Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Crop Discrimination and Classification Using Sentinel-2A and Machine Learning Techniques

Pradnya L. Awate[1], Ajay D. Nagne[2]
[1]*Dr. G. Y. Pathrikar College of Computer Science & IT, MGM University, Chhatrapati Sambhajinagar-431003, MH, India.*

ARTICLE INFO

Received: 30 Dec 2024
Revised: 19 Feb 2025
Accepted: 27 Feb 2025

ABSTRACT

Precision agriculture and agricultural monitoring rely heavily on accurate crop identification. This study examines the ability of Support Vector Machine and Random Forest classifiers to distinguish cotton, wheat, maize, and sugarcane using Sentinel-2 imagery. Both classifiers were trained and tested using ground truth data, with accuracy determined by the User's Accuracy, Producer's Accuracy, Overall Accuracy, and Kappa coefficient. The outcome revealed RF as having a better OA (83.33%) and Kappa value (0.77) than SVM (66.66% OA, 0.55 Kappa), reflecting a better classification performance. RF also had a better validation accuracy (97.91%) than SVM (95.83%). Although both models performed well in classifying crops, moderate UA values reveal room for improvement. The research emphasizes the efficacy of RF for crop categorization and recommends future enhancement with the use of sophisticated machine learning methods and multi-temporal data fusion for more efficient agricultural surveillance.

Keywords: Crop Classification, Random Forest, Support Vector Machine, NDVI

INTRODUCTION

Food insecurity must be addressed through sustainable agriculture in response to increased food demand caused by population growth and urbanization. With limited arable land, advanced technology is required [1]. Remote sensing increases precision farming by optimizing irrigation, crop health monitoring, and reducing climatic risk, resulting in more effective resource usage and yield [2]. Precision farming in Maharashtra utilizes technology and data-driven operations for increased productivity and sustainability [3]. Machine learning, such as RF, SVM, XGBoost, and deep models like CNNs and RNNs, is applied to further crop identification based on remote sensing, maximization of the use of resources and enhancing agricultural production [4].

This research investigates the use of machine learning-based temporal analysis in crop classification using Sentinel-2 images over the geographical vicinity of Chhatrapati Sambhaji Nagar, Maharashtra. More precisely, we want to identify cotton, wheat, maize, and sugarcane crops, and we've identified RF and SVM as two novel algorithms for machine learning that may aid us do so. The fundamental goal of our research is to evaluate the geographic generalizability of these algorithms as well as to identify an effective feature selection scheme—these efforts will result in improved crop discrimination and classification accuracy at various spatial resolutions. The research paper is well-organized, beginning with a review of related literature in Section 2, followed by a brief description of the specific study area for investigation in Section 3. Section 4 focuses on ground data collection. part 5 is the preferred approach, as defined, while part 6 contains the results and discussion. The study's final analysis may be found in Section 7, which covers the entire research endeavor.

RELATED WORK

Numerous investigations have demonstrated the use of machine learning algorithms for crop classification; some of the authors' work is discussed here. This study explores crop classification using Sentinel-2 imagery and various machine learning techniques to enhance crop monitoring and area estimation. It addresses challenges like spectral similarities and environmental factors by employing methods such as Random Forest (RF), SVM, KNN, ISODATA, and K-means clustering, with NDVI aiding spectral feature identification. RF achieved the highest accuracy (87.71%), followed by ISODATA (85.01%), demonstrating its effectiveness for classification [5]. Additionally, pre-harvest crop mapping using Sentinel-2 and machine learning achieved 77.2% accuracy eight weeks before harvest, highlighting the potential for early-season predictions [6]. Crop acreage analysis in Tamluk, West Bengal, using KNN and RF,

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

yielded high accuracies (97.16% and 97.22%), useful for agricultural management [7]. Furthermore, a 3D CNN model incorporating spatio-temporal Sentinel-2 data outperformed traditional methods, effectively capturing crop growth dynamics [8]. These findings support advanced machine learning applications in precision agriculture, improving crop classification accuracy and facilitating better agricultural decision-making. The following sections will look at the technique, datasets, and results of crop classification machine learning methods.

STUDY AREA

The Sillod (20.3079° N, 75.6528° E) was chosen as the research region due to was an agricultural educational area located in the district of Chhatrapati Sambhaji Nagar, formerly known as Aurangabad, in Maharashtra, India. The map of the research region was built using the shapefile data from Figure 1, which is available on the Survey of India website. Cotton, wheat, maize and sugarcane are the principal crops grown in this region, which is made feasible by the lush soil of the Deccan plateau. In 2023, temporal images for August, September, and October were obtained in Google Earth Engine after filtering the Sentinel-2A image collection to exclude cloud data. The images were crucial in classifying cotton, wheat, maize and sugarcane for the reason that, cotton and sugarcane were planted in June, with maize and wheat sown in October. The platform's JavaScript API makes it easier to collect and evaluate data, resulting in more accurate crop growth stage tracking.

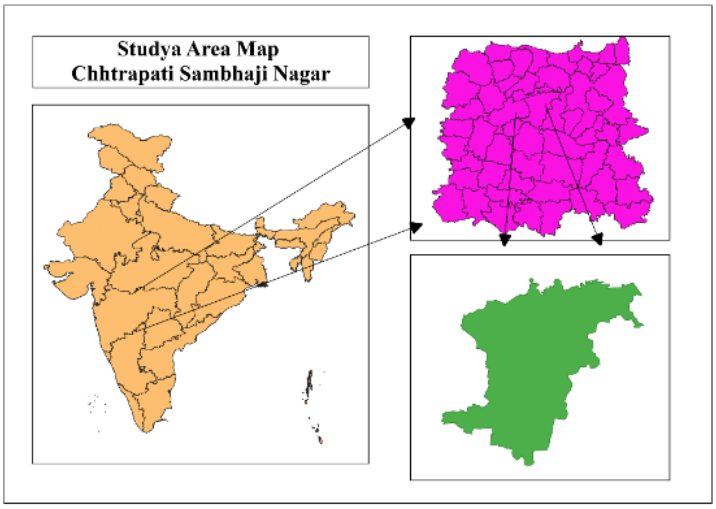


Fig 1: Study Area

GROUND DATA COLLECTION

Upon identifying the location, we carried out ground data collection in Nillod village, Sillod, Maharashtra. Representative samples of crops such as cotton, soybean, and sugarcane were meticulously collected to ensure that different leaf sizes were covered within the canopy. GPS coordinates were taken systematically for accurate mapping and subsequent analysis using satellite imagery.

METHODOLOGY

Figure 2 depicts the process of development stage-wise. We processed the data after acquiring Sentinel-2A imagery by cloud masking, calculating NDVI, creating composite images, and clipping. Temporal images were stacked, and then classification was done using RF and SVM with training data and LULC maps. Accuracy assessment was done to ensure credible results, resulting in a critical analysis in the results and discussion section. The study ends with the conclusion of key findings and their implications.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

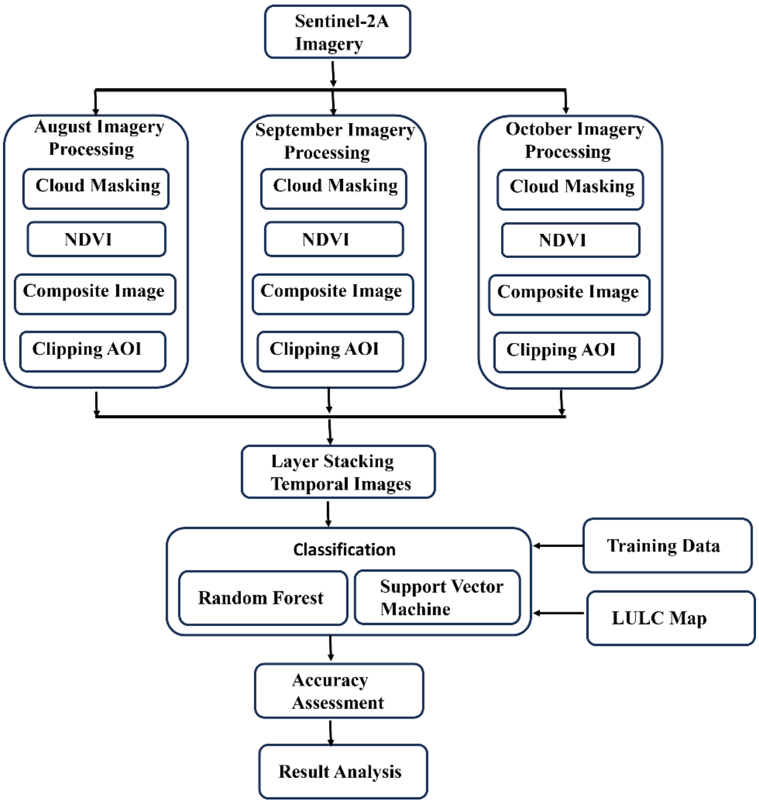


Fig 2: Preferred Methodology

DATA SET

Between August and October, Sentinel-2A imagery data from the Copernicus Open Access Hub was analysed in Google Earth Engine. The 13-band data consisting of 10m, 20m, and 60m resolutions carries rich spectral details that are pivotal in accurate crop classification and analysis.

PROCESSING

After acquiring images, we need to perform further necessary processing on it, which are as follows,

CLOUD MASKING

Cloud masking detects and eliminates clouds and their shadows from satellite images so that precise remote sensing data can be obtained. Cloud masking enhances data quality for uses such as crop classification by reducing cloud-borne spectral distortions.

NDVI CALCULATION

NDVI is a primary crop index mathematical formula given in equation 1, that measures photosynthetic activity and the health of the crop as a function of red and near-infrared reflectance. NDVI values vary between -1 (no vegetation) and +1 (heavy vegetation) [9].

$$NDVI = \frac{(NIR-RED)}{(NIR+RED)} \quad (1)$$

COMPOSITE IMAGE

A composite NDVI image, produced by stacking several NDVI images across time, minimizes cloud and shadow effects, maximizes land cover classification accuracy, and enhances vegetation and crop health analysis. It is most helpful for the identification of temporal change and long-term vegetation trend monitoring.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

CLIPPING AOI

Clipping the region of interest cuts out particular areas from a composite picture to enhance research precision. It minimizes data size, conserves computing resources, and makes analysis easier, making it valuable for large-scale remote sensing research. It provides a targeted approach by examining only the most critical geographic regions [10].

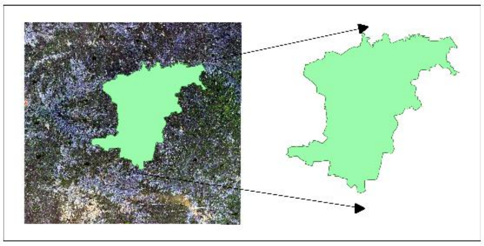


Fig 3: (a) Before Clipping (b) After Clipping AOI

LAYER STACKING TEMPORAL IMAGES

Layer stacking combines images acquired at different times to produce a multi-layered dataset, facilitating change detection and trend analysis. Such a method enables the tracking of crop growth stages, detection of anomalies, and enhanced classification accuracy.

CLASSIFICATION

Classification is used to determine the types of land cover, like crops, in satellite images through the transformation of spectral data into meaningful agricultural data. Machine learning classifiers such as Random Forest and Support Vector Machine were employed in this research for enhanced accuracy and efficiency in identifying crops through pattern learning from training data.

RANDOM FOREST

Random Forest (RF) is an ensemble learner algorithm that works by building a series of decision trees in training and producing the class that is the majority vote of the single trees. RF can be applied widely to crop discrimination because it can manipulate high-dimensional remote sensing data as well as intricate relationships among spectral features. RF minimizes overfitting through averaging several trees, enhancing classification accuracy and noise robustness. The algorithm operates by choosing random subsets of training data and attributes to extend individual trees [11]. Mathematically, the RF classification is based on equation 2:

$$\hat{y} = \arg\max_{c} \sum_{t=1}^{T} I(h_t(x) = C) \qquad (2)$$

where $\hat{y}$ is the predicted class, $h_t(x)$ is the decision of the t-th tree, T is the total number of trees, c represents the possible classes, and I(·) is an indicator function that returns 1 if the condition is met and 0 otherwise. This method is highly effective for crop discrimination as it can capture subtle spectral variations between different crops [12].

SUPPORT VECTOR MACHINE

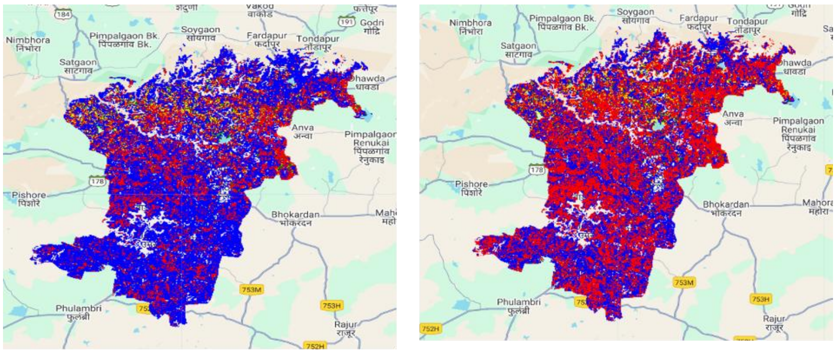
Support Vector Machine (SVM) is a strong supervised learning algorithm that can be utilized for crop discrimination, especially with high-dimensional datasets. SVM discovers the best hyperplane that can separate various classes of crops with maximum distance in the feature space. SVM takes advantage of kernel functions to project non-linearly separable data into a high-dimensional space in which a linear separator can be used [13]. Decision boundary in SVM is given as in equation 3:

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i k(x_i, x) + b) \qquad (3)$$

where α_i are the Lagrange multipliers, y_i represents class labels, x_i are support vectors, $k(x_i, x)$is the kernel function (e.g., linear, polynomial, or radial basis function), and b is the bias term. By maximizing the margin between different classes, SVM ensures high classification accuracy for crop types, distinguishing between species based on

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

subtle spectral differences. Its ability to handle non-linearly separable data makes it particularly useful for remote sensing applications in precision agriculture [14].



(a) (b)

Fig 4: (a) Random Forest Classification (b) Support Vector Machine Classification

TRAINING DATASET AND LULC INPUT

We trained SVM and RF classifiers with ground truth data for cotton, soybean, and sugarcane obtained from field visits. These precise and varied data points enabled the classifiers to learn distinctive spectral features of each crop, enhancing classification accuracy based on Sentinel-2 imagery. We also utilized the 2020 and 2021 ESA World Cover land use and land cover datasets as training inputs. These data, which were taken from Sentinel-1 and Sentinel-2 data, supplied high-resolution land cover data, making our crop classification outcome more reliable and accurate.

ACCURACY ASSESSMENT

Accuracy validation is important to assess the effectiveness of crop classification with machine learning models. We applied validation indicators including User's Accuracy (UA), Producer Accuracy (PA), Overall Accuracy (OA), and Kappa coefficient, all based on confusion matrices at ground validation sites. UA calculates the chance that a predicted class is accurate, PA measures the chance of accurately predicting a sample's true class, OA is the overall proportion of accurate predictions, and the Kappa coefficient adjusts for chance agreement. We also calculated training and validation matrices to evaluate the model's performance on training and unseen validation data. These matrices offer insights about accuracy, precision, and recall to ensure model generalization. These evaluation metrics collectively assess the performance of our crop classification model in separating various crops from Sentinel-2 images.

RESULT AND DISCUSSION

As we employed two machine learning algorithms, the following table 1 compares the performance of random forest and support vector machine.

Classifier	UA	PA	OA	Kappa Coefficient	Training Accuracy	Validation Accuracy
RF	66.66%	85.05%	83.33%	0.77	100%	97.91%
SVM	54.55%	85.71%	66.66%	0.55	100%	95.83%

Table 1: Result of RF and SVM classifiers

The results of the classification of the crops using the RF (Random Forest) and SVM (Support Vector machine) classifiers differ in their performance for various accuracy measures. RF achieved a higher Overall Accuracy (OA) of 83.33% versus 66.66% for SVM, demonstrating an overall superior ability to correctly classify the four crops. The Kappa coefficient for RF was also higher (0.77) than for SVM (0.55), indicating more agreement between predicted and actual classifications above random chance. The User's Accuracy (UA) for RF was also higher (66.66%) than for SVM (54.55%), suggesting RF had less misclassifications as false positives. The Producer's Accuracy (PA) was similar for the two classifiers. RF was 85.05% and SVM was 85.71%, which indicates both classifiers produced similar results when accurately classifying true crop samples.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The robustness of RF is also supported by the training and validation accuracies. RF demonstrated almost perfect training accuracy (100%) and a substantial validation accuracy (97.91%), whereas SVM attained a perfect training accuracy (100%) but had a slightly lower validation accuracy (95.83%). The disparities between the training and validation accuracies indicate SVM has a higher chance of misclassifying the data in new applications. Ultimately, these results indicate RF is better at classifying cotton, wheat, maize, and sugarcane than SVM and therefore presents a more compelling option for crop discrimination using Sentinel-2 imagery. Nevertheless, the moderate UA values suggest that further investigations to fine-tune the model with feature selection or hyperparameter tuning could enhance dataset concatenation.

CONCLUSION

This research assessed the performance of Support Vector Machine (SVM) and Random Forest (RF) classifiers for crop classification based on Sentinel-2 imagery. The findings proved RF was superior to SVM in total classification accuracy with an OA of 83.33% and a more significant Kappa coefficient of 0.77, reflecting good agreement with ground truth data. Besides, RF had improved UA-PA balance and hence proved to be more effective for discrimination between cotton, wheat, maize, and sugarcane. The accuracy during training and validation also assured the strength of RF with validation accuracy being 97.91% against SVM's 95.83%. Both classifiers efficiently identified crop classes; however, comparatively lower values of UA indicate improvement potential to decrease misclassification error. Future research may investigate more sophisticated machine learning methods, including deep learning models or hybrid strategies, for better classification performance. Moreover, the integration of multi-temporal information and optimization of feature selection can provide even better accuracy and enable more accurate crop differentiation for extensive agricultural monitoring.

REFRENCES

[1] Erickson, B., & Fausti, S. W. (2021). The role of precision agriculture in food security. Agronomy Journal, 113(6), 4455-4462.
[2] Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. Remote sensing, 12(19), 3136.
[3] Phade, G., Kishore, A. T., Omkar, S., & Suresh Kumar, M. (2023). IoT-Enabled Unmanned Aerial Vehicle: An Emerging Trend in Precision Farming. Drone Technology: Future Trends and Practical Applications, 301-324.
[4] Shaikh, T. A., Mir, W. A., Rasool, T., & Sofi, S. (2022). Machine learning for smart agriculture and precision farming: towards making the fields talk. Archives of Computational Methods in Engineering, 29(7), 4557-4597.
[5] Khan, A., Patil, C. H., Vibhute, A. D., & Mali, S. (2024). Mapping Crop Types At a 10 m Scale Using Sentinel-2 Data and Machine Learning Methods. International Journal of Computing and Digital Systems. https://doi.org/10.12785/ijcds/1501123
[6] Maponya, M. G., Van Niekerk, A., & Mashimbye, Z. E. (2020). Pre-harvest classification of crop types using a Sentinel-2 time-series and machine learning. Computers and electronics in agriculture, 169, 105164.
[7] Hudait, M., & Patel, P. P. (2022). Crop-type mapping and acreage estimation in smallholding plots using Sentinel-2 images and machine learning algorithms: Some comparisons. The Egyptian Journal of Remote Sensing and Space Science, 25(1), 147-156.
[8] Ji, S., Zhang, C., Xu, A., Shi, Y., & Duan, Y. (2018). 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote Sensing, 10(1), 75.
[9] Awate, P., & Nagne, A. (2023, May). Enhanced Vegetation Cover Assessment Using Sentinel-2A: A Unified Perspective on NDVI, SAVI, and GNDVI. In International Conference on Information Science and Applications (pp. 257-271). Singapore: Springer Nature Singapore.
[10] Kathar, S. P., Nagne, A. D., Awate, P. L., & Bhosle, S. (2023, December). Comparative Study of Supervised Classification for LULC Using Geospatial Technology. In International Conference on Soft Computing and its Engineering Applications (pp. 79-93). Cham: Springer Nature Switzerland.
[11] Geetha, V., Punitha, A., Abarna, M., Akshaya, M., Illakiya, S., & Janani, A. P. (2020, July). An effective crop prediction using random forest algorithm. In 2020 international conference on system, computation, automation and networking (ICSCAN) (pp. 1-5). IEEE.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

[12] Paul, A., Mukherjee, D. P., Das, P., Gangopadhyay, A., Chintha, A. R., & Kundu, S. (2018). Improved random forest for classification. IEEE Transactions on Image Processing, 27(8), 4012-4024.
[13] Kok, Z. H., Shariff, A. R. M., Alfatni, M. S. M., & Khairunniza-Bejo, S. (2021). Support vector machine in precision agriculture: a review. Computers and Electronics in Agriculture, 191, 106546.
[14] Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review. Journal of Data Analysis and Information Processing, 8(4), 341-357.