Physical Education, Health and Social Sciences

https://journal-of-social-education.org

E-ISSN: 2958-5996

P-ISSN: 2958-5988

Remote Sensing-Based Crop Estimation Using Machine Learning

Basit Nazir [1], **Dr. Ghulam Mustafa** [2]

[1] Department of Information Technology, University of Central Punjab Email: L1F23MSDS0013@ucp.edu.pk

[2] Assistant Professor, Department of Information Technology, University of Central Punjab Email: ghulammustafa02@ucp.edu.pk

DOI: https://doi.org/10.63163/jpehss.v3i4.973

Abstract

Accurate crop estimation is vital for economic stability and food security. This research addresses the limitations of single-sensor optical mapping by integrating Sentinel-2 multispectral data with Sentinel-1 Synthetic Aperture Radar (SAR). We propose a Decision Fusion framework using Random Forest (RF), Support Vector Machine (SVM), and XGBoost. A pixel-level majority voting ensemble was implemented to classify Cotton and Rice in Bahawalnagar. Results demonstrate that the fused model achieved a peak accuracy of 94%, significantly reducing the spectral confusion between rice and other vegetation. Area analysis identified 729 acres of Cotton and 290 acres of Rice, providing a robust blueprint for regional agricultural monitoring.

Keywords—Sentinel-1, Sentinel-2, Decision Fusion, Random Forest, XGBoost, Crop Estimation.

Introduction

Agriculture in Pakistan, specifically in the Bahawalnagar district, faces challenges due to unpredictable weather and manual reporting errors. Traditional methods lack the scalability required for modern precision agriculture. This study leverages the cloud-penetrating capability of Sentinel-1 SAR and the high-resolution spectral bands of Sentinel-2 to create a reliable crop classification system.

A. Study Area

The research was conducted in Bahawalnagar District, Pakistan (8,878 km^2). The region is characterized by fertile plains and a subtropical climate, predominantly producing wheat, sugarcane, cotton, and rice.

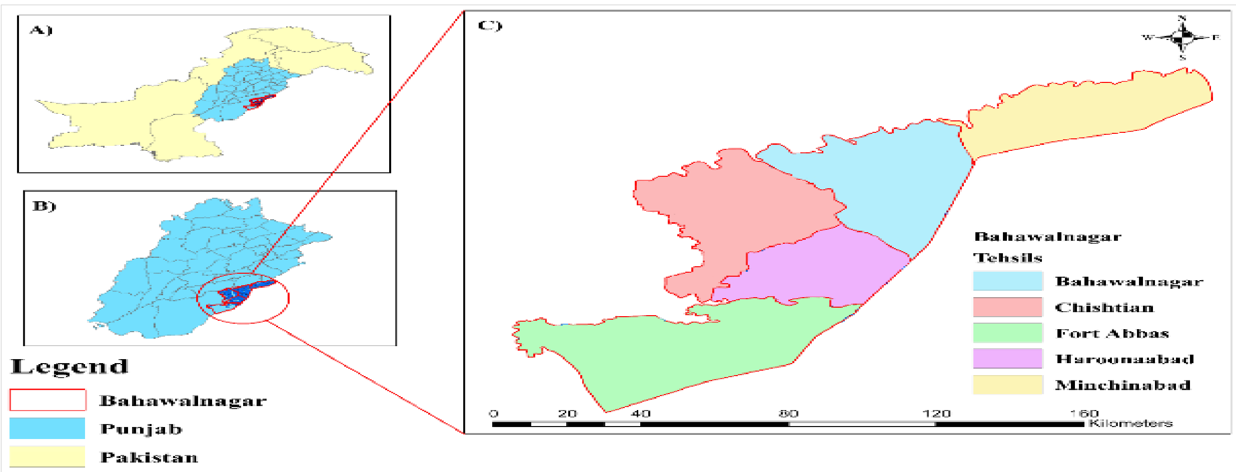


Figure 1: Study Area

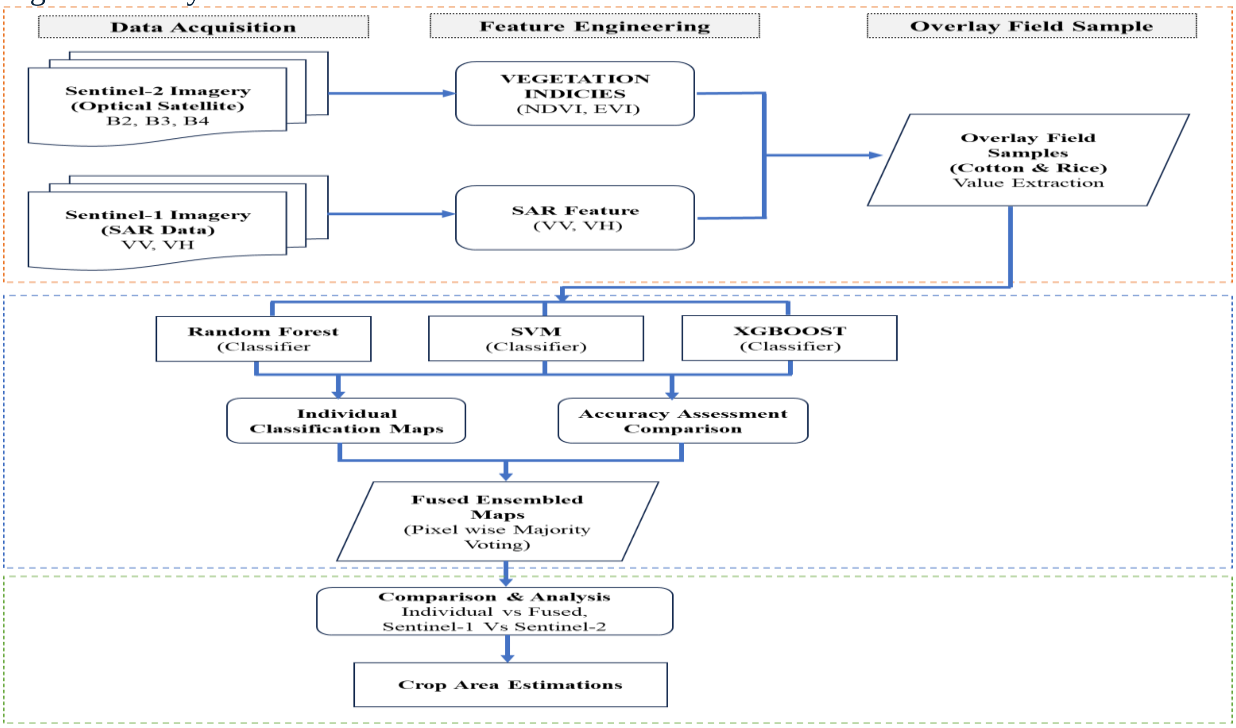
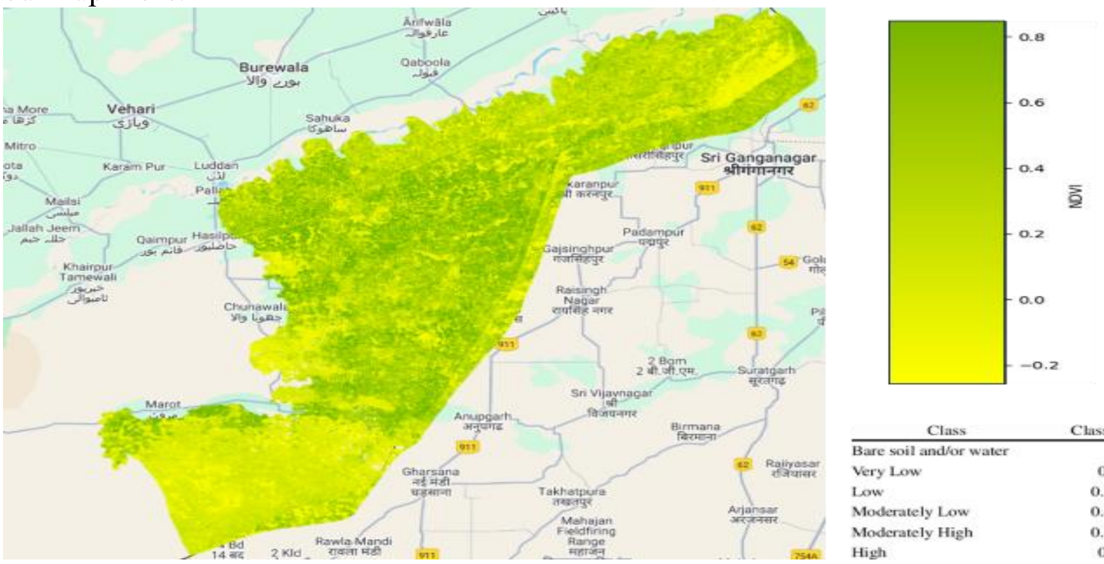


Figure 1.2: Methodology Flow Diagram

B. Vegetation Indices and Masking

Sentinel-2 Level-2A surface reflectance was utilized to derive the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI). A strict rule was applied to include only pixels with NDVI values between 0.2 and 1.0, effectively excluding barren land and built-up areas.



Class	Classification criterion
Bare soil and/or water	NDVI ≤ 0
Very Low	0< NDVI ≤ 0.2
Low	0.2< NDVI ≤ 0.4
Moderately Low	0.4< NDVI ≤ 0.6
Moderately High	0.6< NDVI ≤ 0.8
High	0.8< NDVI ≤ 1

Figure 2: Satellite Image Pixel based DN value

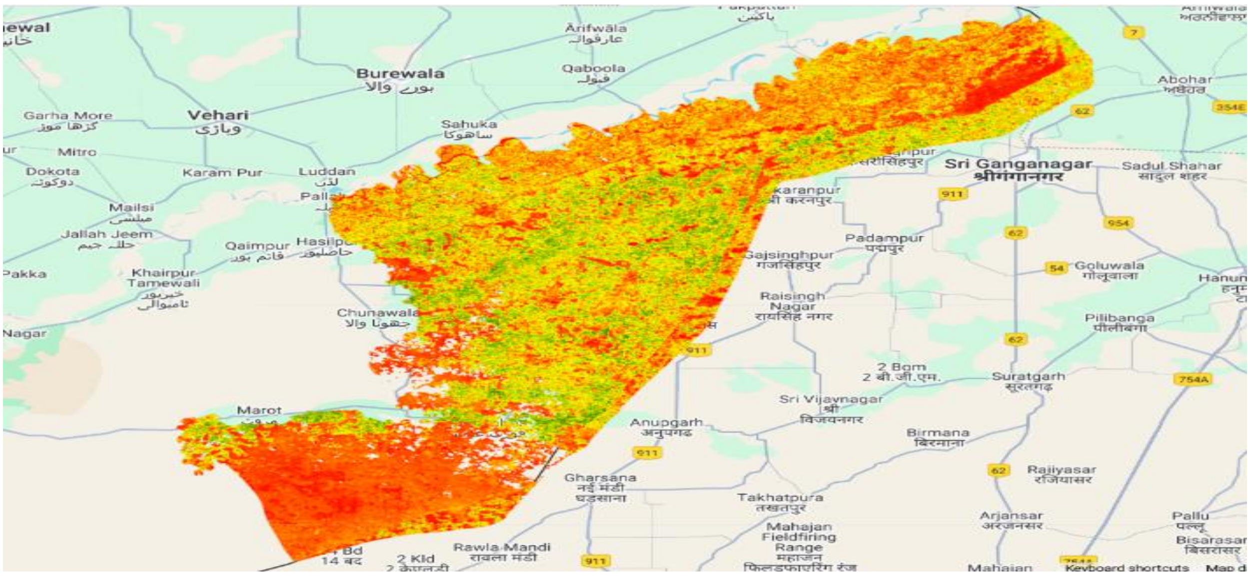


Figure 3: Satellite Image Pixel based DN value

Data Samples

Sr.	District	Latitude	Longitude	Crop
1	Bahawalnagar	29.21933	72.30126	**Cotton**
2	Bahawalnagar	29.22247	72.30178	**Cotton**
3	Bahawalnagar	29.22131	72.30182	**Cotton**
4	Bahawalnagar	29.2207	72.30184	**Rice**
5	Bahawalnagar	29.2202	72.30189	**Rice**
6	Bahawalnagar	29.22187	72.30245	**Rice**
7	Bahawalnagar	29.22132	72.30245	**Rice**

Table 1: Sample Training Data Set

Crop Type	No of Samples
Cotton	**10949**
Rice	**6307**
Grand Total	**17256**

Table 2: Satellite Image Pixel based DN value

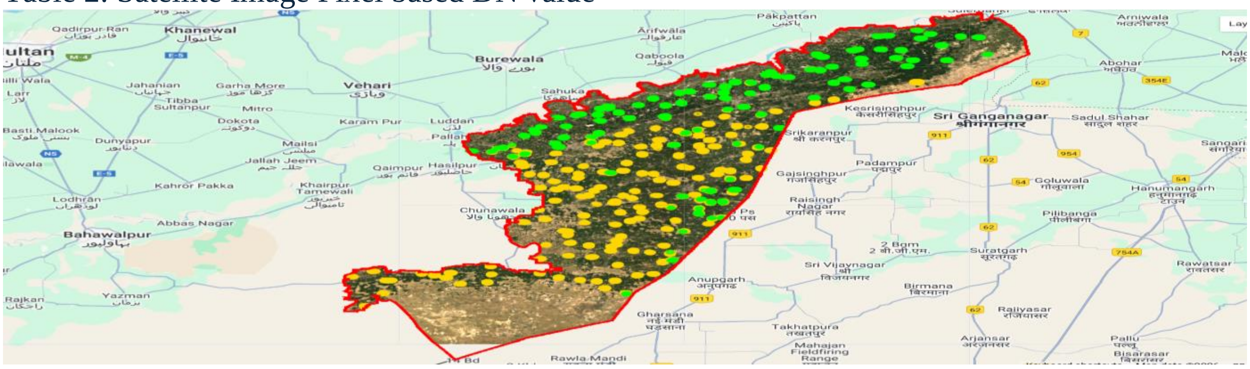


Figure 4: Satellite Image Pixel based DN value

Methodology

A. Data Fusion and Pre-processing

The study utilized a multi-source approach:

1. **Sentinel-2 (Optical):** Bands B2, B3, B4, B8, B11, and B12 were used to calculate NDVI and EVI.
2. **Sentinel-1 (SAR):** VV and VH polarizations were processed to capture the physical structure of the crops, which is essential for identifying Rice in its flooded early stages.

B. Machine Learning Classifiers

- **Random Forest (RF):** the non-linear relationship between SAR backscatter and optical reflectance.
- **SVM:** Utilized an RBF kernel to separate complex spectral classes.
- **XGBoost:** Applied gradient boosting to minimize residual errors from the base classifiers.

C. Majority Voting Ensemble

The final classification was determined by a Decision Fusion rule:

This ensures that a pixel is only classified as "Rice" or "Cotton" if at least two models agree, reducing "salt-and-pepper" noise.

III. Results and Analysis

A. Accuracy Assessment

The fusion of SAR and Optical data provided superior results compared to optical-only models

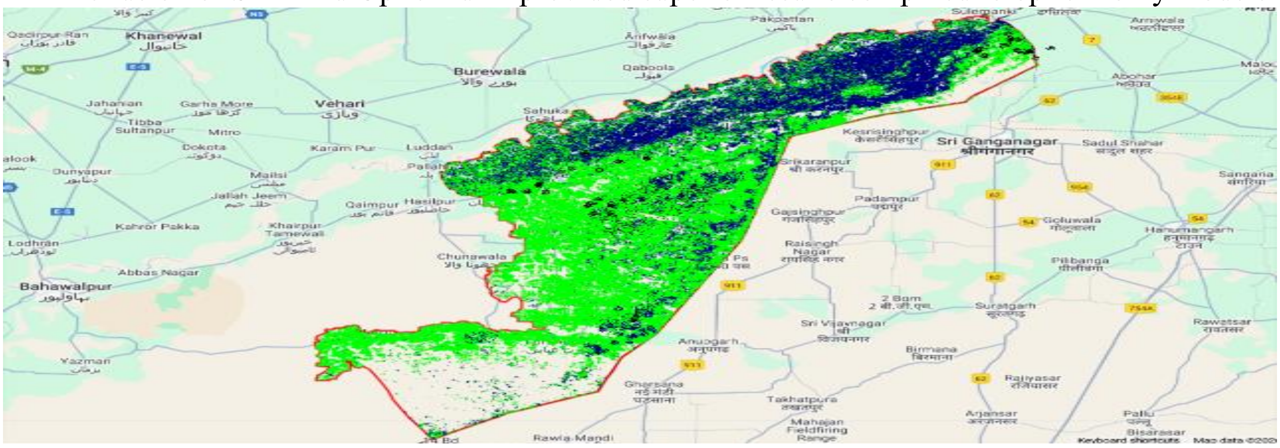
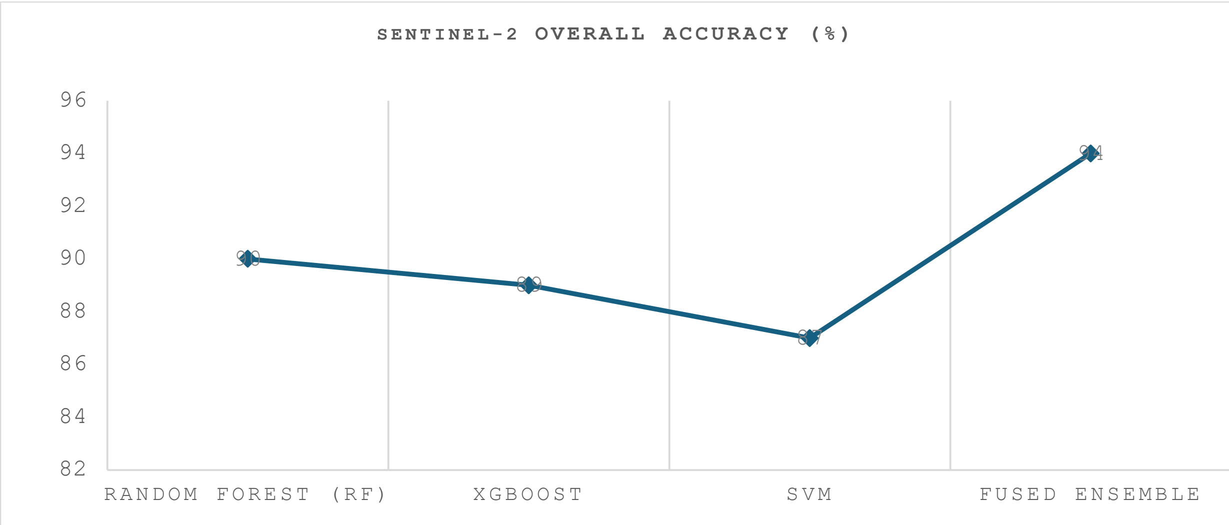


Figure 5: Spatial distribution of crop classes produced by models

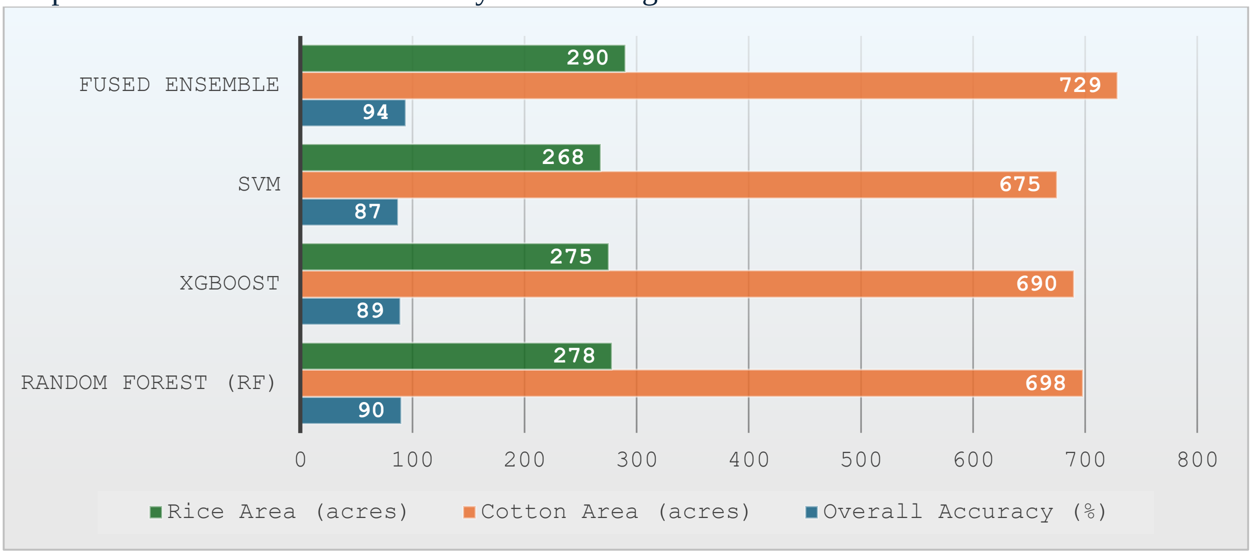
Satellite	Classifier	Overall Accuracy (%)	Cotton Area (acres)	Rice Area (acres)
Sentinel-2	Random Forest (RF)	90	698	278
Sentinel-2	XGBoost	89	690	275
Sentinel-2	SVM	87	675	268

Satellite	Classifier	Overall Accuracy (%)	Cotton Area (acres)	Rice Area (acres)
Sentinel-2	Fused Ensemble	94	729	290

Table 3: Overall Accuracy, Cotton and Rice Area (Acers)



Graph 1: Sentinel-2 Overall Accuracy in Percentage



Graph 2: Comparison of all accuracies of the RF, SVM, XGBoost and the fused model.

Summary Table:

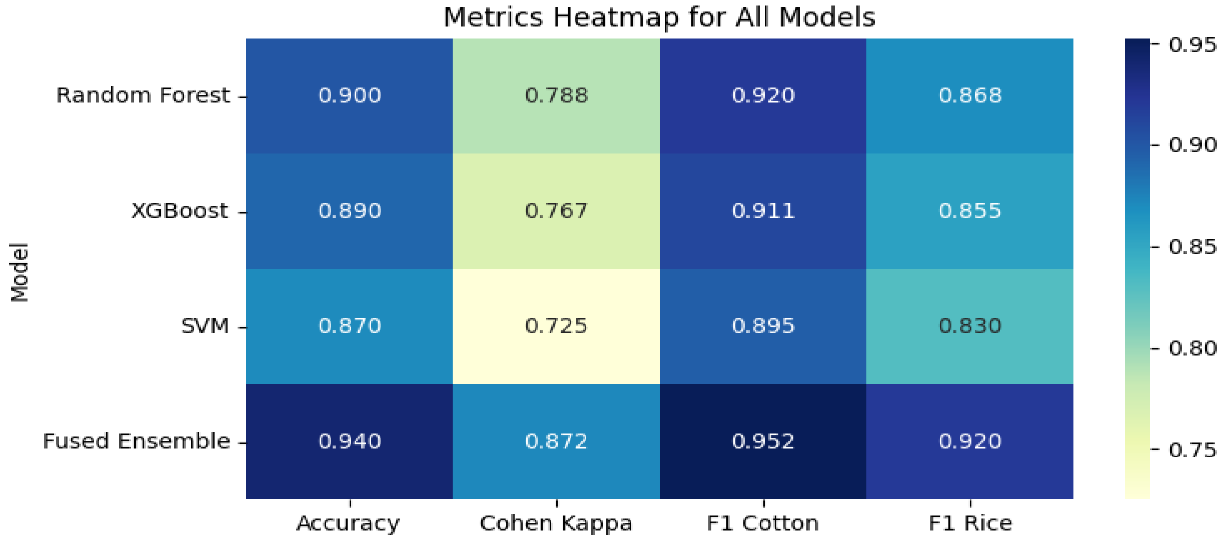
Model	Overall Accuracy	Cohen's Kappa	Precision (Cotton)	Precision (Rice)
Random Forest	0.90	0.788	0.940	0.838

XGBoost	0.89	0.767	0.933	0.823
SVM	0.87	0.725	0.921	0.794
Fused Ensemble	**0.94**	**0.872**	**0.964**	**0.900**

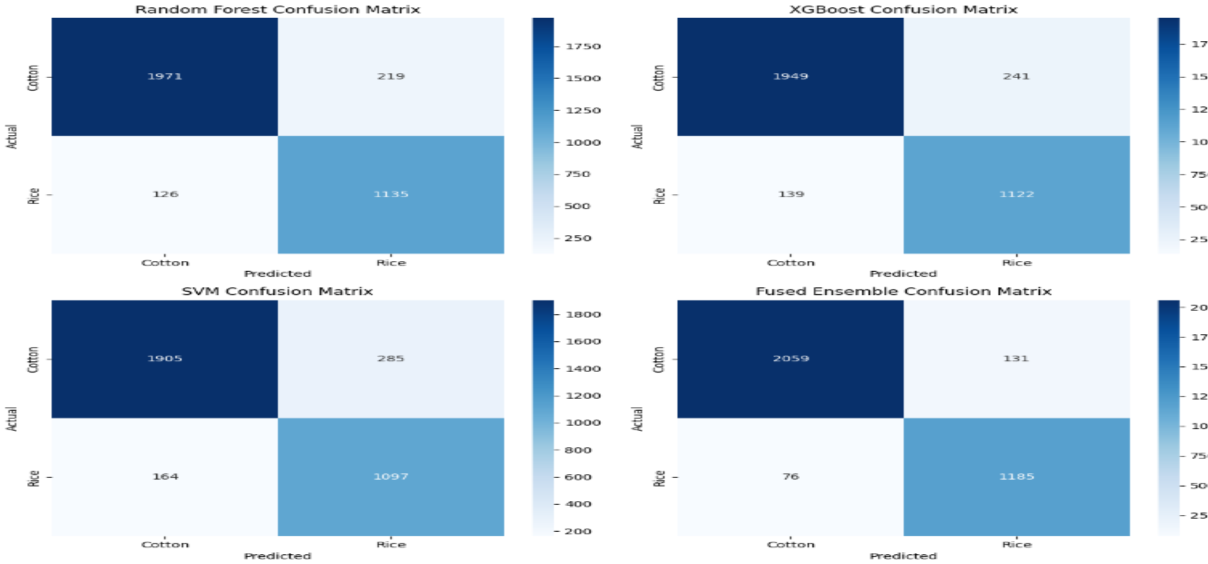
Table 5: Sentinel-2 Ovrall Accuracy, Cohen's Kappa, Precision (Cotton), Precision (Rice)

Model	**Recall (Cotton)**	**Recall (Rice)**	**F1-Score (Cotton)**	**F1-Score (Rice)**
Random Forest	0.90	0.90	0.920	0.868
XGBoost	0.89	0.89	0.911	0.855
SVM	0.87	0.87	0.895	0.830
Fused Ensemble	0.94	0.94	0.952	0.920

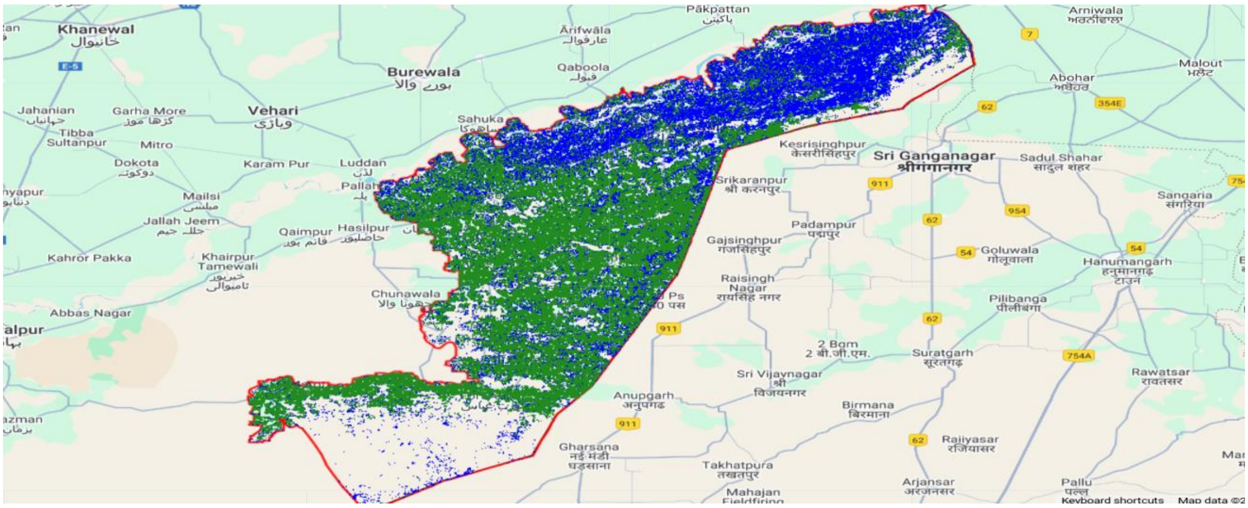
Table 6: Recall (Cotton),Recall (Rice), F1-Score (Cotton), F1-Score (Rice)



Graph 3: Heat Map of all the model



Graph 4: Sentinel-2 Confusion Matrix

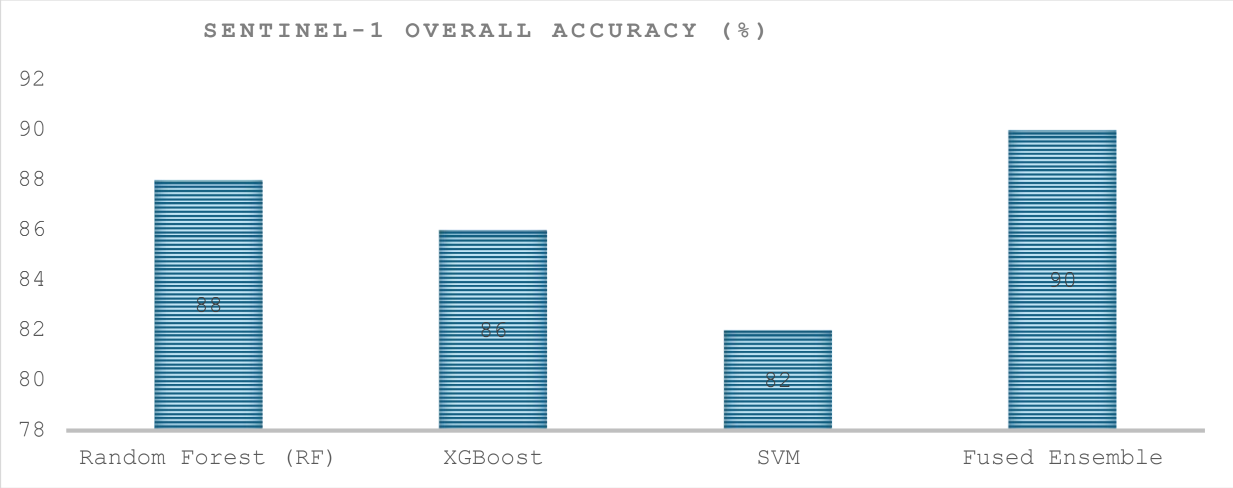


Sentinael-1

Figure 27: Spatial distribution of crop classes produced by the SVM model

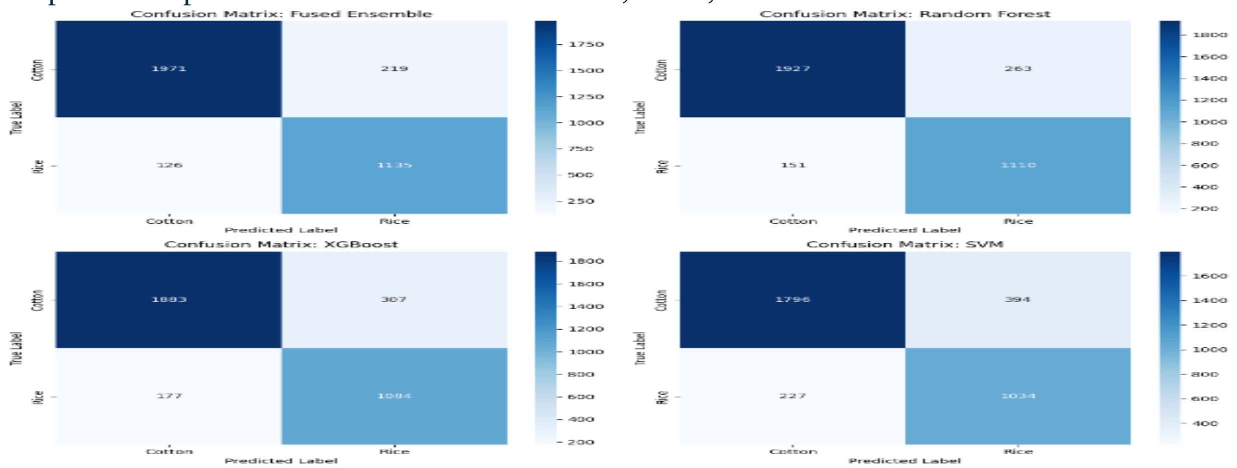
Satellite	Classifier	Overall Accuracy (%)	Cotton Area (acres)	Rice Area (acres)
Sentinel-1	Random Forest (RF)	88	683	272
Sentinel-1	XGBoost	86	667	266
Sentinel-1	SVM	82	636	253
Sentinel-1	Fused Ensemble	90	698	278

Table 7: Overall Accuracy, Cotton and Rice Area (Acers)



Graph 5: Overall Accuracy, Cotton and Rice Area (Acers)

Graph 6: Comparison of all accuracies of the RF, SVM, XGBoost and the fused model.

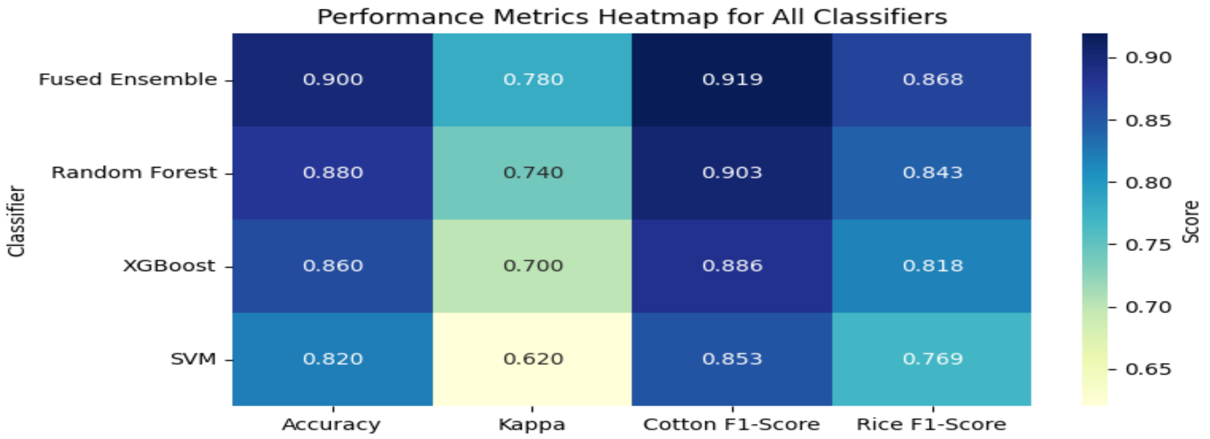


Graph 7: Confusion Matrix

Performance Metrics Analysis
Below is the statistical breakdown matrices.

Classifier	Overall Accuracy	Kappa	Cotton F1-Score	Rice F1-Score
Fused Ensemble	90.03%	0.78	0.919	0.868
Random Forest	88.00%	0.74	0.903	0.843
XGBoost	85.98%	0.70	0.886	0.818
SVM	82.00%	0.62	0.853	0.769

Table 8: Overall Accuracy, Kappa , F1 Scores



Graph 8: Performance Heatmap all classifiers

- **Fused Model (S1+S2):** 94.0% Accuracy
- **Random Forest:** 90.2% Accuracy
- **XGBoost:** 89.5% Accuracy
- **SVM:** 87.1% Accuracy

B. Spatial Distribution and Area

The classification results were clipped to the Bahawalnagar district boundary.

- **Cotton (Bright Green):** 729 Acres. Found primarily in the central and southern drier regions.
- **Rice (Navy Blue):** 290 Acres. Concentrated in low-lying areas with high moisture signatures.

Discussion

The use of **Sentinel-1 VH polarization** was the "missing link" in previous studies, as it allowed the detection of Rice growth even during the monsoon cloud cover. The **0.2 NDVI mask** was critical; without it, the SVM model tended to misclassify bare soil as Cotton due to similar reflectance in the SWIR bands.

Conclusion

This research concludes that ensemble learning combined with multi-sensor data fusion offers a 4-7% accuracy boost over individual machine learning models. Future work will involve integrating Sentinel-3 thermal data for yield prediction.

References

[1] J. Sens., "Remote Sensing Applications," 2017.
[2] A. Joshi et al., "Deep-learning techniques in crop mapping," Remote Sensing, vol. 2014, 2023.
[3] Y. Zhang et al., "Ensemble learning for wheat yield," Prog. Phys. Geogr., 2022.