PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol 11(sp4): 01–09
https://doi.org/10.14719/pst.5165

HORIZON e-Publishing Group

RESEARCH ARTICLE

Object-based image analysis and machine learning for mapping cashew plantations in Ariyalur district, Tamil Nadu

Karthikkumar Alaguvel[1], Kumaraperumal Ramalingam[1]*, Pazhanivelan Sellaperumal[2], Muthumanickam Dhanaraju[1], Ragunath Kaliyaperumal[2], Selvakumar[2] & Nivasraj Moorthi[1]

[1]Department of Remote Sensing and Geographic Information System, Tamil Nadu Agricultural University, Coimbatore 641 003, India
[2]Centre for Water and Geospatial Studies, Tamil Nadu Agricultural University, Coimbatore 641 003, India

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

OPEN ACCESS

ARTICLE HISTORY
Received: 20 September 2024
Accepted: 14 October 2024
Available online
Version 1.0 : 23 December 2024
Version 2.0 : 28 February 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
Alaguvel K, Ramalingam K, Sellaperumal P, Dhanaraju M, Kaliyaperumal R, Selvakumar, Moorthi N. Object-based image analysis and machine learning for mapping cashew plantations in Ariyalur district, Tamil Nadu. Plant Science Today.2024;11(sp4):01-09. https://doi.org/10.14719/pst.5165

Abstract

An object-based image analysis (OBIA) approach provides a comprehensive method for delineating homogeneous segments based on spectral characteristics, geometry, and spatial imagery structures. The present study utilizes OBIA and machine learning (ML) techniques to map cashew plantations in Ariyalur district of Tamil Nadu, India. Sentinel-2 Multi-Spectral Instrument (MSI) imagery, acquired during the 2023 kharif season, was employed as the primary data source due to its high spatial and spectral resolution, suitable for detailed land cover mapping. The OBIA methodology involved multi-resolution segmentation using eCognition software to delineate homogeneous image objects based on spectral, spatial, and contextual characteristics. Machine learning algorithms, including random forest (RF), support vector machine (SVM), and decision tree (DT), were evaluated to improve classification accuracy. The SVM demonstrated the best superior performance, achieving an overall accuracy of 92.1% and a kappa coefficient of 0.85. The results underscore the effectiveness of ML techniques in conjunction with OBIA for precise cashew plantation mapping while contributing to improved land use/land cover mapping, agricultural resource management, and sustainable development within the region.

Keywords

image segmentation; machine learning; OBIA; rule set; sentinel 2 MSI

Introduction

Cashew (*Anacardium occidentale* L.) is one of India's top foreign exchange-earning perennial horticultural crops. In addition to generating valuable foreign exchange earnings of approximately Rs 4,000 crore annually, cashews support the rural economy by giving 11.5 million people, particularly women, stable jobs in the agricultural and processing sectors. (Source: Directorate of Cashew and Cocoa Development). In Tamil Nadu, cashew plantations are primarily located in Ariyalur, Cuddalore, Pudukkottai, and Sivagangai districts. Accurate crop area mapping is vital for supporting policy-making, ensuring food security, and providing critical data for resource allocation and agricultural planning. These maps assist in yield prediction, crop production statistics, crop rotation records, soil productivity, crop stress analysis, damage evaluation, and agricultural activity monitoring (1).

Remote sensing and geospatial technologies are essential for accurate crop area estimation, facilitated by high-frequency satellite images and the development of advanced classification algorithms. Over the last decade, researchers have conducted numerous studies to enhance the accuracy and efficiency of crop classification (2).

Pixel-based classification methods have shown limitations in inaccurately mapping horticultural crops due to their significant spatial heterogeneity (3). Various studies indicate that traditional pixel-based classifiers do not use spatial information to classify imagery, making OBIA more efficient than pixel methods (4). Object-Based Image Analysis has proven effective for mango mapping, with studies indicating high classification accuracies. For instance, indices use achieved an overall accuracy of 89% in mango orchard mapping (5). While traditional visual interpretation techniques have been useful, advanced object-based classifiers have demonstrated promising results for classifying high-resolution data for natural resource mapping (6-8). Using Sentinel series data at a regional scale has improved crop species mapping, reducing errors in crop area estimation and enhancing agricultural practices (9). Machine learning algorithms have emerged as a powerful tool for land cover analysis, enabling precise crop classification. Integrating ML techniques enhances understanding of environmental dynamics and supports sustainable resource management. Machine learning algorithms, such as SVMs and RFs, improve land use land cover (LULC) classification accuracy by analyzing satellite imagery to effectively identify and categorize different land cover types (10). The Sentinel data used in crop discrimination and crop area estimation produced an accuracy of 4 to 5 percent greater than Landsat data (11). The high spatial resolution (10-60 m.) allows for precise mapping of agricultural fields, effectively identifying crop types and monitoring growth stages (12). The amalgamation of Sentinel-2 data with ML and object-based classification methodologies has exhibited considerable promise for diverse applications, including land cover categorization and habitat mapping. This synergistic approach combines remote sensing data with advanced algorithms to enhance the accuracy and efficiency of environmental monitoring. Accurately mapping and monitoring these plantations is essential for effective agricultural management, sustainable development, and informed decision-making regarding land use and resource allocation. However, pixel-based remote sensing methods often lack the spatial and spectral accuracy needed for cashew plantations (13). This study uses an OBIA approach combined with ML techniques to provide precise land cover mapping. The combination of ML and OBIA enhances data analysis and decision-making in land use management, facilitating accurate monitoring, predictive modeling, and improved resource allocation for sustainable agriculture. OBIA techniques coupled with ML classifiers have demonstrated high accuracy in land use classification and achieved kappa values of 0.90 in various urban environments (14). The high-resolution capabilities of Sentinel-2 imagery, along with the strengths of OBIA, enable detailed identification of homogeneous segments of cashew plantations based on their spectral and spatial characteristics. The primary aim of this study is to map cashew plantations in the Ariyalur district and evaluate the performance of various ML algorithms.

Classifier model

The different ML algorithms used for plantation mapping (15). These algorithms can effectively classify satellite images based on training data. Effective crop recognition can be accomplished through the use of classification algorithms such as RF, SVM, and DT, which use remote sensing data, image processing techniques, and ML. The integration of these methods significantly enhances accuracy and efficiency in agricultural applications (16).

Several classification algorithms are available for application on segmented objects, with optimal features for object-based crop recognition. With the rapid development of ML in recent years, image classification algorithms and tools have become increasingly accessible, making their application in crop recognition more widespread. These explore three typical ML algorithms for image object classification RF, SVM, and DT.

Decision tree (DT)

In DT, the root of the tree is used to predict the class level of a dataset. The advantage of the DT is its ease of understanding and presentation, requiring minimal data preparation (17). Each branch of the tree represents the results of the tests conducted at the internal node. A branch corresponds to a specific value or range of values for the tested attribute. The leaf nodes represent the final decision or prediction after the data have passed through all the relevant internal nodes. These nodes contain the predicted class or value. The DT algorithm is used to solve regression and classification problems. Studies have demonstrated that DT algorithms are central to Developing object-based rule set classification. Decision tree algorithms for object-based land cover classification. This approach is particularly useful for creating detailed and interpretable rule sets for classifying different land cover types from Landsat 8 imagery (18).

Decision tree has several disadvantages, including the likelihood of generating an inefficient solution and overfitting. To address the latter, tree pruning is typically employed, involving the removal of one or more gap layers (i.e., branches). Pruning reduces classification accuracy for training data but increases accuracy for dealing with unknowns in general (19).

Random forest (RF)

Random forest is a collective classifier consisting of numerous autonomous DT that perform RF classification tasks. The RF generates multiple random DT through bootstrapping and random feature selection, leveraging predictions from previously constructed trees. Every tree is trained using a subset of the input data. The remaining data, known as out-of-bag samples, are used for unbiased validation. The final classification result is determined by majority voting across all trees. Random forest is a highly robust classifier that effectively addresses overfitting and noise. It can also efficiently process high-dimensional data

(20). RF can struggle with high-dimensional data, leading to overfitting and misclassification, particularly when distinguishing between similar land use and land cover classes.

Support vector machine (SVM)

Support vector machine is a supervised learning classification and regression analysis method. This technique establishes an ideal decision boundary referred to as a hyperplane. The SVM algorithm determines the most effective hyperplane that differentiates various classes. This hyperplane separates the data from the training data, which separates the data from the training data based on the minimum misclassification of pixels based on the minimum misclassification of pixels. The SVM adopted an iterative approach to create the optimal hyperplane and differentiate the patterns of the training dataset. The effectiveness of the SVM model depends on crucial parameters such as kernel functions, cost, and gamma, which are critical in determining the support vectors and handling nonlinear data patterns (21). The performance of SVM is highly dependent on the choice of kernel function and the tuning of its parameters. Incorrect selections can lead to suboptimal classification results (22).

Materials and Methods

Study area

Ariyalur District, a significant cashew-growing region in Tamil Nadu, is geographically located between 10.54° and 11.30° N latitude and 78.40° to 79.30° E longitude (Fig. 1). Lacking major natural boundaries, the district is bordered by the Kollidam river to the north and the Vellar river to the south. The district's soils are predominantly ferruginous, with a clayey texture that varies from red on the surface to yellow on the lower horizon. These well-drained soils are generally free of salt and carbonates, with relatively low organic matter, nitrogen, and phosphorus content but adequate potash and lime. Red clay soils are prevalent in the Sendurai, T. Palur, Andimadam, and Jeyankondam blocks, while black soils are found in the Thirumanur and Ariyalur blocks.

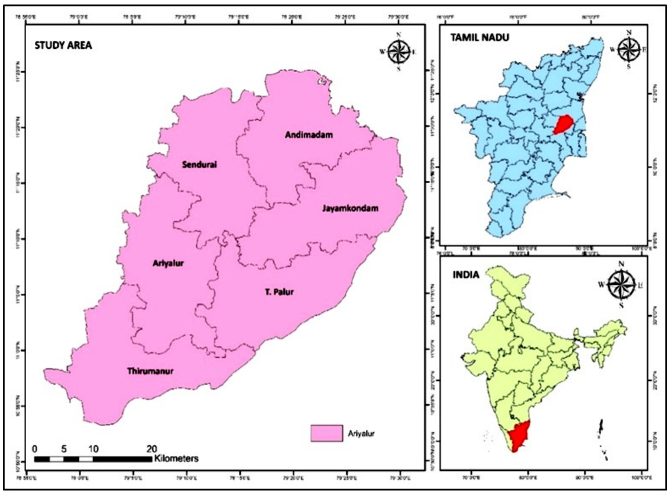

Fig. 1. Study area map.

Datasets and geoprocessing tools

Sentinel-2 data, with its high spatial resolution of 10 meters, is ideal for crop mapping. Multispectral Sentinel-2 imagery, acquired during the 2023 kharif season and exhibiting less than 5% cloud cover, was accessed through the Copernicus Data Space Ecosystem (23). The Sentinel-2 bands used for the false color composite (FCC) include B8 near Infrared (NIR) at 842 nm, B4 (red) at 665 nm, and B3 (green) at 560 nm. The FCC enhances vegetation, rendering it a bright red color, and facilitating the differentiation between vegetated and non-vegetated areas. This combination is beneficial for analyzing vegetation health and crop conditions, as healthy vegetation strongly reflects NIR (24).

ArcGIS10.1 and QGIS 3.38.0 were utilized to handle spatial datasets and perform Geographical Information System (GIS) operations. eCognition 9.0 software was employed for object-based image classification, and Excel was used for deriving statistical analyses.

Land use land cover (LULC) classification

Land use land cover studies are critical because they form the base layer for many earth-based applications. LULC maps are valuable for analyzing landscape composition and structure, identifying changes in the landscape, and assessing transformations across environmental gradients (25). Crop level mapping is necessary to yield forecasts, create statistics from agricultural data, track crop rotation, map soil productivity, identify crop stress, assess crop damage, and track farming activity (26). This research generated a LULC map (Fig. 2) using an RF approach with maize, rice, cashew, forest, built-up areas, and water bodies.

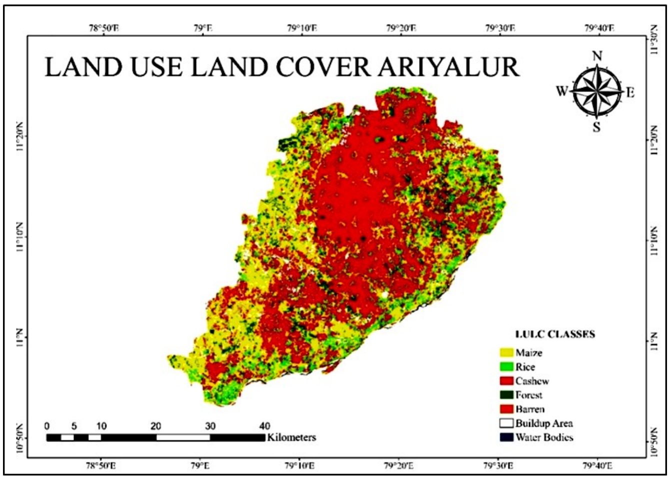

Fig. 2. Land use land cover map for the Ariyalur district.

Object-based segmentation

The input Sentinel-2 data was processed into a false color composite using the image mixture layer in eCognition software. The process involved segmentation using the multiresolution segmentation technique, a bottom-up region merging approach classified as a region-based algorithm. This technique begins by treating each pixel as a separate object and then progressively merges smaller objects into larger segments in iterative steps (27).

Segmentation creates spectrally homogeneous objects associated with real-world features on the ground (28). Previous research has highlighted the challenges in determining optimal segmentation parameters (29). In image segmentation, selecting an appropriate segmentation scale is crucial, ensuring that the resulting objects accurately represent specific land use types. However, mixed pixels can complicate segmentation, particularly in areas with small objects or heterogeneous landscapes.

The segmentation parameters, such as scale (Sc), shape (Sh), and compaction (Cm), are typically determined through trial-and-error approaches (30). However, a formal method for determining optimal scale factors involves the Estimation of Scale Parameter (ESP) tool (31). The ESP tool suggested 70 for Sc, 0.3 for Sh, and 0.6 for Cm in this research. The segmentation parameters, including scale (Sc), shape (Sh), and compaction (Cm), are often determined through trial-and-error methods (Fig. 3). Using these scale parameters, the multiresolution algorithm in eCognition Developer 9.0 was employed to segment the images into spectrally homogeneous objects.

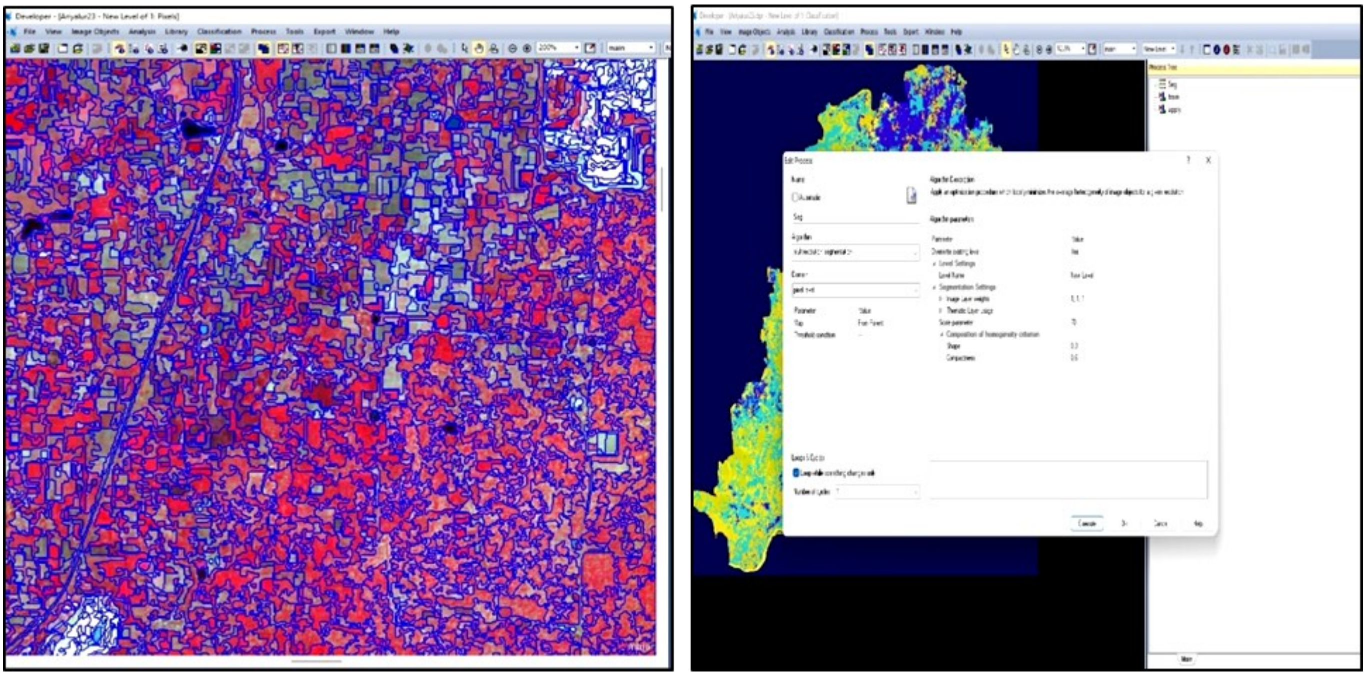

Fig. 3. eCognation Segmentation Process.

Object feature extraction

Each segmented object was analyzed to extract various attributes, including spectral (mean or median values for each spectral band), textural (roughness, contrast, or homogeneity), geometric (shape and compactness), and contextual features (relationship with neighboring objects). Feature extraction is a crucial process in many fields, simplifying the identification and classification of relevant data. Recent advancements in eCognition software have significantly improved feature extraction capabilities in urban and agricultural environments. Rule-based techniques and algorithm-based classification have enhanced land cover mapping, while feature space optimization has led to higher accuracy in agricultural applications (32). LISS-IV high-resolution and Cartosat-1 data were utilized for mapping horticultural plantations using a semi-automatic object-oriented feature extraction model approach in ENVI software. The feature extraction model incorporated textural, contextual, spectral, and NDVI information for classifying plantation crops with an object-oriented technique. The results achieved a reasonable classification accuracy of 75-80% (33). The eCognition software for object-based classification and representative objects for the mango crop were carefully delineated and used to train the module. Several iterations were performed to increase the classification accuracy to 91.2% (34).

Training sample extraction

Manually identified and labeled training samples were collected for each land cover class: water, crop, and forest. These training samples were used to train the classifier, enabling it to learn the distinguishing features of each class. In eCognition, a class hierarchy organizes object classes in a parent-child structure, facilitating the systematic classification of image objects. General categories are placed at higher levels, while specific categories are positioned at lower levels.

Classification and post-classification refinement

The trained models were applied to classify each object into non-cashew categories. Each RF, DT, and SVM algorithm generated an individual classification map (Fig. 4-6). The ML algorithms utilized the training data to classify image objects into predefined classes based on the extracted features for rule-set classification. A common procedure for classifying segmented objects using a rule-set algorithm involves assigning objects to classes based on prior knowledge and nearest-neighbor training samples (35).

Decision tree (DT) strength

The DT methodology is widely employed to make predictions. Researchers commonly opt for this technique due to its simplicity and clarity in identifying patterns within extensive and limited datasets and its ability to forecast values (36). The DT algorithm utilizes if-then-else rules to create classification pathways.

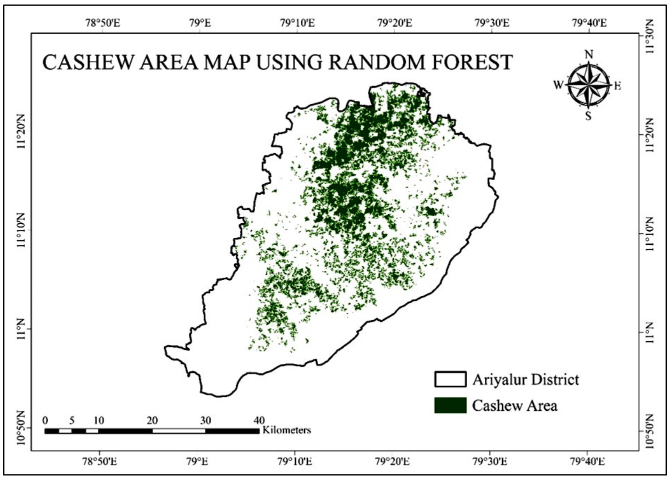

Fig. 4. Cashew area map using random forest algorithm.

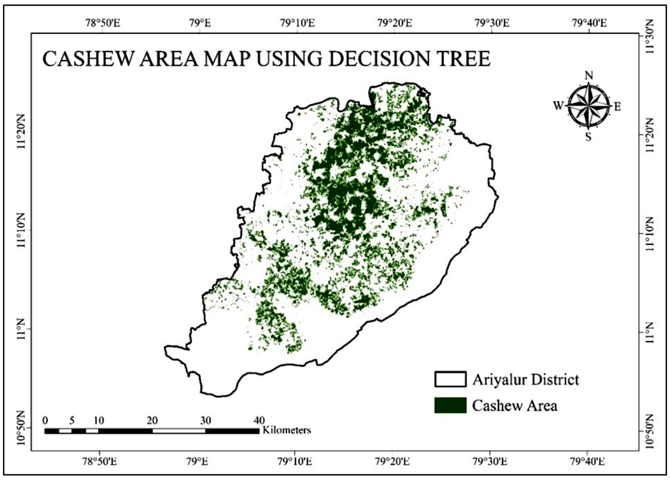

Fig. 5. Cashew area map using decision tree algorithm.

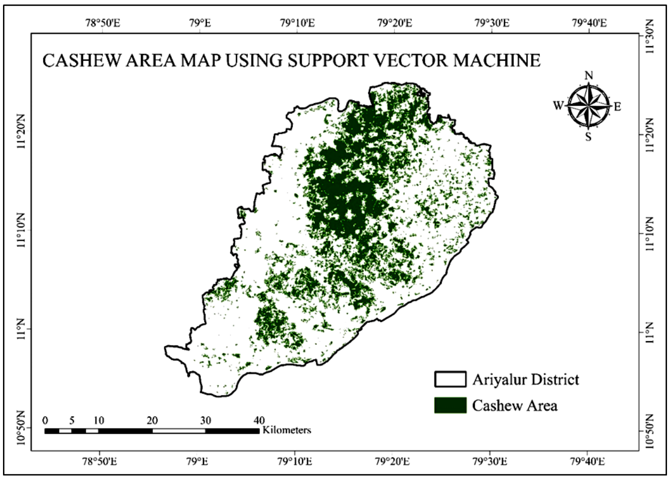

Fig. 6. Cashew area map using support vector machine algorithm.

Decision tree (DT) weakness

Decision tree classification algorithms face challenges such as instability in rule generation, sensitivity to small data changes, and difficulty in providing consistent, interpretable rules for decision-making (37).

Random forest (RF) strength

The RF algorithm, widely recognized in supervised learning, constructs a forest of multiple trees, enabling regression and classification techniques. Its foundation lies in the DT algorithm, where multiple trees are formed and combined to produce the desired output (38).

Random forest (RF) weakness

Random forest can be challenging due to its complexity, parameter selection, and the need for effective variable importance measures, especially in high-dimensional settings with sparse data (39).

Support vector machine (SVM) strength

The SVM is widely applied across multiple domains, including face detection, text categorization, and image classification, due to its robust ability to handle complex datasets. The method is grounded in the theoretical framework of structural risk minimization, which aims to balance model complexity and training error for improved generalization performance (40).

Support vector machine (SVM) weaknesses

SVMs require careful selection of kernel functions to transform data into higher dimensions, which can complicate implementation (41). Ground truth data was used to train the RF, DT, and SVM models. The features extracted from the segmented objects served as input variables, while the ground truth labels (cashew or non-cashew) acted as the output classes. The classified map was refined by merging objects to ensure coherent and meaningful areas (42). Spatial filters were applied to remove noise, and post-classification smoothing techniques were employed to generate the crop area map. A plantation crop comparison between object-based and pixel-based classification methods revealed the superiority of the object-based approach.

Accuracy assessment and validation

A confusion matrix evaluated classification accuracy by comparing the object-based classified image with ground truth data. The kappa coefficient was also calculated to assess the agreement between the two classifications. The accuracy of various object-based supervised classification methods was assessed in this study using an error matrix with validation samples. The results of object-based classification for cashew plantations using these different supervised classification techniques (Fig. 4-6). The classified images were statistically validated using ground truth data, generated through an error matrix and kappa statistics to assess classification accuracy. A total of 120 ground truth locations were identified and documented for classification, subsequently verified through visual interpretation using satellite imagery available on Google Earth Pro. As described by Lillesand, an error matrix was generated based on the agreement and disagreement of classified objects (43). This matrix was then used to calculate the kappa coefficient and overall accuracy (44).

Results

Overall accuracy trends with different machine learning algorithms

The performance of Ml classifiers- DT, SVM, and RF was evaluated to estimate the block-level cashew area in the Ariyalur district (Tables 1 and 1a-1c). Among these algorithms, the SVM algorithm achieves 92.1% overall accuracy (OA) with a kappa coefficient of 0.85, followed by the RF

Table 1. Individual classification results of efficient performing algorithm

District	Model	Overall accuracy (%)	Kappa coefficient
Ariyalur cashew area	Random forest	89.6	0.79
	Decision tree	81.1	0.62
	Support vector machine	92.1	0.85

Table 1a. Random forest accuracy assessment

		Predicted class from the map		
Actual class from		Cashew	Non-cashew	Accuracy
	Cashew	73	7	91.3%
	Non-cashew	6	36	85.7%
	Reliability	92.4	83.7%	89.6%
Overall accuracy				89.6%
Kappa index				0.79

Table 1b. Decision tree accuracy assessment

		Predicted class from the map		
Actual class from		Cashew	Non-cashew	Accuracy
	Cashew	68	12	85.0%
	Non-cashew	11	31	73.8%
	Reliability	86.1%	72.1%	81.1%
Overall accuracy				81.1%
Kappa index				0.62

Table 1c. Support vector machine accuracy assessment

		Predicted class from the map		
Actual class from		Cashew	Non-cashew	Accuracy
	Cashew	76	4	95.0%
	Non-cashew	6	36	85.7%
	Reliability	92.7%	90.0%	92.1%
Overall accuracy				92.1%
Kappa index				0.85

algorithm with an overall accuracy of 89.6% and a kappa value of 0.79.

The DT classifier exhibited the lowest performance estimate, with an OA of 81.1% and a kappa coefficient of 0.62. However, it exhibited the lowest performance estimate in the cashew area in the Ariyalur district.

The results were compared against the statistical area reported by the Department of Economics and Statistics (2023), which records 31,678 hectares of cashew plantations (Table. 2). The comparison revealed that the SVM classifier achieved accuracy, with a deviation of 3.10% from the statistical cashew area. The RF method resulted in a moderate deviation of 5.95%, whereas the DT method exhibited the most significant deviation at 11.22%, indicating overestimation. In a similar study, it was reported that the mango areas in major blocks of Krishnagiri district were 9,077.9 hectares, while the statistical area reported by the Department of Economics and Statistics was 9,746.2 hectares (34). The estimated area deviated by 6.85% from the statistical data.

Table 2. Comparison of estimated and statistical cashew areas with deviations for different classification methods

District	Estimated area (in ha)	Statistical area (in ha)	Deviation (in %)
Decision tree	31678	28481	11.22
Support vector machine	31678	30727	3.10
Random forest	31678	29900	5.95

Block-level classification area

In Ariyalur district, a cashew plantation area map was generated using ML algorithms for block-wise area estimation (Table 3). In the Andimadam block, according to the estimated cashew area, it was 11,208 ha RF, followed by 11,505 ha SVM and 10,507 ha DT. In the Sendurai block, the SVM estimated the highest cashew area to be 10,476 ha, the RF estimated 10,208 ha, and the DT estimated 9,615 ha. For the Jayamkondam block, the RF model predicted 2,607 ha, slightly greater than the 2,520 ha, predicted by the SVM model, whereas the DT model estimated a larger area of 3,321 ha.

Table 3. Block-wise distribution of cashew area in the Ariyalur district (in ha)

Blocks	RF	SVM	DT
Andimadam	11208	11505	10507
Sendurai	10208	10476	9615
Jayamkondam	2607	2520	3321
Ariyalur	2929	2778	2023
T.Palur	2403	2531	2202
Thirumanur	545	917	813
Total	29900	30727	28481

In the Ariyalur block, the RF was classified as 2,929 ha, the SVM predicted 2,778 ha, and DT estimated the lowest area at 2,023 ha. For the T. palur block, the SVM provided the highest estimate of 2,531 ha, followed by the RF model at 2,403 ha and the DT model at 2,202 ha. In the Thirumanur block, the SVM again predicted the highest area (917 ha), the RF model estimated at 545 ha, and the DT model predicted 813 ha.

Overall, the total cashew area predicted across all blocks was 30,727 ha by SVM, 29,900 ha by RF, and 28,481 ha by the DT. These results indicate that the SVM consistently predicted a higher cashew area across most blocks, while the DT provided lower estimates.

These results highlight that the SVM consistently produced higher cashew area estimates across most blocks, demonstrating its effectiveness for cashew classification in the Ariyalur district compared with the other algorithms. This superior performance suggests that an SVM may be more suitable for cashew area delineation in this region. Moreover, although valid simultaneously, the DT and RF classifiers achieved relatively lower accuracies.

Discussion

Performance of different ML algorithms

This study indicated that the SVM was the best-performing algorithm for cashew plantation classification, achieving an OA of 92.1% with a kappa coefficient of 0.85. The RF closely followed, with an OA of 89.6% and a kappa value of 0.79. The DT classifier exhibited the lowest performance, with an OA of 81.1% and a kappa coefficient of 0.62. Decision tree creates simple decision boundaries, which may not capture the complexity of object features, leading to less accurate classification. They can also overfit the training data, resulting in poor results.

These results support the effectiveness of combining ML algorithms and OBIA for precise cashew area maps. In a comparable study, SVM is an effective instrument for defining cashew regions, with an OA of 88.6% and a kappa coefficient of 0.86 in crop monitoring utilizing Sentinel 2 data (45).

In some cases, SVM also demonstrates strong performance, achieving accuracies of 95.48% in the eucalyptus classification (46) and 89.88% for betel palms (47). Algorithm performance based on input data and source parameters. In addition, RF required more feature values than SVM to achieve maximum accuracy, attributed to SVM's strength in performing higher accuracy using limited input data. We also found that the SVM classifier was more sensitive to influence the data redundancy compared with the RF classifier.

The segmentation scale parameter is one of the critical factors for quality segmentation. The scale parameter properties the spatial scale of segmentation, as it is positively related to the size of objects. Multi-segmentation-based classification sometimes suffers from error propagation at the object level across varying object scales, potentially impacting the accuracy and consistency of the classification results. Object-based classification was correctly applied using rule-set parameters for segmentation, allowing the objects to be accurately classified (48).

In a similar study, it was reported that SVM produced the most accurate land cover maps, with a kappa coefficient of 0.916, whereas the value was 0.909 for RF (49). Other studies have suggested that classification and regression trees can outperform SVM in crop classification. SVM outperformed in the Lake Urmia Basin (Iran), achieving 91.4% accuracy (50). Conversely, in the Munneru river basin (India), the RF was the best-performing algorithm, with an accuracy of 94.5% (51). Similarly, RF outperformed the SVM in LULC classification in Botswana's greater Gaborone planning area (52).

Additionally, the accuracy is heavily dependent on the quality of training samples. Thus, maintaining high-quality input samples is crucial, and adopting advanced, efficient algorithms is recommended for achieving accurate crop classifications. Cashew plantations are mostly suitable for object-based classification in future work.

Conclusion

An integration of OBIA segmentation and ML classifiers was explored for cashew classification from sentinel 2 data. The optimal scale parameter for segmentation was determined through quantitative measures, ensuring effective delineation of objects. Feature selection significantly improved segmentation and classification accuracy while reducing the computational cost of image analysis. Using an object-based approach, the SVM classifier achieved an overall accuracy of 92.1% and a kappa coefficient of 0.85%. This proposed classification procedure demonstrates its potential for large-area cropland mapping and agricultural monitoring. This study demonstrates ML algorithms, such as SVM, RF, and DT, for classifying cashew plantation areas in the Ariyalur district. Among the classifiers, the SVM performed best with an overall accuracy of 92.1% and a kappa coefficient of 0.85, whereas the RF algorithm followed, DT had a lower accuracy of 81.1%.

The algorithm's accuracy depends on the crop type, data characteristics, and environmental conditions. The SVM classifier is suitable for dealing with complex data boundaries. It is, therefore, most effective for cashew classification, while the ability of the RF model to capture spatial variations proved valuable for block-level analysis. A hybrid approach that combines the strengths of different models could be considered to improve accuracy.

This study highlights the importance of selecting appropriate ML algorithms for specific crops and regions. Future research could explore the integration of multiple algorithms to improve cashew area estimation and other land use classifications.

This study evaluated their effectiveness in delineating the topography of the complicated terrain of Ariyalur district. The classifiers also substantiated their utility in segregating land use and land cover at the regional and sub-regional levels with optimal classification accuracy. The thematic maps generated for cashew cultivation provide valuable insights that can inform several policies related to cashew exports and incentives for processing industries. Several policy decisions can be advocated with higher precision and judicial use of resources.

Additional restrictions included the recognition of juvenile plantations, the differentiation of plantations within agro-horticultural systems, and the management of intercropping scenarios. To achieve accurate object segmentation, it was crucial to provide the correct scale parameter; otherwise, it was challenging. Additionally, it was essential to assign the training classes correctly, as incorrect assignments could lead to misclassification.

For algorithm improvement, future studies could integrate and assess additional features to better discriminate crop types with similar phenollogical characteristics. Incorporating more spectral indices, texture measures, and temporal data could enhance the models ability to differentiate between crops with similar growth patterns, leading to more precise classification outcomes.

Acknowledgements

The authors thank the Professor and 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. PS handled the software work. MD generated the map. RK was responsible for ground truth data collection. S and NM 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 interests to declare.

Ethical issues: None

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT by Open AI to enhance language clarity and readability. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

1. Gokool S, Mahomed M, Kunz R, Clulow A, Sibanda M, Naiken V, et al. Crop monitoring in smallholder farms using unmanned aerial vehicles to facilitate precision agriculture practices: A scoping review and bibliometric analysis. Sustainability.2023;15(4):3557. https://doi.org/10.3390/su15043557
2. Acharki S, Frison PL, Veettil BK, Pham QB, Singh SK, Amharref M, Bernoussi AS. Land cover and crop types mapping using different spatial resolution imagery in a Mediterranean irrigated area. Environ Monit Assess. 2023;195:1309. https://doi.org/10.1007/s10661-023-11877-4
3. Shrivastava RJ, Gebelein JL. Land cover classification and economic assessment of citrus groves using remote sensing. ISPRS J Photogramm Remote. 2007;61(5):34153. https://doi.org/10.1016/j.isprsjprs.2006.10.003
4. Verma HC, Ahmed T, Rajan S. Mapping and area estimation of mango orchards of Lucknow region by applying knowledge based decision tree to Landsat 8 OLI Satellite Images. Int J Innov Technol Explor.2020;9(3):362735. https://doi.org/10.35940/ijitee.B8109.019320
5. Stephen S, Haldar D, Patel NR. Impact of various vegetation indices on mango orchard mapping using object-based image analysis. J Geomat. 2022;16(2):159-66. https://doi.org/10.58825/jog.2022.16.2.45
6. Mueller M, Segl K, Kaufmann H. Edge-and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. Pattern Recognit. 2004;37(8):1619-28. https://doi.org/10.1016/j.patcog.2004.03.001
7. Bock M, Xofis P, Mitchley J, Rossner G, Wissen M. Object-oriented methods for habitat mapping at multiple scales–Case studies from Northern Germany and Wye Downs, UK. J Nat Conserv. 2005;13(2-3):75-89. https://doi.org/10.1016/j.jnc.2004.12.002
8. Gamanya R, De Maeyer P, De Dapper M. An automated satellite image classification design using object-oriented segmentation algorithms: A move towards standardization. Expert Syst Appl. 2007;32(2):616-24. https://doi.org/10.1016/j.eswa.2006.01.055
9. Tran KH, Zhang HK, McMaine JT, Zhang X, Luo D. 10 m crop type mapping using Sentinel-2 reflectance and 30 m cropland data layer product. Int J Appl Earth Obs. 2022;107:102692. https://doi.org/10.1016/j.jag.2022.102692
10. 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
11. Forkuor G, Conrad C, Thiel M, Ullmann T, Zoungrana E. Integration of optical and synthetic aperture radar imagery for improving crop mapping in Northwestern Benin, West Africa. Remote Sens. 2014;6(7):6472-99. https://doi.org/10.3390/rs6076472
12. Kramarczyk P, Hejmanowska B. Unet neural network in agricultural land cover classification using SENTINEL-2. Int Arch Photogramm Remote Sens Spatial Inf Sci. 2023;XLVIII-1/W3-2023:85-90. https://doi.org/10.5194/isprs-archives-XLVIII-1-W3-2023-85-2023
13. Ouattara A, Zenobe ED, Sangne CY, Frazier BE, Azagoh MC, Konan NA. Determining cashew acreages in a fragmented landscape: Object vs. pixel based classification. Online J Earth Sci. 2010;4(1):13-18.
14. Xing J, Li Z, Li J, Du S, Li W, Zhang C. A classification framework with multi-level fusion of object-based analysis and convolutional neural network: A case study for land use classification in mining areas. Geo-Spat Inf Sci. 2024;1-7. https://doi.org/10.1080/10095020.2024.2336594
15. Breiman L. Random forests. Machine Learning. 2001;45:5-32. https://doi.org/10.1023/A:1010933404324
16. Koppaka R, Moh TS. Machine learning in Indian crop classification of temporal multi-spectral satellite image. In: 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM); 2020 Jan 3; Taichung, Taiwan. IEEE; 2020. p.1-8. https://doi.org/10.1109/IMCOM48794.2020.9001718
17. Belgiu M, Drăguţ L. Random forest in remote sensing: A review of applications and future directions. ISPRS J Photogramm Remote Sens. 2016;114:24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011
18. Phiri D, Simwanda M, Nyirenda V, Murayama Y, Ranagalage M. Decision tree algorithms for developing rulesets for object-based land cover classification. ISPRS Int J Geo. 2020;9(5):329. https://doi.org/10.3390/ijgi9050329
19. Pal M, Mather PM. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens Environ. 2003;86(4):554-65. https://doi.org/10.1016/S0034-4257(03)00132-9
20. Teluguntla P, Thenkabail PS, Oliphant A, Xiong J, Gumma MK, Congalton RG, et al. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. ISPRS J Photogramm. Remote Sens. 2018-40. https://doi.org/10.1016/j.isprsjprs.2018.07.017
21. Talukdar S, Singha P, Mahato S, Shahfahad, Pal S, Liou YA, Rahman A. Land-use land-cover classification by machine learning classifiers for satellite observations-A review. Remote Sens. 2020;12(7):1135. https://doi.org/10.3390/rs12071135

22. Li RH, Belford GG. Instability of decision tree classification algorithms. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2002; p. 570-5. https://doi.org/10.1145/775047.775131

23. Europe's Eyes on Earth [Internet]. The Programme of the European Union; 2023 [cited 2024 July 1]. https://www.copernicus.eu/sites/default/files/202307/Copernicus%20General%20EN.pdf

24. Sabthapathy M, Kaliaperumal R, Pazhanivelan S, Velmurugan S. Cashew area mapping using Sentinel-2 in Ariyalur district of Tamil Nadu, India. Ecol Environ Conserv. 2022; 11:68-68.

25. 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

26. Adrian J, Sagan V, Maimaitijiang M. Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine. ISPRS J Photogramm Remote Sens. 2021;175:215-35. https://doi.org/10.1016/j.isprsjprs.2021.02.018

27. Baatz M, Benz U, Dehghani S, Heynen M, Holtje A, Hofmann P, Lingenfelder I, Mimler M, Sohlbach M, Weber M, Willhauck G. eCognition professional user guide, version 4.0. München (Germany): Definiens Imaging GmbH; 2004.

28. Hussain M, Chen D, Cheng A, Wei H, Stanley D. Change detection from remotely sensed images: From pixel-based to object-based approaches. ISPRS J Photogramm Remote Sens. 2013;80:91-106. https://doi.org/10.1016/j.isprsjprs.2013.03.006

29. Li M, Ma L, Blaschke T, Cheng L, Tiede D. A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. Int J Appl Earth Obs Geoinf. 2016;49:87-98. https://doi.org/10.1016/j.jag.2016.01.011

30. Rasuly A, Naghdifar R, Rasoli M. Monitoring of Caspian Sea coastline changes using object-oriented techniques. Procedia Environ Sci. 2010;2:416-26. https://doi.org/10.1016/j.proenv.2010.10.046

31. 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;(6):859-71. https://doi.org/10.1080/13658810903174803

32. Rana M, Kharel S. Feature extraction for urban and agricultural domains using ecognition developer. Int Arch Photogramm Remote Sens Spatial Inf. 2019; XLII-3/W64:609-15. https://doi.org/10.5194/isprs-archives-XLII-3-W6-609-2019

33. Hebbar R, Ravishankar HM, Trivedi S, Subramoniam SR, Uday R, Dadhwal VK. Object oriented classification of high resolution data for inventory of horticultural crops. Int Arch Photogramm Remote Sens Spatial Inf Sci. 2014;XL-8:745-9. https://doi.org/10.5194/isprsarchives-XL-8-745-2014

34. Harish M, Kaliaperumal R, Pazhanivelan S, Sivakumar C, Kumaraperumal R. Mango area mapping using very high-resolution satellite data in major blocks of Krishnagiri district, Tamil Nadu, India. Int J Environ Clim Chang. 2023;13(9):8118. https://doi.org/10.9734/ijecc/2023/v13i92302

35. Chabalala Y, Adam E, Ali KA. Machine learning classification of fused Sentinel-1 and Sentinel-2 image data towards mapping fruit plantations in highly heterogenous landscapes. Remote. 2022;14(11):2621. https://doi.org/10.3390/rs14112621

36. Chaudhary D, Vasuja ER. A review on various algorithms used in machine learning. Int J Sci Res Comput Sci Eng Inf. 2019;5(2):915-20. https://doi.org/10.32628/CSEIT1952248

37. Aasi E, Vasile CI, Bahreinian M, Belta C. Classification of time-series data using boosted decision trees. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2022 Oct 23; Kyoto, Japan. IEEE; 2022. p. 1263-8. https://doi.org/10.1109/IROS47612.2022.9982105

38. Biau G, Scornet E. A random forest guided tour. TEST. 2016;25:197-227. https://doi.org/10.1007/s11749-016-0481-7

39. Natek S, Zwilling M. Student data mining solution-knowledge management system related to higher education institutions. Expert Syst. 2014;41(14):6400-07. https://doi.org/10.1016/j.eswa.2014.04.024

40. Li RH, Belford GG. Instability of decision tree classification algorithms. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '02). Edmonton, Alberta, Canada: Association for Computing Machinery; 2002. p. 570–75.

41. Du KL, Swamy MNS. Support vector machines. In: Neural networks and statistical learning. London: Springer; 2014. p.319–45. https://doi.org/10.1007/978-1-4471-5571-3_16

42. Hedayati A, Vahidnia MH, Behzadi S. Paddy lands detection using Landsat-8 satellite images and object-based classification in Rasht city, Iran. Egypt J Remote Sens Space Sci. 2022;25(1):73-84. https://doi.org/10.1016/j.ejrs.2021.12.008

43. Lillesand T, Kiefer RW, Chipman J. Remote sensing and image interpretation. Hoboken: John Wiley & Sons; 2015.

44. 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

45. Rawat S, Saini R. Cropland mapping using single date Sentinel-2 imagery using machine learning classifiers. In: Proceedings of the 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM); 2022. p. 1–7. https://doi.org/10.1109/ICACCM56405.2022.10009406

46. Lin X, Ren C, Li Y, Yue W, Liang J, Yin A. Eucalyptus plantation area extraction based on SLPSO-RFE feature selection and multi-temporal sentinel-1/2 data. Forests. 2023;14(9):1864. https://doi.org/10.3390/f14091864

47. Luo H, Li M, Dai S, Li H, Li Y, Hu Y, et al. Combinations of feature selection and machine learning algorithms for object-oriented betel palms and mango plantations classification based on Gaofen-2 imagery. Remote Sens. 2022;14(7):1757. https://doi.org/10.3390/rs14071757

48. Hunt Jr ER, Li L, Yilmaz MT, Jackson TJ. Comparison of vegetation water contents derived from shortwave-infrared and passive-microwave sensors over central Iowa. Remote Sens Environ. 2011;115(9):2376-83. https://doi.org/10.1016/j.rse.2011.04.037

49. Xue H, Xu X, Zhu Q, Yang G, Long H, Li H, et al. Object-oriented crop classification using time series sentinel images from google earth engine. Remote Sens. 2023;15(5):1353. https://doi.org/10.3390/rs15051353

50. Feizizadeh B, Omarzadeh D, Kazemi Garajeh M, Lakes T, Blaschke T. Machine learning data-driven approaches for land use/cover mapping and trend analysis using Google Earth Engine. J Environ Plann Man. 2023;66(3):665-97. https://doi.org/10.1080/09640568.2021.2001317

51. Loukika KN, Keesara VR, Sridhar V. Analysis of land use and land cover using machine learning algorithms on google earth engine for Munneru River Basin, India. Sustainability. 2021;13(24):13758. https://doi.org/10.3390/su132413758

52. Ouma YO, Keitsile A, Nkwae B, Odirile P, Moalafhi D, Qi J. Urban land-use classification using machine learning classifiers: Comparative evaluation and post-classification multi-feature fusion approach. Eur J Remote Sens.2023;56(1):2173659. https://doi.org/10.1080/22797254.2023.2173659