

A Comparative Analysis of Machine Learning Algorithms for Crop Type Mapping Using Sentinel-2 Imagery: Assessment of Spatial Distribution Patterns of Paddy and Maize in Semi-Arid Region

GopiraghunadhPVS^{1†} and Anjireddy M¹

¹A Centre for Environment, IST, JNTUH, Ashok Nagar, Kukatpally Housing Board Colony, Kukatpally, Hyderabad, Telangana 500085

†gopiraghunadh@gmail.com

ORCID IDs -0000-0002-7635-6359

Abstract: Remote sensing-based classification of paddy and maize crops is challenging due to spectral similarities and complex cropping systems, particularly in semi-arid regions. This study investigates the potential of multiple machine learning algorithms i.e., Forest (RF), Support Vector Machine (SVM), XGBoost, and Light GBM for paddy and maize crops classification in Mahabubabad district of Telangana, India, using Sentinel-2 imagery. Different remote sensing datasets including multi-temporal Sentinel-2 data acquired during the kharif and rabi of 2023-2024 year and SRTM digital elevation are used for training the machine learning models. Sentinel-2 data derived vegetation indices and phenological metrics, and SRTM DEM derived slope parameters are used along with 850 georeferenced crop sites for training and validation of the models. The performance of these methods are assessed using different accuracy measures. The results indicate, XGBoost outperformed other machine learning models with overall accuracy 92.3%, followed by RF and other methods. The spatial pattern analysis of classification accuracy depicts classification errors are mainly related to field size and crop phenology. This study highlights the usefulness of machine learning approaches in classifying staple crops such as paddy and maize. Additionally provides insights on the impact of different auxiliary parameters on classification accuracy. Overall the framework implemented in this study can be useful for enhancing the accuracy of crop type mapping in other regions with similar agro-ecological conditions.

Keywords: Crop type mapping; Machine learning; Spatial pattern.

1. INTRODUCTION

India is primarily agriculture-based country and 1/3rd of nation's GDP is dependent on agriculture (Kaur, 2013). With the global shift in market economies, reliable information on agriculture is having more importance than before. Accurate and timely mapping of different crops, especially the staple crops like paddy and maize, gives vital insights for the policy makers, agricultural managers, and farmers for trade policy formulation, yield prediction and managing water resources (Zhao et al., 2021; See et al., 2015; Gumma et al., 2011). Conventionally crop statistics such as acreage, yield etc., are calculated based on the land revenue system. Shortcomings of conventional systems include lack of real-time monitoring, delay in reporting, non-sampling errors, inadequate data for forecasting the data in response to change in weather and crop health conditions, etc. (Sagar & K, 2018). These limitations are overcome to certain extent by making use of remote sensing data sets such as aerial photographs, satellite imagery, etc. (Singh, 2017; Khanal et al., 2020; Mandal, 2016). These datasets facilitate more sophisticated approaches allowing accurate quantification of crop areas, thereby addresses related agrarian challenges (Migdall et al., 2018).

In recent years, the availability of high-resolution and open-access satellite data, combined with the rapid advancements in machine learning techniques, has opened new opportunities for large-scale, automated, and efficient crop area estimation (Li et al., 2022). Crop discrimination with remote sensing datasets is primarily based on spectral response and is affected by sensor characteristics as well as time of observation (Arafat et al., 2013). Different crops exhibit diverse spectral responses in spectral bands of various sensors onboard different satellites (Montibeller et al., 2019; Surase et al., 2018). The

interaction of electromagnetic radiation with crops is primarily influenced by chlorophyll and water content in optical wavelength region, whereas in microwave region, it is influenced by the dielectric properties of the crop and crop pattern, height, etc (Silleos et al., 1992). The general steps in crop acreage estimation includes delineation of crop lands, crop type identification, estimation of crop area and accuracy assessment of crop area and identification (Hudait& Patel, 2022; David & Sturza, 2010). Crop mapping in multispectral optical and thermal remote sensing imagery is carried out using visual image interpretation and digital image classification techniques (Büchi et al., 2018; Thomson & Sullivan, 2006). With the advancements in remote sensing approaches and computational resources, remote sensing datasets are extensively used to supplement various estimates and observations related to crops (such as crop mapping, yield prediction, crop health monitoring, etc.) made from field observations using ground-based surveys. This helps in reducing the amount of ground observations needed while providing the comprehensive details for various applications.

Though multiple types of crops are possible to map with remote sensing datasets, mapping staple crops i.e., paddy and maize are more important in India. These crops form essential means of food for multiple northern and southern Indian states. Many studies have demonstrated the effectiveness and the potential of combining satellite data and machine learning methods for crop area mapping for different crops including maize and paddy (K et al., 2022; She et al., 2020; Dhillon et al., 2023). In India, cultivation is mainly carried out over small and fragmented land holdings by large number of individual farmers leading to heterogeneity in crop types within small areas. This field size heterogeneity presents challenges for remote sensing applications, especially when using medium to coarse resolution satellite data. Usage of multi-temporal high-resolution satellite imagery such as Sentinel-2 in different crop phases can be advantageous in these situations for mapping different crops. The multispectral and temporal characteristics of sentinel-2 datasets allows to capture different spectral patterns exhibited by different crops and their variation be distinguished in time domain (Singh et al., 2019; Ghosh et al., 2018). Understandably, there has been an increasing number of studies focusing on the use of Sentinel-2 and other high-resolution imagery in agricultural research (Gascon, 2018). Several investigations have shown that optical multi-temporal satellite imagery facilitates crop type mapping across different climates and diverse cropping systems. In a recent study, Sentinel-1 and Sentinel-2 datasets used are in a convolution neural networks approach for crop classification in China (Wang et al., 2023) have exhibited excellent results in classifying maize, soybean, peanuts, and other crops. Similarly, many studies are carried out using different machine learning for crop area classification.

Despite the advantages of using high resolution satellite imagery for crop classification, their usage is still limited in India. For example, although different machine learning algorithms are used in crop classification, comparison studies are still not comprehensive. Accurate classification of crops such as paddy and maize are still challenging due to their spectral similarities of different crops at various growth stages, the effects of heterogeneous soil moisture conditions, and the presence of arbitrary cropping patterns. Moreover, the evolution of plant growth phenology across time and the influence of drought stress under semi-arid conditions make this classification process more challenging. Also, the interaction between field size, crop phenology, and classification accuracy is poorly understood indicating the need for research in this direction. Limitations in computational resources, complexities associated with handling large datasets poses challenges for effective implementation of such practices. Additionally, during the monsoon season, the weather conditions including clouds and rain constrains the usability of optical and infrared remote sensing images for crop classification and area estimation (Kordi& Yousefi, 2022; Shen et al., 2022; Kyere et al., 2020; Dadhwal et al., 2002). The hardware requirement and computation resources can be overcome to certain extent by making use of cloud computing geospatial platforms such as google earth engine, which provides direct access to multiple remote sensing datasets along with robust computational resources for all users at free of cost.

The main aim of this study is to compare and better understand the functionality of different machine learning algorithms for differentiating between paddy and maize crops using Sentinel-2 data, to assess

the spatial variability of the crops. The study is implemented over Mahabubabad district, Telangana, India, which exhibits semi-arid agriculture conditions. This research work presented in this study demonstrates the potential of machine learning for crop mapping and emphasizes details about various real-world challenges for agricultural stakeholders operating in semi-arid environments. This research aims to improve the understanding of the challenges and solutions for crop mapping in semi-arid regions by means of implementing different machine-learning methods. The results will provide insights on multiple factors that affect the crop classification towards the improvement of sustainable agricultural practices in similarly agriculture constrained geo-climatic settings.

The study attempts to achieve the following objectives:

1. To compare the performance of machine learning algorithms in classifying paddy and maize using Sentinel-2 data.
2. To analyze the spatial distribution of paddy and maize crops across the study area.
3. To examine the impact of environmental variables on crop classification accuracy.

The manuscript is structured in the following manner. Followed by introduction, section 2 provides the brief description about the study area and datasets. The detailed methodology is provided in section 3. The results and discussion are outlined in section 4. Finally, the conclusions are presented in section 6.

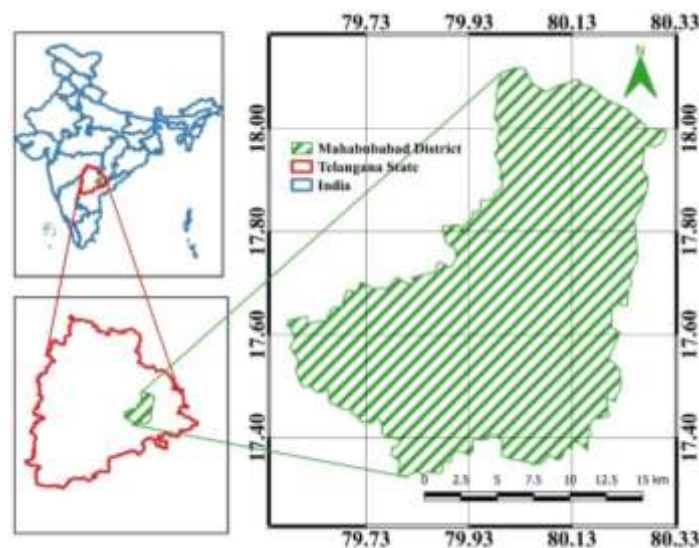


Fig. 1: Geographical extent of Mahabubabad region.

2. MATERIALS AND METHODS

2.1. Study region

Mahabubabad district is located in Telangana state, India extending between the latitudes and longitudes of 17°36' to 18°0' N and 79°30' to 80°15' E encompassing an area of approximately 2,569 square kilometres (See in Fig. 1). The study region experiences a semi-arid climate characterized by seasonal variation in weather conditions. Temperature reaches up to 45 °C in summer, whereas in winter the temperature ranges between 15–28 °C. The annual rainfall is around 900 mm and most of it is received in monsoon months (from June–September) and is important for agricultural activities in this region.

Mahabubabad's agriculture includes two major cropping seasons i.e., Kharif (June–November) and Rabi (December–March). Paddy cropping is more prevalent in Kharif with monsoon rains, whereas maize is cultivated in both crop seasons. However, in some parts of Mahabubabad paddy is cultivated even in rabi season based on the availability of water. Small to medium field sizes are most common in the district with land holding varying approximately between 1-2 hectares. Irrigation for the cultivation is

supported by a network of tanks, canals and ground water, which induces heterogeneity in agriculture practices within the district. The terrain is mainly characterized by rolling plains and few hillocks with topographical elevation ranging between 200 to 500 m above the mean sea level over the entire region. Different soil types, such as red sandy loams and black cotton soils, coupled with the irrigation facilities leads to the cultivation of variety of crops and farming techniques across the district.

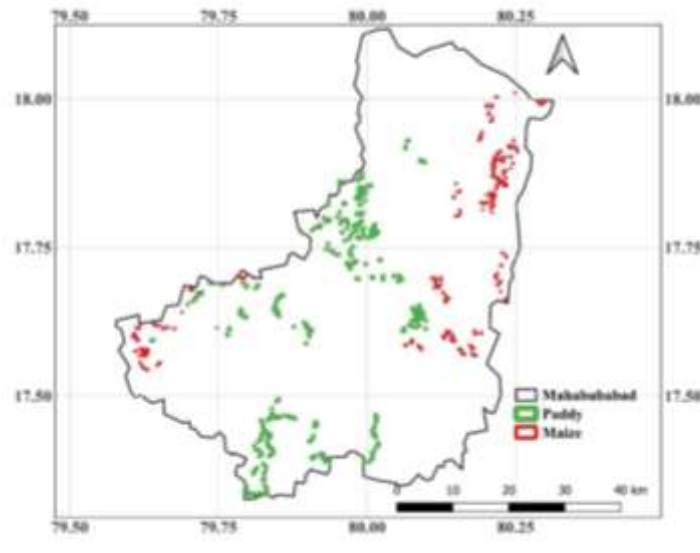


Fig. 2: Training datasets

2.2. Datasets

The Ground truth data consists of 874 field polygons which include 554 paddy fields, and 320 maize fields collected using field surveys (as shown in Fig. 2) during the study period are considered in this study. These polygons are used for training and validating the different machine learning algorithms. In addition to ground observations, multi-temporal remote sensing data i.e., Sentinel-2 satellite images were obtained for the 2023-24 agricultural season i.e., June 2023-March 2024. During the study window, remote sensing imagery are selected based on key phenological stages for both paddy and maize crops. Additionally, images with high cloud are not considered as they are less useful and can lead to misclassification of the crops. Between June 2023 to March 2024, ten Sentinel-2 scenes were selected to cover both Kharif and Rabi seasons. These details of the Sentinel-2 imagery used in the study are shown in Table 1.

In addition to Sentinel-2 imagery, Digital Elevation Model (DEM) created from the Shuttle Radar Topography Mission (SRTM) data is used as an auxiliary dataset for enhancing the accuracy of the model. From DEM, slope map of study region is created as shown in figure 3. Integrating these auxiliary datasets in the machine learning approaches facilitates the analysis of paddy and maize crop distribution and their performance with respect to different environmental and topographical factors.

Table 1: Details of Sentinel-2 imagery used in the study.

Acquisition Date	Cloud Cover (%)	Growth Stage (Paddy)	Growth Stage (Maize)	Season
June 15, 2023	5.2	Sowing/Early	Sowing	Kharif
July 20, 2023	8.4	Vegetative	Vegetative	Kharif
August 25, 2023	12.1	Reproductive	Tasseling	Kharif
September 30, 2023	3.5	Maturity	Grain filling	Kharif
December 15, 2023	2.1	Sowing	Early vegetative	Rabi
January 20, 2024	4.3	Vegetative	Reproductive	Rabi
February 15, 2024	3.8%	Reproductive/Heading	Grain filling	Rabi
March 20, 2024	2.5%	Maturity/Harvest	Maturity	Rabi

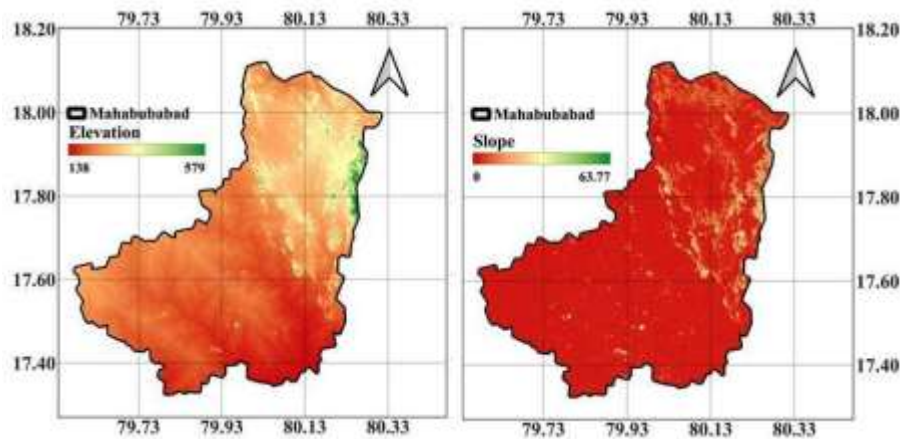


Fig. 3: Maps representing a) Topographic elevation and b) Slope of study region

2.3 Methodology

2.3.1 Preprocessing

The methodological framework starts with the data preparation task i.e., selection of suitable bands from Sentinel-2 images and their preprocessing. The details of different steps followed in the methodology are shown as a flowchart in the figure 4. The bands were selected based on the spatial resolution and their usefulness to vegetation mapping

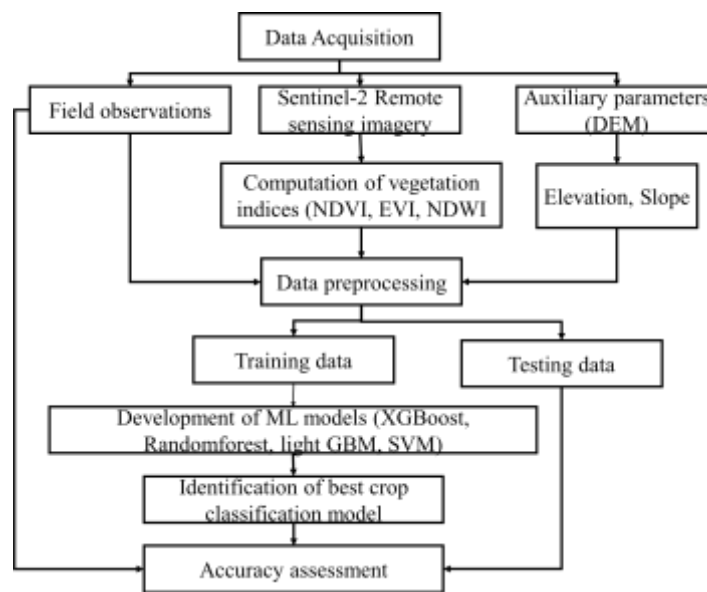


Fig 4: Methodology

and analysis as suggested by Sentinel-2 crop mapping studies. The selected bands include visible (B2, B3, and B4) and near-infrared (B8) bands at 10 m resolution, and red-edge bands (B5, B6, B7) originally available at 20 m resolution. These bands are resampled to 10 m using bilinear interpolation to preserve the consistency and spatial continuity across the data. Following the resampling, radiometric calibration and atmospheric correction is performed on all selected bands using the Sen2Cor processor to produce surface reflectance values. The entire preprocessing and selection are implemented in a cloud-based platform google earth engine.

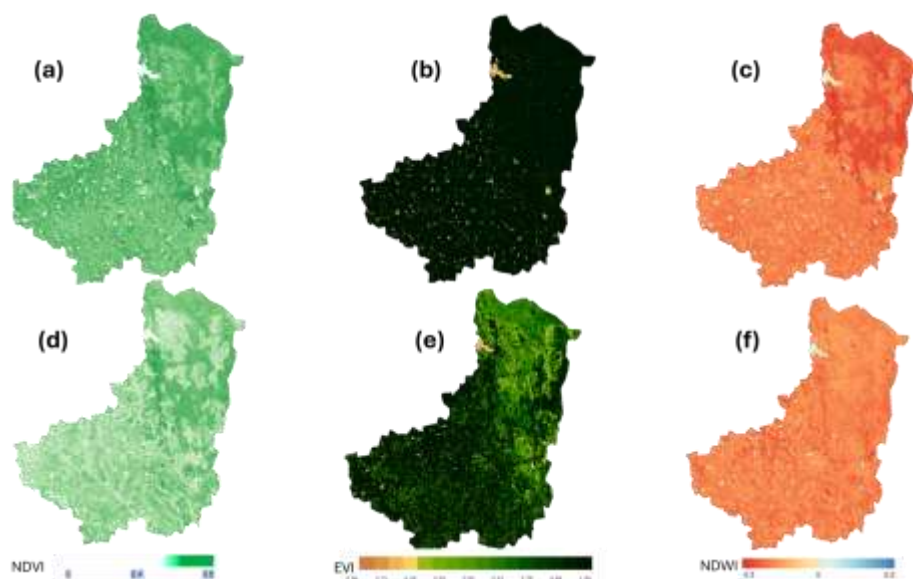


Fig 5: Vegetation Indices used for developing machine learning models (in Kharif): a) NDVI b) EVI c) NDWI; (in Rabi) d) NDVI e) EVI f) NDWI;

2.3.2. Computation of vegetation indices

Three different vegetation indices were calculated and used as input features in the machine learning approaches to improve crop discrimination. These indices include Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Normalized Difference Water Index (NDWI). The average indices over the study region during the Kharif and Rabi seasons are depicted in Fig.5. Overall vegetation greenness can be quantified using the Normalized Difference Vegetation Index (NDVI). It is calculated as a normalized band combinations ratio derived from near-infrared (B8) and red (B4) bands and is used for quantifying the vegetation health and greenness. EVI exhibits sensitivity to high biomass regions and reduce the impact of soil background effects and atmospheric interference. Water content and stress conditions of crops can be assessed using NDWI by utilizing the spectral contrast between green and near-infrared bands. Apart from the vegetation indices, additional features were generated using the auxiliary datasets. These features include elevation, slope, aspect, and topographic wetness index (TWI) which are calculated from the digital elevation model. These features provide contextual information about the local terrain conditions that affect the drainage and crop growth.

2.3.3. Training of machine learning models

Different machine learning models compared in this study are Random forest, Support Vector Machine (SVM), XGBoost and Light GBM. These selected models in this study have already demonstrated their effectiveness in crop classification in many previously reported studies. These models are configured with the following settings for performing classification of paddy and maize crops. Random forest approach is implemented with 500 trees, with tree depth of six, and minimum no of samples to split are eight. Random forest algorithm adopts an ensemble learning scheme for training the model to generate balanced results which can be generalized. SVM classifier is implemented with Radial Basis Function (RBF) kernel for training the model. SVM model is highly suitable for distinguishing classes which are non-linear in nature and can be separated mainly with non-linear conditions. XGBoost and Light GBM approaches emphasize on gradient boosting with a focus on preventing overfitting using regularization parameters.

The remote sensing observations (i.e., Sentinel-2 data), derived vegetation indices, topographical parameters were paired based on the geolocation. This dataset was split into training and validation sets with proportions of 70% and 30% respectively. Testing data were excluded while training the machine learning models. Following that, a 5-fold cross-validation method was performed while training the

model to evaluate its stability and generalization. Grid search optimization was employed for hyperparameter optimization of different model parameters for each algorithm to determine optimal configurations. The parameters optimized for random forest are the number of trees, maximum depth and minimum samples per leaf. Whereas for SVM model, C parameter and RBF kernel coefficient (i.e., gamma value) are selected with hyperparameter tuning. Learning rate, maximum tree depth, and number of estimators are optimized in the case of XGBoost and Light GBM approaches. Hyperparameter optimization for each algorithm was performed using grid search optimization and evaluated using cross-validation based on the classification accuracy metrics.

2.3.4. Evaluation of the models

A comprehensive evaluation framework is established for determining quantitative accuracy metrics and spatial patterns of accuracy and errors to assess the performance of different machine learning classification models. Training samples dataset is used for determining the accuracy of the machine learning approaches. Classification accuracy is quantified using multiple statistical measures such as overall accuracy, Kappa coefficient, etc., to evaluate and compare between machine learning models implemented in this study. Overall classification accuracy as well as class-wise accuracy are calculated for each machine learning model. Overall accuracy is calculated as the total number of correct pixels divided by all the pixels in validation dataset. Additionally, class-wise accuracies are computed for both paddy and maize crops to obtain insights into the model performance, as these have heterogeneous spectral and temporal dynamics.

In addition to classification accuracy, Kappa coefficient and F1 score are used to evaluate the accuracy of the classification. Kappa metric is particularly useful in this study, where the class distributions are imbalanced between paddy vs maize cultivation. F1 score is the harmonic mean of precision and recall, provides details about the performance of a classification model with respect to both false positives and false negatives. For each classifier, we also provide a detailed confusion matrix that allows us to analyze the patterns of misclassifications between crop types.

Table 2: Accuracy metrics for different classification algorithms

Algorithm	Overall Accuracy (%)	Kappa Coefficient	F1-Score	Producer's Accuracy (%)	User's Accuracy (%)
XGBoost	92.3	0.89	0.91	93.5/91.1	92.8/90.7
Random Forest	89.7	0.86	0.88	90.8/88.6	89.9/88.2
Light GBM	88.4	0.84	0.87	89.2/87.6	88.7/87.1
SVM	86.5	0.82	0.85	87.3/85.7	86.9/85.2

The classification accuracy was further assessed by systematically analysing the spatial pattern of classification error over the study area. This analysis helps understanding the region and conditions where the classification model is performing better or poor.

The impact of field size on the classification accuracy is also determined by comparing the misclassification amounts when related with field size. This analysis gives insights into the goodness of classification models especially in heterogeneous cropping communities. Similarly, the impact of elevation, slope and soil type are also determined by analysing the classification accuracy with auxiliary parameters. Furthermore, temporal stability analysis is conducted on the selected machine learning model while performing crop classification by analysing the classification accuracy on different time frames (i.e., classification of different date imagery) during the study period. This analysis across times provides insight into periods of interest for crop discrimination and demonstrates the effect of phenological stages on classification performance. The overall crop classification assessment provides comprehensive details about the performance and statistical validation of different machine learning crop classification models and depicts the spatial patterns of the selected model performance.

3. RESULTS AND DISCUSSIONS

3.1. Comparison of different machine learning models performance

The comparison of different machine learning algorithms—XGBoost, Random Forest, Light GBM, and Support Vector Machine (SVM) models' performance highlights the difference in their classification abilities in discriminating the paddy and maize crops. The crop classification results from different models are shown in Figure 6. The results of comparative analysis based on overall accuracy, Kappa coefficient, F1-score, producer's accuracy, and user's accuracy are summarized in Table 2. The overall and crop level classification accuracy are shown in Figure 7.

Among the compared models, XGBoost has consistently outperformed the others, achieving the highest overall accuracy (92.3%) and a Kappa coefficient of 0.89, indicating strong agreement with reference field observations. The model also exhibited high accuracy when evaluated for individual crops, with producer's accuracy of 93.5% for paddy and 91.1% for maize, and user's accuracy of 92.8% and 90.7%, respectively. These metrics indicates majority of the XGboost predictions are highly accurate and reliable. The random forest model ranks second with an overall accuracy of 89.7%, and a slightly lower class-wise performance compared to XGBoost.

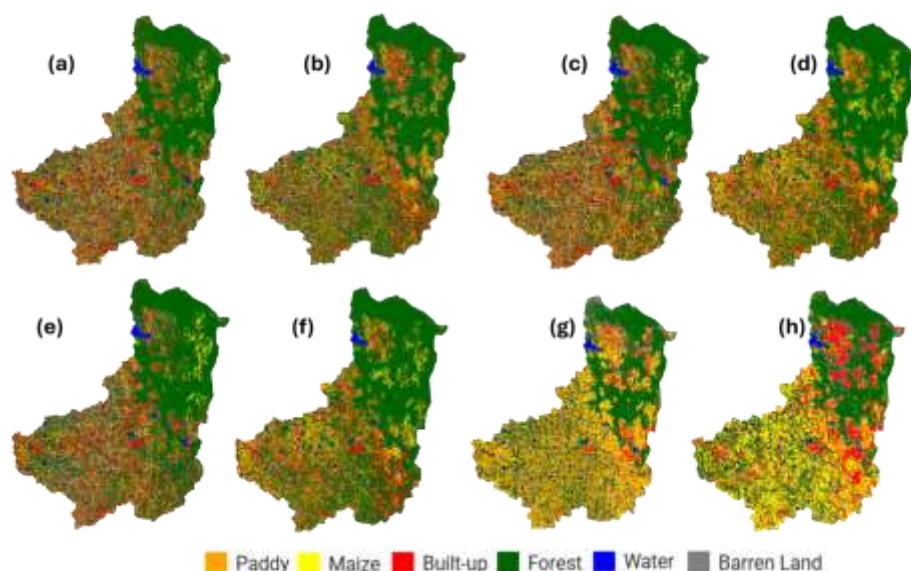


Fig 6: Crop classification over mahabubabad: a) XG Boost b) Random forest c) Light GBM d) SVM for Kharif season; e) XG Boost f) Random forest g) Light GBM h) SVM for Rabi season

However, for both the crops producer's and user's accuracy are above 88%, indicating the reliability of random forest model. Light GBM and support vector machine have shown overall accuracies of 88.4% and 86.5%. While Light GBM is computationally efficient, its accuracy is lower than XGBoost and random forest models. SVM model has shown lowest performance among the compared models. Despite its theoretical strengths, the decision boundary used in the SVM model could not achieve better performance than the other models. These models' performance hierarchy remained consistent across different metrics i.e., kappa coefficient, F1-score, producer's and user's accuracy for validation as shown in Table 2. Class-level accuracy and overall metrics also exhibit similar pattern with XGBoost model outperforming the other approaches as shown in the Figure 7.

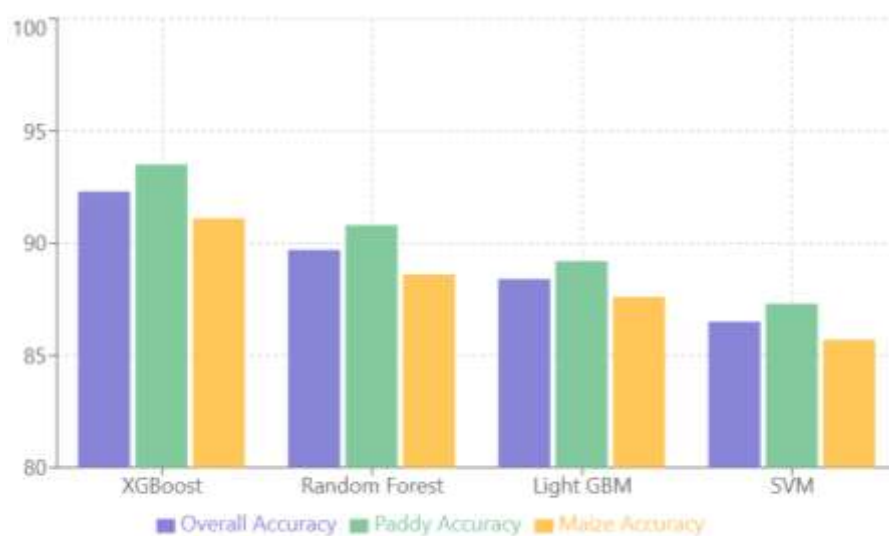


Fig 7: Comparative performance of machine learning algorithms showing overall accuracy and crop-specific classification accuracies for paddy and maize.

3.2. Analysis of classification accuracy with field size

The spatial patterns in classification errors and accuracy details are further assessed based on various factors such as field size, location, etc. With regard to location no distinguishable pattern. The classification errors are randomly distributed over the entire region. Whereas, field size has some association with classification accuracy. Therefore, overall accuracy is assessed against the field size using the reference data for the best performing model i.e., XGBoost. The spatial distribution characteristics of classification accuracy did not exhibit any specific regional variations over Mahabubabad region. However, high-accuracy in this region generally correspond with contiguous large-scale farmed areas, whereas errors are observed across fragmented agricultural landscapes. The results indicate, classification accuracy decreased in areas having mixed cropping patterns and with smaller field sizes (< 0.5 hectare). When field size is 0.1-0.5 hectare, the overall accuracy is less than 88% and varied between 77-88%. Whereas for fields above 0.5 hectare, the overall accuracy is more consistent varying in a small range between 88-92% for the XGBoost model (See in Fig 8). This is particularly pronounced in areas of mixed cropping, since the complexity of such agricultural patterns poses challenge for spectral classification of crops with remote sensing datasets.

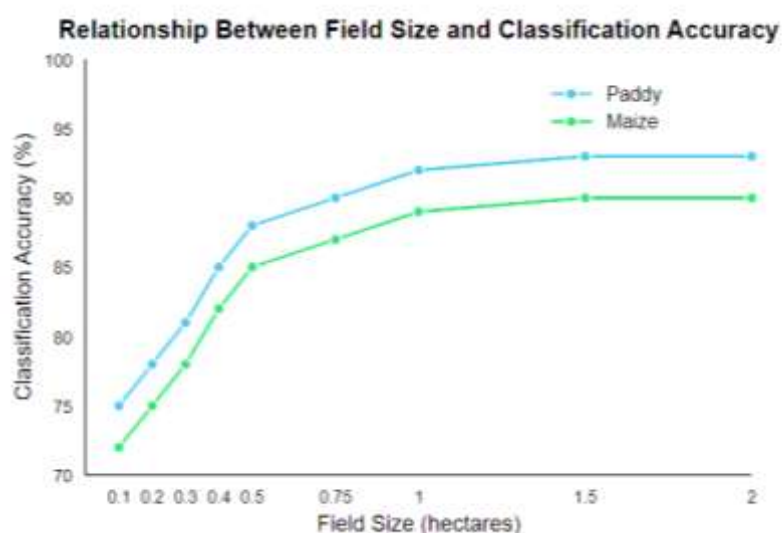


Fig 8: Field size impact map

3.3. Temporal analysis of accuracy

The results from temporal analysis indicates misclassification increases during crop transition periods, particularly in late-season maize fields where leaf senescence patterns lead to spectral confusion with harvested paddy fields. Temporal trends in classification performance can be observed in the seasonal accuracy assessment.

The classification accuracy varied between 80-90% for both the crops depending on the time the imagery is acquired for crop classification in both kharif and rabi seasons as represented in the figure 9. The classification accuracy is observed to be highest during the peak growing periods for both maize and paddy in kharif as well as rabi seasons; However, the overall classification uncertainty increases during the transition period between Kharif and Rabi. These insights are essential to determine the image acquisition timing for accurately classifying the crops for various applications.

3.4. Analysis with respect to auxiliary parameters

The classification accuracy is significantly affected by environmental factors, like temperature and humidity. The performance of XGBoost model is analysed with topographic and auxiliary parameters i.e., slope, soil type, and field size. The results indicates that overall accuracy decreases by 8-12% on slopes larger than 15 degrees. The details regarding the impact of auxiliary parameters on crop classification with XGBoost model are summarized in Table 3. Further, accuracy is analysed with underlying soil types of agricultural landscapes. The results depict regions with black cotton soils reporting 7% higher accuracy than red sandy soils. The larger fields have exhibited higher accuracy compared to smaller fields. The overall accuracy decreased by 15 to 20 percent when field size is less than 0.5 ha.

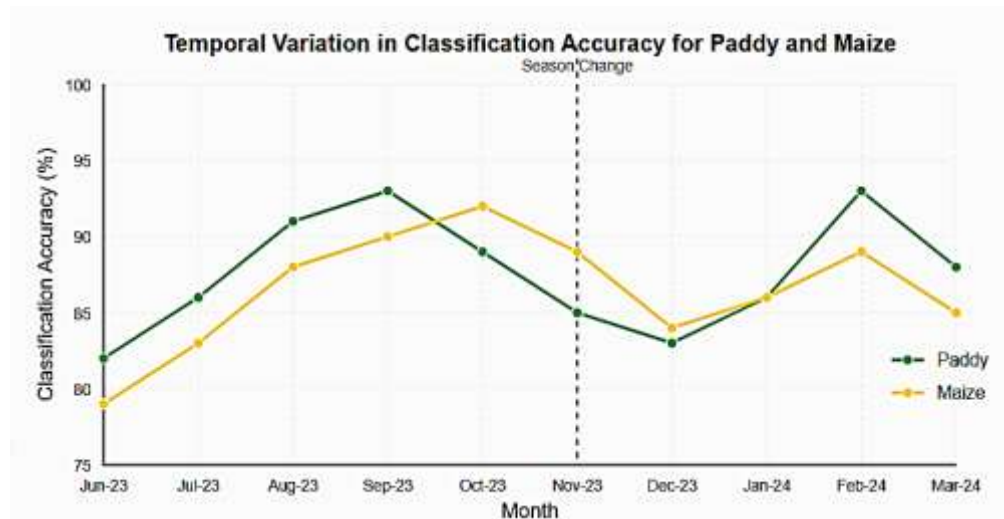


Fig.9. Temporal variation in classification accuracy for paddy and maize throughout the growing season

Table 3: Summary of different factors affecting classification accuracy

Auxiliary Parameter	Impact on Accuracy (%)	Key Observations
Slope > 15°	-8 to -12	Higher misclassification on steeper slopes
Black Cotton Soil	+7	Better accuracy compared to red soils
Field Size < 0.5 ha	-15 to -20	Significant decrease in small fields

4. CONCLUSIONS AND SUMMARY

This study implements different types of machine learning approaches in the mapping of paddy and maize crops in a semi-arid region i.e., Mahabubabad. The results from this study demonstrates that among the different approaches tested, XGBoost and Random Forest have high classification accuracy (92.3% and 89.7% respectively). Overall, these results demonstrate a significant advancement over

conventional classification methods, especially in peri-urban and rural regions where mixed crop systems and spatial variability in field sizes are common.

The methodology implemented in this study offers different useful insights in agricultural remote sensing. In addition to Sentinel-2 data, different auxiliary parameters are used in this for addressing the environmental effects in semi-arid regions where crop spectra are impacted by moisture stress. The classification accuracy is further investigated with respect to different parameters such as slope, soil type, field size, etc. This analysis indicated that field size and soil type have strong implications on classification accuracy in semi-arid regions such as Mahabubabad. The variability in accuracy of crop classification is more prevalent in areas with mixed cropping patterns. Although XGBoost has higher accuracy than Random Forest, the later approach is much simpler and computationally less demanding. The results indicate that although increased accuracy can be achieved through more complex implementations, the choice of crop classification methods should be determined with consideration of the operating environment conditions. Despite the various useful insights, the study has certain limitations which can improved in the future research. Other crop patterns are not considered in this study, which can lead to classification errors in areas with mixed crop patterns. Additionally, the impact of other auxiliary parameters such as streams, rainfall can be investigated. Cloud cover poses challenges during the monsoon period. This can be overcome with help of active microwave remote sensing datasets. Future studies can investigate the synergy between active microwave and multispectral remote sensing datasets for crop classification. The usefulness of deep learning models to deal with complex spatial and spectral classification can be. Additionally, further studies should aim to develop hybrid approaches that leverage the strengths of multiple algorithms while overcoming the specific limitations found in different agriculture landscapes. Overall, the investigation presents insights that are useful for understanding crop mapping dynamics in semi-arid regions, particularly through analyzing spatial patterns and accuracy variations by season.

The research fills the void between scientific understanding of these relationships and practical applications in the operational context of agricultural remote sensing by quantifying the relationships and offering usable checklists for farmers to implement on the ground. These findings underscore the need to integrate additional factors within the technical research phase of new operational crop mapping systems beyond algorithmic performance aspects, which in this case focus on practical implementation constraints. The results indicate that although increased accuracy can be achieved through more complex implementations, the ultimate choice of classification methods should be determined with consideration of the operating environment. Further studies should aim to create hybrid approaches that leverage the strengths of multiple algorithms whilst also overcoming the specific limitations found in specific landscape contexts. The overall conclusion is that the mapping performance of crop type in semi-arid regions needs to strike the right balance between accuracy, computational efficiency, and practical applicability. This study shows that data balance of these classes can be achieved as long as they are appropriately selected with respect to algorithm choice, feature engineering and environmental classification performance determinants.

Author Contributions: “Conceptualization, P.V.S.G.R. and M.A.R.; methodology, P.V.S.G.R. and M.A.R.; software, X.X.; validation, P.V.S.G.R.; formal analysis, P.V.S.G.R.; investigation, P.V.S.G.R.; resources, P.V.S.G.R. and M.A.R.; data curation, P.V.S.G.R; writing—original draft preparation, P.V.S.G.R; writing—review and editing, P.V.S.G.R; visualization, P.V.S.G.R.; supervision, M.A.R.; project administration, M.A.R. All authors have read and agreed to the published version of the manuscript.”

Funding: This research received no external funding.

Institutional Review Board Statement: ~NA~

Informed Consent Statement: Not applicable.

Acknowledgments: The author sincerely than European Space Agency (ESA) and National Aeronautics and Space Administration (NASA) for freely providing different remote sensing datasets used in this study.

Conflicts of Interest: "The authors declare no conflicts of interest."

REFERENCES

1. Arafat, S.M., Aboelghar, M. and Ahmed, E.F., 2013. Crop discrimination using field hyper spectral remotely sensed data. *Advances in Remote Sensing*, 2(2), pp.109–117.
<https://doi.org/10.4236/ars.2013.22009>
2. Büchi, L., Wendling, M., Mouly, P. and Charles, R., 2018. Comparison of visual assessment and digital image analysis for canopy cover estimation. *Agronomy Journal*, 110(1), pp.266–272.
<https://doi.org/10.2134/agronj2017.11.0679>
3. D. S. Thosar, R. D. Thosar, P. B. Dhamdhere, S. B. Ananda, U. D. Butkar and D. S. Dabhade, "Optical Flow Self-Teaching in Multi-Frame with Full-Image Warping via Unsupervised Recurrent All-Pairs Field Transform," 2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI), Wardha, India, 2024, pp. 1-4, doi: 10.1109/IDICAIEI61867.2024.10842718.
4. P. B. Dhamdhere, D. S. Thosar, S. B. Ananda, R. D. Thosar, D. S. Dabhade and U. D. Butkar, "A Semantic Retrieval System Using Imager Histogram Computation to reduce Trademark infringement based on the conceptual similarity of text," 2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI), Wardha, India, 2024, pp. 1-6, doi: 10.1109/IDICAIEI61867.2024.10842701.
5. Butkar, U. (2016). Identity Based Cryptography With Outsourced Revocation In Cloud Computing For Feedback Management System For Educational Institute. *Internatinal Journal Of Advance Research And Inovative Ideas in Education*, 2(6).
6. Ghosh, P., Mandal, D., Bhattacharya, A., Nanda, M.K. and Bera, S., 2018. Assessing crop monitoring potential of Sentinel-2 in a spatio-temporal scale. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5, pp.227–231.
<https://doi.org/10.5194/isprs-archives-xlii-5-227-2018>
7. Gumma, M.K., Nelson, A., Thenkabail, P.S. and Singh, A.N., 2011. Mapping rice areas of South Asia using MODIS multitemporal data. *Journal of Applied Remote Sensing*, 5(1), 053547.
<https://doi.org/10.1117/1.3619838>
8. Hudait, M. and Patel, H.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(2), pp.273–280. <https://doi.org/10.1016/j.ejrs.2022.01.004>
9. K, V., Halappanavar, N.H., Mukkannavar, S.R. and Nishant, N., 2022. Comparative study of machine learning approaches on cropland mapping. 2022 6th International Conference on Computing, Communication, Control and Automation (ICCUBE), pp.1–6.
<https://doi.org/10.1109/csitts57437.2022.10026372>
10. A. K. Bhaga & Et.al "Internet Traffic Dynamics in Wireless Sensor Networks," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1081-1087, doi: 10.1109/ICACITE57410.2023.10182866.

11. Khanal, S., Kc, K., Fulton, J.P., Shearer, S.A. and Ozkan, E., 2020. Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sensing*, 12(22). <https://doi.org/10.3390/rs12223783>
12. Kordi, F. and Yousefi, H., 2022. Crop classification based on phenology information using time series of optical and SAR images. *Remote Sensing Applications: Society and Environment*, 26, p.100812. <https://doi.org/10.1016/j.rsase.2022.100812>
13. Kyere, I., Astor, T., Graß, R. and Wachendorf, M., 2020. Agricultural crop discrimination in a heterogeneous low-mountain region using multi-temporal and multi-sensor data. *Computers and Electronics in Agriculture*, 178, p.105864. <https://doi.org/10.1016/j.compag.2020.105864>
14. Li, D., Gajardo, J., Volpi, M. and Defraeye, T., 2022. Using machine learning to generate open-access cropland maps in the Indian Himalayan region. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2203.14673>
15. Mandal, S.K., 2016. Role of remote sensing and GIS in agriculture. *Journal of Advances in Agriculture*, 7(1), pp.1105–1111.
16. Migdall, S., Brüggemann, L. and Bach, H., 2018. Earth observation in agriculture. In: W. Wagner and B. Székely, eds., *Remote Sensing and GIS for Sustainable Forestry*, Cham: Springer, pp.205–230. https://doi.org/10.1007/978-3-319-74805-4_9
17. Montibeller, B., Del’Arco Sanches, I., Barreto Luiz, A.J., Gonçalves, F. and Aguiar, D.A., 2019. Spectral-temporal profile analysis of maize, soybean and sugarcane using Landsat-8 data. *Brazilian Journal of Agriculture*, 94(3), pp.361–375. <https://doi.org/10.37856/bja.v94i3.3612>
18. Rao, P., Zhou, W., Bhattarai, N., Srivastava, A.K., Singh, B., Poonia, S.P., Lobell, D.B. and Jain, M., 2021. Mapping crop type of smallholder farms using Sentinel-1, Sentinel-2, and Planet imagery. *Remote Sensing*, 13(10), p.1870. <https://doi.org/10.3390/rs13101870>
19. Sagar, B.M. and K, C.N., 2018. Agriculture data analytics in crop yield estimation: A critical review. *Indonesian Journal of Electrical Engineering and Computer Science*, 12(3), pp.1087–1093. <https://doi.org/10.11591/ijeecs.v12.i3.pp1087-1093>
20. See, L., Fritz, S., You, L., Ramankutty, N., Herrero, M., Justice, C.O., Becker-Reshef, I., Thornton, P.K., Erb, K.H., Gong, P., Tang, H., van der Velde, M., Ericksen, P.J., McCallum, I., Kraxner, F. and Obersteiner, M., 2015. Improved global cropland data as an essential ingredient for food security. *Global Food Security*, 4, pp.37–45. <https://doi.org/10.1016/j.gfs.2014.10.004>
21. She, B., Yang, Y., Zhao, Z., Huang, L., Liang, D. and Zhang, D., 2020. Identification and mapping of soybean and maize crops based on Sentinel-2 data. *International Journal of Agricultural and Biological Engineering*, 13(6), pp.142–150. <https://doi.org/10.25165/ijabe.v13i6.6183>
22. Shen, Y., Zhang, J., Yang, L., Zhou, X., Li, H. and Zhou, X., 2022. A novel operational rice mapping method using multi-source satellite images and object-oriented classification. *Agronomy*, 12(12), p.3010. <https://doi.org/10.3390/agronomy12123010>
23. Silleos, N.G., Misopolinos, N. and Perakis, K., 1992. Relationships between spectral indices and crop discrimination. *Geocarto International*, 7(2), pp.43–51. <https://doi.org/10.1080/10106049209354365>
24. Singh, J., Mahapatra, A., Basu, S. and Banerjee, B., 2019. Assessment of Sentinel-1 and Sentinel-2 imagery for crop classification in Indian region during kharif and rabi crop cycles. In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. <https://doi.org/10.1109/IGARSS.2019.8900491>

25. Singh, J., Devi, U.C., Hazra, J. and Kalyanaraman, S., 2018. Crop-identification using Sentinel-1 and Sentinel-2 data for Indian region. In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. <https://doi.org/10.1109/IGARSS.2018.8517356>
26. Singh, S., 2017. Application of geospatial techniques in crop inventory: A review. *International Journal of Research*, 4(1), pp.152–158.
27. Surase, R.R., Kale, K.V., Solankar, M.M., Varpe, A.B., Gite, H.R. and Vibhute, A.D., 2018. Crop discrimination based on reflectance spectroscopy using spectral vegetation indices (SVI). In: *Proceedings of the International Conference on Intelligent Systems Design and Applications*, pp.311–320. https://doi.org/10.1007/978-981-13-9187-3_27
28. Thomson, S.J. and Sullivan, D.G., 2006. Crop status monitoring using multispectral and thermal imaging for aerial platforms. *ASABE Technical Library*, Paper No. 063009. <https://doi.org/10.13031/2013.20657>
29. Wang, X., Fang, S., Yang, Y., Du, J. and Wu, H., 2023. A new method for crop type mapping using multi-source and multi-temporal Sentinel imagery. *Remote Sensing*, 15(9), p.2466. <https://doi.org/10.3390/rs15092466>
30. Zhao, R., Li, Y. and Ma, M., 2021. Mapping paddy rice with satellite remote sensing: A review. *Sustainability*, 13(2), p.503. <https://doi.org/10.3390/su13020503>