Environment and Ecology 42 (4C) : 2037—2043, October—December 2024
Article DOI: https://doi.org/10.60151/envec/HYXU4637
ISSN 0970-0420

Sentinel-1A Data Analysis for Rice Classification Utilizing Random Forests and Support Vector Machine

Pritam Meshram, Kishan Singh Rawat, Vinod Kumar Tripathi

Received 16 August 2024, Accepted 13 November 2024, Published on 27 December 2024

ABSTRACT

Rice is known to be one of the most important crops in India and many other nations, particularly in Asia, therefore accurate rice area estimation has an important role in many activities, ranging from human nutrition to environmental concerns. As a result, the determination of cultivation area remains a hot topic among researchers from numerous disciplines, planners, and decision makers. Using Sentinel-1A SAR (Synthetic Aperture Radar) satellite data, this study attempts to evaluate the effectiveness of random forest (RF) and support vector machines (SVM) algorithms for rice crop classification. According to the findings, rice fields can easily be distinguished from other crops in the research area by using the temporal characteristics of the rice crop as reflected in the VH backscatter patterns. The total precision and Kappa coefficient produced by RF showed 85.7% and 0.74, respectively, when the classification outcomes were compared to the ground reference data. These values were somewhat higher than those obtained by SVM (81.2% overall accuracy and 0.68 Kappa coefficient). The government's rice area statistics were used to compare the analysis results; the relative difference in rice area for RF and SVM, respectively, came out to be +1.40% and -4.63%. In summary, the RF algorithm is highly recommended for the accurate differentiation of rice fields from neighbouring classes in conditions of identical climate, soil, and topography with similar methods of cultivation. On the other hand, Sentinel-1A SAR provides a valuable data set at cost-free for similar studies.

Keywords Sentinel-1A, Rice classification, Random forest (RF), VH Backscatter, Support vector machine (SVM).

Pritam Meshram[1], Kishan Singh Rawat[1]*, Vinod Kumar Tripathi[3]

[1]PhD Student, [2]Professor
[1,2] Department of Civil Engineering Graphic Era deemed to be University, Dehradun, Uttarakhand, India

[3]Associate Professor
Department of Agricultural Engineering, Banaras Hindu University (BHU), Varanasi 221005, India

Email: ksr.kishan@gmail.com
*Corresponding author

INTRODUCTION

Approximately two-thirds of India's population depends on agriculture for their livelihood. Among the cultivated land in India, rice is a major food crop, with the percentage varying from approximately one-third to nearly half. According to the FAO, 70% of rural households in India still primarily rely on agriculture (https://www.nfsm.gov.in). In order to meet the food demands and socioeconomic development of the people in the study region, rice cultivation has increased due to pressures resulting from rapid population expansion. Intensification of rice crops benefits farmers, but it also raises concerns about environmental degradation and decreased agricultural production from degraded soil. So, keeping an eye on rice cultivation developing effective plans for exporting rice grains and ensuring the country's food security requires identifying places where people may get the food they need while protecting the environment.

Low-coarse spatial resolution imagery from ENVISAT and Moderate Resolution Imaging Spectroradiometer (MODIS) were employed in earlier studies on rice crop monitoring (Bouvet *et al.* 2009). The capacity to measure crops phenology parameters for mapping and monitoring rice crops depends on the high temporal resolution of satellite data, which is crucial for understanding the temporal aspects of phenology of rice crops throughout the growing seasons. The typical dimensions of rice fields in the research area were comparatively small (less than one hectare), which presents a challenge for the use of MODIS NDVI satellite product for crop classification and mapping. These challenges are mostly related to mixed-pixel difficulties and cloud cover contamination of the data.

The Sentinel-1A remote sensing satellite (launched on April 3, 2014) allows the mapping of small rice fields at various scales due to its high spatial resolution (10*10 meter) and 12-day (Table 1) revisit period for imaging mode. The data was available for free around the world. There are two C-band polarization modes available for this all-weather radar imaging sensor (VH and VV). Radar imaging systems record backscatter intensities that vary based on object size, shape, orientation, and dielectric constant, unlike optical data. Thus, backscatter intensities can be used to differentiate crops with distinct canopy elements (like leaves, stems, and fruits) and cropping techniques ((like planting or sowing) (Haack 2007). Since studies have indicated that C-band signals mostly interacts with the upper section of canopy layers, biophysical properties of the canopy have been collected using radar data from C-band satellites (Chakraborty *et al.* 2005). The inherent limitations of optical data, such as cloudy atmosphere, have been surmounted with the help of microwave (SAR) data. As a result, SAR data has proven to be a useful tool for crop monitoring, particularly during the rice-growing season when the season is rainy (monsoon) and the atmosphere is generally overcast. According to recent studies (Mishra *et al.* 2017), estimated paddy acreage and yields in the state of Jharkhand using RADARSAT data, and they claimed that the data was more accurate, dependable, and time-efficient than the conventional method. They also found that the accuracy of predicted yields ranged from 52 to 72%, while the accuracy of paddy acreage during the kharif season varied from 78 to 90%. According to (Mansaray *et al.* 2017) In China, have mapped paddy acreage at local scales using multi-temporal Sentinel-1A (SAR) data. The overall classification accuracy was 88%, indicating the operational applicability of Sentinel-1A data in paddy discrimination. This study indicated that the temporal backscatter coefficients of paddy increased steeply from flooding/planting to tillering/booting and then decreased slightly at heading. In a recent study (Torbick *et al.* 2017) created a framework for rice monitoring that maps the spatial extent of rice fields using an RF classifier and multiscale moderate-resolution images. A different study (Nguyen *et al.* 2016) utilized the vertical transmitted and horizontal received (VH) polarization of Sentinel-1A to map paddy rice at a regional scale in the Mekong Delta with an overall accuracy 87.2%. Using machine learning methods like random forests (RF) & support vector machines (SVM), this study investigated whether temporal Sentinel-1A VH backscatter information was appropriate for identifying rice areas in the study region. The possible use of VH backscatter information for rice monitoring had not been validated by previous studies employing Sentintel-1A data.

The statistical machine learning theory (Breiman 2001) is the foundation for the development of the random Forest (RF). This theory has been widely utilized in numerous domains, such as ecology, medical imaging, land-use/cover (LUC) monitoring, and the environment (Otukei and Blaschke 2010), this classifier performed better than decision trees, binary hierarchical classifiers, artificial neural networks, linear discriminant classifiers, and linear discriminant analysis. RF classifier for classification of *kharif* rice using Sentinel-1 SAR data has achieved acceptable results which compared with observed estimated area in the east Vidharbha zone of Maharashtra (Meshram and Rawat 2023). In contrast, another well-known machine learning algorithm that is often utilized for LUC classification in remote sensing, results from mapping produced by RF and SVM classifiers were equally trustworthy (Ghosh *et al.* 2014).

The primary purpose of this research is to create a method for mapping rice area in the eastern zone of Maharashtra jointly utilizing temporal Sentinel-1A VH backscatter data. Data for the rice cropping seasons of 2022 were processed using random forest (RF) and support vector machine (SVM) in order to assess the mapping performance.

MATERIALS AND METHOD

Study area

Bhandara district, situated in the central part of India (Fig. 1), is renowned as a major rice-producing

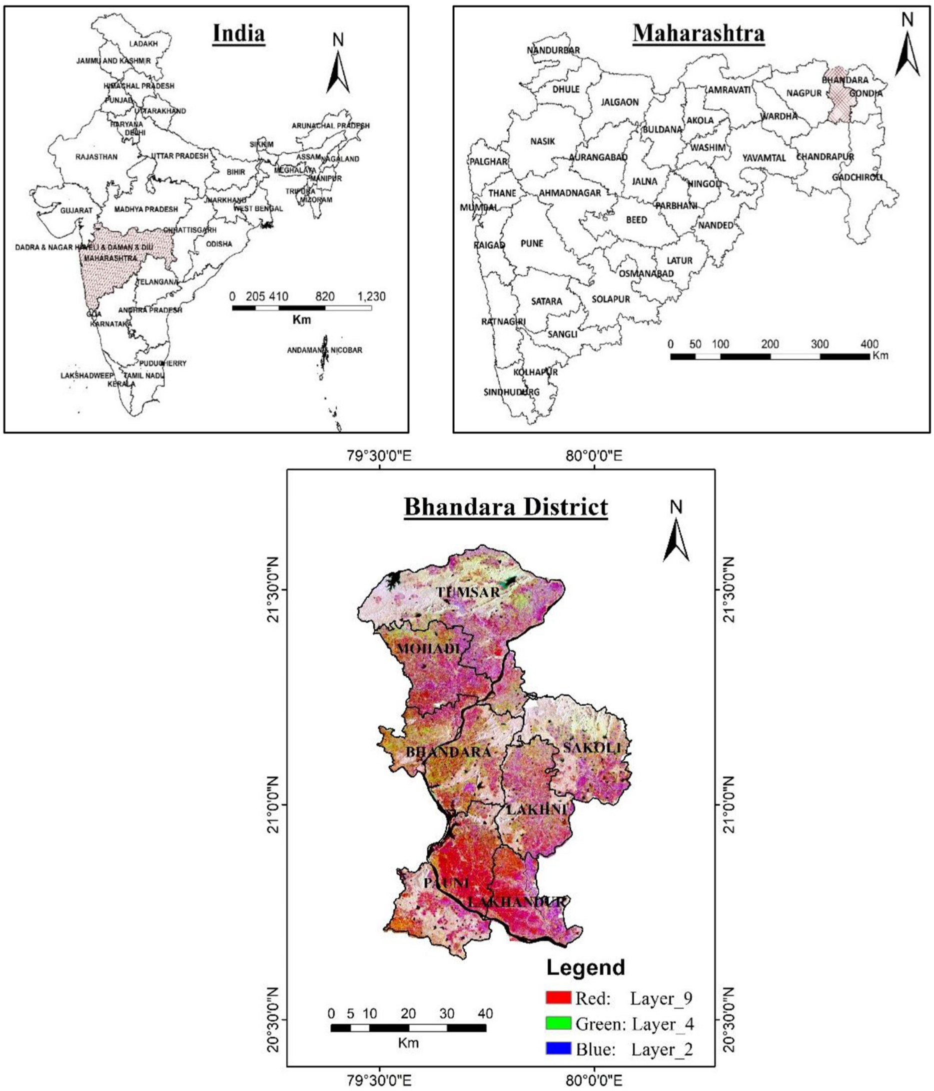

Fig. 1. Location map of Bhandara district in Maharashtra, India.

region in the Vidarbha region and is often known as the 'RICE BOWL' of Maharashtra State. Spanning an area of 3716 square kilometres, it shares its boundary with Nagpur, the second capital of Maharashtra. The study area receives a significant amount of rainfall, ranging between 1200-1300 mm annually, which is the highest in the Vidarbha region. This abundant rainfall is crucial for the flourishing agricultural practices in the district. The region is blessed with two primary rivers, namely the Wainganga and Surnadi, which play a vital role in the agricultural activities of the district. These rivers provide irrigation resources for the farmers, who practice both rain-fed and irrigated agriculture. The irrigation facilities are made available through canals, wells, and tanks, ensuring that the agricultural lands are adequately watered. Soil in Bhandara district is broadly classified into three categories: deep soil comprising 78.9% of the area, shallow soil covering 13%, and moderately deep soil accounting for 8%. This diversity in soil types allows for a wide range of crops to be cultivated, further enhancing the agricultural productivity of the region.

Data collection

Sentinel-1A satellite data

The European Space Agency (ESA) released Sentinel-1 data that spanned the 2022 rice growing season, which ran from June to November. The sensor includes four imaging techniques-interferometric wide sweep (IW), strip map (SM), wave (WV), and extra wide swath (EW) and a 12-day revisit period, the Sentinel-1A was released on April 3, 2014. The frequency range in which the Sentinel-1A operates is 5.405 GHz (C-band) (Table 1), or around 5.6 cm for radar. The S1A_IW_GRDH component was used in this study, and it has a swath range of about 250 km & a spatial accuracy of 10 m. It contains VH and VV polarizations.

Table 1. Specification of Sentinel-1A.

Specification	Sentinel-1A Synthetic Aperture Radar (SAR)
Acquisition duration	June to November 2022
Acquisition orbit	Ascending & Descending
Imaging Mode	IW
Polarisation	VH
Imaging frequency	C-band (5.4 GHz)
Spatial resolution	10 meters
Revisit cycle	12 days

Ground reference data

For the analysis of remote sensing data to distinguish between distinct land uses and land coverings, the acquisition of field survey data is crucial. The two main purposes of conducting field survey activities are (a) Obtaining relevant data and information that may be used as inputs and references when analysing data from remote sensing and (b) To ensure the correctness of what has been discovered and categorised using data from remote sensing. MAPInr, an Android app for smartphones, was used to collect ground data. This app marks the field with a polygon feature. In tabular form, the basic information collected during field survey includes crop type, field location, field size, crop health condition, crop growth stage, crop health condition, field photographs, ground cover, crop transplantation date, expected harvest date, and soil state, among other things. In the year 2022, between July, August and September, field survey data was collected.

METHODS

This investigation used three phases of data processing to categorise the crop of rice in the research area: Pre-processing of Sentinel-1A, (2) categorization of rice crops applying support vector machines (SVM) & random forests (RF), and (3) evaluation of the accuracy of the classification results. This is depicted in a flowchart of the technique (Fig. 2).

Data pre-processing

Sentinel–1Radiometric calibration was done to transform digital pixel readings of VH amplitude as sigma naught (σ°) values which can be directly related to the image's radar backscatter because the data contained errors caused by variations in transmitted power, system noise, receiver gains, and antenna illumination pattern. The σ° data were filtered with the Lee filter (3×3 moving window) since they contain speckles that complicate the analysis of features. Distances in photographs are distorted due to topographical

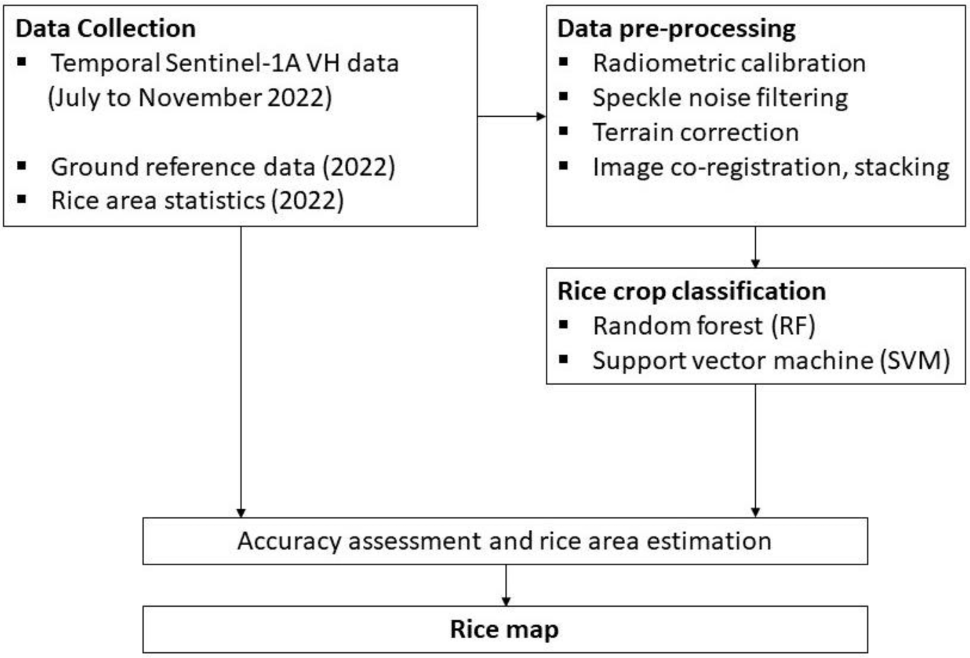

Fig. 2. Methodology flowchart.

variations, the inclination of the satellite sensor, and of side-looking geometry of the photos. Thus, terrain correction was applied to compensate for these inconsistencies. In this work, terrain correction was performed using SRTM DEM (3s resolution). After orthorectification, bilinear interpolation was used to resample the images to a scale of 10 metres. These processing steps were all completed with SNAP 9.0, an ESA-developed program. Next, a subset of the data was used for the research area.

Random forests (RF)

The RF is a highly effective classifier based on a learning technique. The approach employs tree-based classifiers {h(x, Θk), k =1, …}, where {\k} represents independent, identically distributed random vectors and x serves as the input vector. Deterministic bootstrapping is used to build the decision tree, leaving residual data points for validation. The most popular class is then voted on. The RF employs bootstrapping with replacement to increase the diversity of classification trees. Based on which class received the most votes among all the trees in the collection, each pixel is assigned. Compared to boosting techniques, this classifier is faster, quantifies variable importance, is less prone to noise or overfitting, and does not make any distributional predictions about predictors (Rawat *et al.* 2024). To optimise model accuracy while using RF for the classification of images, specify two parameters: the quantity of trees and predictors (Rawat et al. 2024). Higher computation costs are associated with larger tree counts. Training data were analysed, and we found that 100 trees are sufficient to generate RF for classification and maintain consistent out-of-bag mean squared error (OOB-error). Default values were utilized in other RF settings.

Support vector machines (SVM)

A supervised classification technique called Support Vector Machine (SVM) is based on statistical learning theory and usually produces accurate classification results from noisy and complex data. It divides the categories using a decision surface which optimizes the margin of separation. The surface is often referred to as the ideal hyperplane, and the support vectors are the data points that are closest to the hyperplane. The

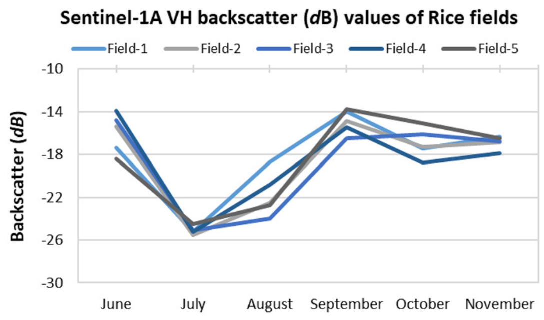

Fig. 3. Sentinel-1A VH backscatter (dB) rice crop profile.

most important components of the training set are the support vectors. By using nonlinear kernels, you can modify SVM to function as a nonlinear classifier. In its most basic form, SVM is a binary classifier, however by merging several binary SVM classifiers, it can perform as a multiclass classifier (generating a binary classifier for every probable pair of classes).

RESULTS AND DISCUSSION

Rice crop VH backscatter intensities increased gradually after sowing, peaking during the heading stage (Fig.3). Backscatter intensities in rice fields observed between -26 to -24 *d*B during the sowing period and -14 to -15 *d*B during the heading stage (about 65 days after planting). A high-density rice field is immediately seeded into wet soil during the sowing duration, with a depth of 2-5 cm of water. This results in relatively low VH backscatter intensities, which may be related to the VH backscatter's cross-polarization, where the pulses partially depolarize.

As the rice plant reaches the heading phase, it grows taller, produces more tillers, and develops leaves. Due to the increased VH contribution from volume scattering of the interior canopies of rice plants, the backscatter intensity during this vegetative phase thus showed a significant rise. Following the heading stage, the rice crop displayed a decline in the number of reproductive and ripening leaves, while backscatter strengths gradually decreased till harvesting. This marked the conclusion of the cropping season as the intensity levelled off. Both in the dry and wet seasons, there was considerable conformity in the backscatter values, and the urban areas and forests, the VH backscatter intensities were comparatively high (- 8 to -10 *d*B). The forest structure including

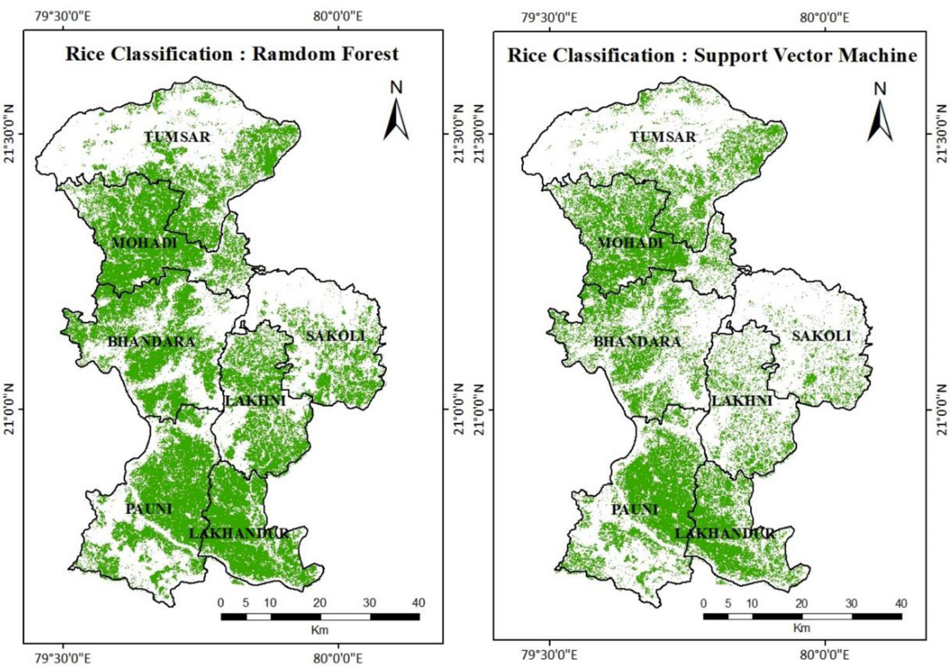

Fig. 4. The Classified Image using RF and SVM.

its canopy cover, tree density, and tree size as well as the climatic factors such as weather, moisture, and weather dynamics determine the degree of sensitivity of the backscatter data to the forest. Throughout the year, water's very low VH backscatter values may be distinguished from those of other LUC types and rice crop.

Similar spatial patterns of rice area were obtained by applying RF and SVM to classify temporal Sentinel-1A VH backscatter data (Fig. 4). Rice occupied the majority of the study region. Rice planting in the study region has increased in recent years in order to increase household income and exports. The accuracy of the mapping techniques was evaluated by comparing the classification maps generated by RF and SVM with the ground dataset used as a reference. It was discovered that the mapping outcomes produced by RF were somewhat better than those produced by SVM. The results of SVM were 81.2% and 0.68, respectively, whereas the overall accuracy & Kappa coefficient obtained by RF were 85.7% and 0.74, respectively. At the district level, it was discovered that the mapping outcomes and government statistics on rice areas were very consistent.

CONCLUSION

In an eastern region of Maharashtra, India, this research examined the use of temporal Sentinel-1A VH backscatter information for rice crop categorization using supervised RF and SVM classifier. The study found that VH backscatter can accurately indicate rice areas in the study location. This study used temporal VH backscatter data to check the correctness of RF & SVM classifiers for mapping rice areas in the region, despite potential concerns with mixed-pixel and border effects lowering accuracy. The results obtained by SVM were 81.20% and 0.68, respectively, whereas the overall accuracy and Kappa coefficient obtained via RF were 85.70% and 0.74, respectively. The mapping outcomes of rice-growing regions, checked with government statistics, confirmed a consistent relationship with relative deviation values of +1.40% and -4.63% for RF and SVM, respectively. This study shows that Sentinel-1A VH data can be used for rice crop classification in an eastern zone of Maharashtra, utilizing RF and SVM. To assess whether the algorithms are appropriate for monitoring rice crops on a broad scale, the procedures might be repeated throughout the rice-growing region of the India.

REFERENCES

Bouvet A, T Le Toan, Lam-Dao Nguyen (2009) Monitoring of the Rice Cropping System in the Mekong Delta Using ENVISAT/ASAR Dual Polarization Data. *Geoscience and Remote Sensing, IEEE Transactions on* 47 (2):517-26.

Breiman L (2001) Random Forests. Machine Learning 45 (1): 5-32. https://doi.org/10.1023/a:1010933404324

Chakraborty M, Manjunath KR, Panigrahy S, Kundu N, Parihar JS (2005) Rice crop parameter retrieval using multi-temporal, multi-incidence angle Radarsat SAR data. *ISPRS Journal of Photogrammetry and Remote Sensing* 59 (5): 310-22. doi: https://dx.doi.org/10.1016/j.isprsjprs.2005.05.001

Ghosh Aniruddha, Joshi PK (2014) A comparison of selected classi fication algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery." *International Journal of Applied Earth Observation and Geoinformation* 26: 298-311. doi:http://dx.doi.org/10.1016/j.jag.2013.08.011

Haack B (2007) A Comparison of Land Use/Cover Mapping with Varied Radar Incident Angles and Seasons. *GIScience & Remote Sensing* 44 (4):305-19. doi: 10.2747/1548-1603.44.4.305

Mishra N, Singh RK, Kumar A, Jeyaseelan A (2017) Rice cultivation monitoring and acreage estimation using RA DARSAT SAR images in Jharkhand. SGVU *Journal of Climate Change and Water* 4 : 1–8.

Mansaray LR, Zhang D, Zhou Z, Huang J (2017) Evaluating the potential of temporal Sentinel-1A data for paddyrice discrim ination at local scales. Remote Sensing Letters 8(10) :967–976. https://doi.org/10.1080/2150704X.2017.1331472

Meshram P, Rawat KS (2023) Detection of paddy fields in Maharashtra's eastern region during the *kharif* season using Sentinel-1A SAR data and geographic information system (GIS). International Conference on Advancement in Computation & Computer Technologies (InCACCT), 10.1109/InCACCT57535.2023.10141834

Nguyen DB, Gruber A, Wagner W (2016) Mapping rice extent and cropping scheme in the mekong delta using Sentinel-1A Data. Remote Sensing Letters 7 (12): 1209–1218. doi:10.1080/2150704X.2016.1225172

Otukei JR, Blaschke T (2010) Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms." *International Journal of Applied Earth Observation and Geoinformation* 12, Supplement 1 (0): S27-S31. doi:10.1016/j.jag.2009.11.002

Rawat KS, Bisht N, Kumar S, Garg N (2024) Statistical comparison of simple and machine learning based Land use and land cover classification algorithms: A case study. *Journal of Water Management Modeling*. DOI: https://doi.org/10.14796/JWMM.H524

Torbick N, Chowdhury D, Salas W, Jiaguo Q (2017) "Monitoring rice agriculture across myanmar using time series Sentinel-1 Assisted by Landsat-8 and PALSAR-2." Remote Sensing 9 (2): 119. doi:10.3390/rs9020119.