydrological modifications on land use/land cover dynamics in Punarbhaba river basin of Indo-Bangladesh. Environ Dev Sustain. 2020;22(1):363–82. https://doi.org/10.1007/s10668-018-0205-0
9. Dutta D, Rahman A, Paul SK, Kundu A. Changing pattern of urban landscape and its effect on land surface temperature in and around Delhi. Environ Monit Assess. 2019;191(9):551. https://doi.org/10.1007/s10661-019-7645-3
10. Bhungeni O, Ramjatan A, Gebreslasie M. Machine-learning algorithms for mapping LULC of the uMngeni catchment area, KwaZulu-Natal. Remote Sens. 2024;16(12):2219. https://doi.org/10.3390/rs16122219
11. Ahmadi K. Assessment of the accuracy of various machine learning algorithms for classifying urban areas through google earth engine: A case study of Kabul city, Afghanistan. Eur J Artif Intell Mach Learn. 2024;3(3):1–8. https://doi.org/10.24018/ejai.2024.3.3.40
12. Aziz N, Minallah N, Hasanat M, Ajmal M. Geographic object-based image analysis for small farmlands using machine learning techniques on multispectral Sentinel-2 data. Proc Pak Acad Sci Phys Comput Sci. 2024;61(1):41–49. https://doi.org/10.53560/PPASA(60-1)795
13. Junior C, Araki H, Macedo CR. Object-Based Image Analysis (OBIA) and Machine Learning (ML) applied to tropical forest mapping using Sentinel-2. Can J Remote Sens. 2023;49(1):2259504. https://doi.org/10.1080/07038992.2023.2259504
14. Ecosystem CDS. Copernicus data space ecosystem Europe's eyes on earth [Internet]. 2023. Available from: https://browser.dataspace.copernicus.eu.
15. Vuolo F, Żółtak M, Pipitone C, Zappa L, Wenng H, Immitzer M, et al. Data service platform for Sentinel-2 surface reflectance and value-added products: System use and examples. Remote Sens. 2016;8(11):938. https://doi.org/10.3390/rs8110938

16. Chaudhary D, Vasuja ER. A review on various algorithms used in machine learning. Int J Sci Res Comput Sci Eng Inf Technol. 2019;5 (2):915–20. https://doi.org/10.32628/cseit1952248
17. Yimer SM, Bouanani A, Kumar N, Tischbein B, Borgemeister C. Comparison of different machine-learning algorithms for land use land cover mapping in a heterogenous landscape over the Eastern Nile river basin, Ethiopia. Adv Space Res. 2024;74(5):2180–99. https://doi.org/10.1016/j.asr.2024.06.010
18. Rudrapal D, Subhedar M. Land cover classification using support vector machine. Int J Eng Res. 2015 Sep 24;V4. https://doi.org/10.17577/ijertv4is090611
19. Adugna T, Xu W, Fan J. Comparison of random forest and support vector machine classifiers for regional land cover mapping using coarse resolution FY-3C images. Remote Sens. 2022;14(3):574. https://doi.org/10.3390/rs14030574
20. Saini R, Ghosh SK. Crop classification on single date Sentinel-2 imagery using random forest and suppor vector machine. Int Arch Photogramm Remote Sens Spat Inf Sci. 2018;5:683–88. https://doi.org/10.5194/isprs-archives-xlii-5-683-2018
21. Smith HL, Biggs PJ, French NP, Smith ANH, Marshall JC. Out of (the) bag-encoding categorical predictors impacts out-of-bag samples. PeerJ Comput Sci. 2024;10:e2445.https://doi.org/10.7717/peerj-cs.2445
22. Bhungeni O, Ramjatan A, Gebreslasie M. Evaluating machine-learning algorithms for mapping lulc of the umngeni catchment area, KwaZulu-Natal. Remote Sens. 2024;16(12):2219. https://doi.org/10.3390/rs16122219.
23. Phinzi K, Ngetar NS, Pham QB, Chakilu GG, Szabó S. Understanding the role of training sample size in the uncertainty of high-resolution LULC mapping using random forest. Earth Sci Inform. 2023;16(4):3667–77.https://doi.org/10.1007/s12145-023-01117-1
24. Wei C, Huang J, Mansaray LR, Li Z, Liu W, Han J. Estimation and mapping of winter oilseed rape LAI from high spatial resolution satellite data based on a hybrid method. Remote Sens. 2017;9 (5):488. https://doi.org/10.3390/rs9050488
25. Maselli F, Chirici G, Bottai L, Corona P, Marchetti M. Estimation of Mediterranean forest attributes by the application of k-NN procedures to multitemporal Landsat ETM+ images. Int J Remote Sens. 2005;26(17):3781–96. https://doi.org/10.1080/01431160500166433
26. Sun D, He M, Zhang D. Land use type identification on the basis of drone imagery and object-oriented methods. In: Bilal M, Tosti F, editors. In:Second International Conference on Geographic Information and Remote Sensing Technology (GIRST 2023). Qingdao, China: SPIE; 2023. p. 101. https://doi.org/10.1117/12.3007579
27. Prabowo NW, Siregar VP, Agus SB. Klasifikasi habitat bentik berbasis objek dengan algoritma support vector machines dan decision tree menggunakan citra multispektral spot-7 di pulau harapan dan pulau kelapa. J Ilmu Dan Teknol Kelaut Trop. 2018;10(1):123–34. https://doi.org/10.29244/jitkt.v10i1.21670
28. Myint SW, Gober P, Brazel A, Grossman-Clarke S, Weng Q. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sens Environ. 2011;115(5):1145–61. https://doi.org/10.1016/j.rse.2010.12.017
29. Grąbczewski K. Techniques of decision tree induction. In: Krzysztof Grąbczewski K, editor. Meta-learning in decision tree induction. Cham: Springer International Publishing; 2014. p. 11–117. https://doi.org/10.1007/978-3-319-00960-5_2
30. Coronel TY, Benitez-Paez F, Padró JC, Drenkhan F, Timaná ME, FallaValdez V. Optimization of remotely sensed variables selection for land cover mapping. SSRN; 2024. https://doi.org/10.2139/ssrn.4853907
31. Veenadhari S, Bharat Mishra Dr, Singh DrC. Soybean productivity modelling using decision tree algorithms. Int J Comput Appl. 2011;27(7):11–15. https://doi.org/10.5120/3314-4549
32. Pradhan R, Ghose MK, Jeyaram A. Land cover classification of remotely sensed satellite data using bayesian and hybrid classifier. Int J Comput Appl. 2010;7(11):1–4. https://doi.org/10.5120/1295-1783
33. Johnson BA, Iizuka K. Integrating OpenStreetMap crowdsourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna de Bay area of the Philippines. Appl Geogr. 2016;67:140–49. https://doi.org/10.1016/j.apgeog.2015.12.006
34. Comber AJ, Law AN, Lishman JR. A comparison of Bayes', Dempster–Shafer and Endorsement theories for managing knowledge uncertainty in the context of land cover monitoring. Comput Environ Urban Syst. 2004;28(4):311–27. https://doi.org/10.1016/S0198-9715(03)00013-9
35. Drăguţ L, Tiede D, Levick SR. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. Int J Geogr Inf Sci. 2010;24(6):859–71. https://doi.org/10.1080/13658810903174803
36. Gao Y, Marpu P, Niemeyer I, Runfola DM, Giner NM, Hamill T, Pontius Jr RG. Object-based classification with features extracted by a semi-automatic feature extraction algorithm-SEaTH. Geocarto Inter. 2011;26(3):211–26. https://doi.org/10.1080/10106049.2011.556754
37. Darwish A, Leukert K, Reinhardt W. Image segmentation for the purpose of object-based classification. In: 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings, Toulouse, France; 2003. p. 2039–41. https://doi.org/10.1109/IGARSS.2003.1294332
38. Nussbaum S, Niemeyer I, Canty MJ. Seath - a new tool for automated feature extraction in the context of object-based image analysis. In: 1st International Conference on Object-based Image Analysis (OBIA 2006), Salzburg, Austria. OISPRS Archives; p. 1–6.
39. Tilahun A. Accuracy assessment of land use land cover classification using google earth. Am J Environ Prot. 2015;4 (4):193. https://doi.org/10.11648/j.ajep.20150404.14
40. Czaplewski RL. Accuracy assessment of maps of forest condition: Statistical design and methodological considerations. In: Wulder MA, Franklin SE, editors. Remote sensing of forest environments: Concepts and case studies. Boston (MA): Springer; 2003. p. 115–40. https://doi.org/10.1007/978-1-4615-0306-4_5
41. Congalton RG. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ. 1991;37(1):35–46. https://doi.org/10.1016/0034-4257(91)90048-B
42. Basheer S, Wang X, Farooque AA, Nawaz RA, Liu K, Adekanmbi T, et al. Comparison of land use land cover classifiers using different satellite imagery and machine learning techniques. Remote Sens. 2022;14(19):4978. https://doi.org/10.3390/rs14194978
43. Singh R, Singh P, Drews M, Kumar P, Singh H, Gupta KA, et al. A machine learning-based classification of LANDSAT images to map land use and land cover of India. Remote Sens Appl Soc Environ. 2021;100624. https://doi.org/10.1016/j.rsase.2021.100624
44. Tassi A, Gigante D, Modica G, Di Martino L, Vizzari M. Pixel vs. object-based landsat 8 data classification in google earth engine using random forest: the case study of Maiella national park. Remote Sens. 2021;13(12):2299. https://doi.org/10.3390/rs13122299
45. Jaber HS, Shareef MA, Merzah ZF. Object-based approaches for land use-land cover classification using high resolution quick bird satellite imagery (a case study: Kerbela, Iraq). Geod Cartogr. 2022;48(2):85–91. https://doi.org/10.3846/gac.2022.14453
46. Garafutdinova L, Kalichkin V, Fedorov D. Object-oriented classification of remote sensing earth images using machine. Bull NSAU Novosib State Agrar Univ. 2024;37–47. https://

doi.org/10.31677/2072-6724-2024-71-2-37-47

47. 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. GI Science Remote Sens. 2021;58(6):914–28. https://doi.org/10.1080/15481603.2021.1947623
48. Sitthi A, Nagai M, Dailey M, Ninsawat S. Exploring land use and land cover of geotagged social-sensing images using naive bayes classifier. Sustain. 2016;8(9):921. https://doi.org/10.3390/su8090921
49. 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. Environ Monit Assess. 2023;195(6):637. https://doi.org/10.1007/s10661-023-11224-7
50. Sundar KSP, Deka PC. Spatio-temporal classification and prediction of land use and land cover change for the Vembanad lake system, Kerala: a machine learning approach. Environ Sci Pollut Res. 2022;29(57):86220–36. https://doi.org/10.1007/s11356-021-17257-0
51. Sathyaseelan M, Ghosh SK, Ojha C. Satellite-based LULC change with special emphasis on SDG using machine learning - A case study on the upper Ramganga Basin. In: IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium. Athens, Greece: IEEE; 2024. p.4159–63. https://doi.org/10.1109/IGARSS53475.2024.10640733
52. Khan S, Bhardwaj A, Sakthivel M. Accuracy assessment of land use land cover classification using machine learning classifiers in google earth engine; a case study of Jammu district. Int arch photogramm remote sens spat Inf Sci. 2024;XLVIII-4–2024:263–68. https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-263-2024
53. Zafar Z, Zubair M, Zha Y, Fahd S, Nadeem A. Performance assessment of machine learning algorithms for mapping of land use/land cover using remote sensing data. Egypt J Remote Sens Space Sci. 2024;27(2):216–26. https://doi.org/10.1016/j.ejrs.2024.03.003
54. Thanh Noi P, Kappas M. Comparison of random forest, k-nearest neighbor and support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors. 2017;18(1):18. https://doi.org/10.3390/s18010018
55. Jasim BS, Jasim OZ, AL-Hameedawi AN. Evaluating land use land cover classification based on machine learning algorithms. Eng Technol J. 2024;42(5):557–68. https://doi.org/10.30684/etj.2024.144585.1638