# Random Forest Regression Approach For Crop Yield Estimationusing SENTINEL-1 SAR, LULC And Climate Data

**Prof. G. Durga, M.E.[1], Keerthana M[2], Dharani R[3], Kesavan V[4], Renuka R[5]**
[1, 2, 3, 4, 5] Dept of Agricultural Engineering
[1, 2, 3, 4, 5] M.A.M College of Engineering, Trichy, Tamil Nadu, India

***Abstract-*** *Crop yield estimation plays a major role in agricultural planning and food security. Traditional methods of crop yield estimation are time-consuming and require extensive field surveys. This study focuses on estimating rice crop yield using Sentinel-1 Synthetic Aperture Radar (SAR) data integrated with Land Use/Land Cover (LULC) and climate parameters in Orathanadu Taluk, Thanjavur District. Sentinel-1 SAR data provides VV and VH polarization backscatter values that help analyze crop growth and moisture conditions. The collected SAR data was preprocessed using radiometric calibration, speckle filtering, terrain correction, and decibel conversion. Climate parameters such as rainfall, temperature, and humidity were integrated with SAR features. A Random Forest Regression model was developed using Google Colab to predict crop yield. The model performance was evaluated using $R^2$ and RMSE values. The obtained results showed high prediction accuracy with an $R^2$ value of 0.9502 and RMSE value of 11.24 kg/ha. Spatial crop yield mapping was performed using GIS techniques to identify high and low yield zones. The study demonstrates that integrating remote sensing data with machine learning provides an efficient and reliable method for crop yield estimation.*

***Keywords:*** Sentinel-1 SAR, Random Forest Regression, Crop Yield Estimation, Remote Sensing, LULC, Climate Data

## I. INTRODUCTION

Agriculture is one of the most important sectors contributing to economic growth and food security, especially in developing countries like India. A large percentage of the population depends directly or indirectly on agriculture for their livelihood. Accurate crop yield estimation is essential for effective agricultural planning, market forecasting, food supply management, and decisionmaking by farmers and government agencies. Traditional crop yield estimation methods mainly depend on field surveys and manual observations, which are time-consuming, labor-intensive, and often limited to small geographic areas.

Recent advancements in remote sensing technologies have provided efficient alternatives for agricultural monitoring. Satellite imagery enables continuous observation of crop growth over large areas with reduced cost and time. Optical satellite data has been widely used for crop monitoring; however, cloud cover during cropping seasons often limits data availability. To overcome this limitation, Synthetic Aperture Radar (SAR) data has gained significant importance because it can capture images under all weather conditions, including cloudy environments and nighttime conditions.

European Space Agency developed Sentinel1 under the Copernicus programmed to provide C-band SAR data for Earth Observation applications. Sentinel-1 offers dual polarization data such as VV and VH backscatter coefficients, which provide valuable information related to crop structure, biomass, and moisture conditions. These backscatter values vary according to crop growth stages and can be effectively used for crop monitoring and yield prediction. Machine learning techniques have become increasingly popular in agricultural studies due to their ability to process large datasets and model complex nonlinear relationships. Among various algorithms, Random Forest Regression is widely used because of its high prediction accuracy, robustness, and ability to handle multiple input variables. It combines multiple decision trees to generate reliable predictions while reducing overfitting issues.

In this study, Sentinel-1 SAR data, Land Use/Land Cover (LULC) information, and climatic parameters such as rainfall, temperature, and humidity are integrated to estimate rice crop yield in Orathanadu Taluk, Thanjavur District, Tamil Nadu, India. The study aims to develop an efficient crop yield prediction model using Random Forest Regression and generate a spatial yield map for better agricultural planning and resource management.

## II. LITERATURE REVIEW

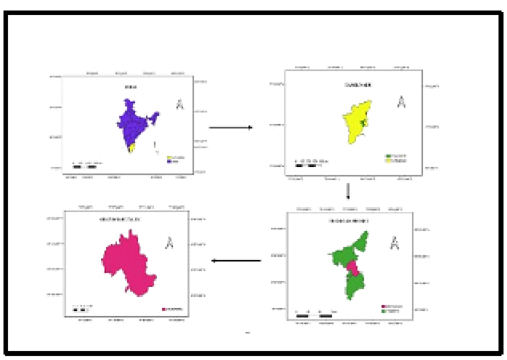
Several researchers have used remote sensing and machine learning for crop yield estimation. Previous studies reported that Sentinel-1 VV and VH backscatter values are highly related to crop biomass and yield prediction. Random Forest Regression has shown better performance compared to traditional statistical models. Integration of SAR data with climatic parameters improves prediction efficiency.

## III. MATERIALS AND METHODS

### Study Area

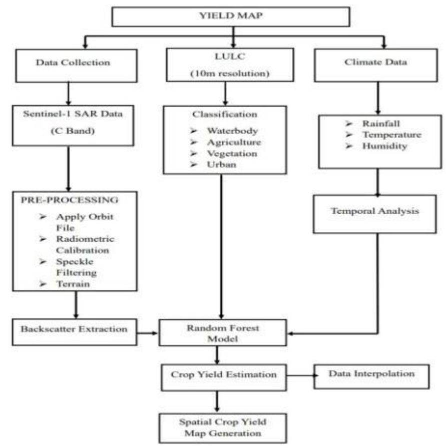
The study was carried out in Orathanadu Taluk, Thanjavur District, Tamil Nadu, India, located in the fertile Cauvery delta region. The study area lies between 10.60° N to 10.75° N latitude and 79.10° E to 79.30° E longitude. The region receives an average annual rainfall of 900–1100 mm, mainly during the Northeast monsoon season.

Agriculture is the major occupation in this area, and rice is the dominant crop cultivated due to fertile alluvial soil and irrigation facilities from the Cauvery River canal system. Due to its intensive agricultural activities, Orathanadu Taluk was selected for crop yield estimation using Sentinel-1 SAR data and machine learning techniques.



### Data Used

- Sentinel-1 SAR data
- LULC data
- Climate data
- Ground truth yield data

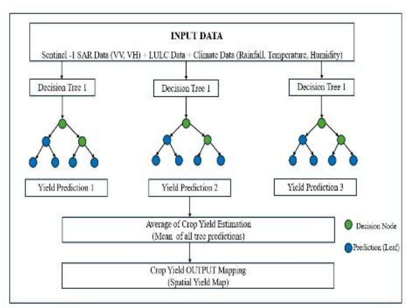



### SAR Preprocessing

- Apply orbit file
- Thermal noise removal
- Radiometric calibration
- Speckle filtering
- Terrain correction

### Random Forest Model

The Random Forest Regression model was used to predict crop yield using VV, VH, rainfall, temperature, and humidity data.




$Y=\frac{1}{N}\sum T_i(x)$ Where:
Y = Predicted yield
N = Number of trees
$T_i(x)$ = Individual tree prediction

## IV. RESULTS AND DISCUSSION

The collected Sentinel-1 SAR data was successfully preprocessed using SNAP through orbit correction, thermal noise removal, radiometric calibration, speckle filtering, terrain correction, and dB conversion. The processed data provided reliable VV and VH backscatter values for crop analysis.

Land Use/Land Cover (LULC) classification was carried out using QGIS to identify agricultural land, water bodies, built-up areas, vegetation, and other land classes. Agricultural regions were extracted for further yield analysis.

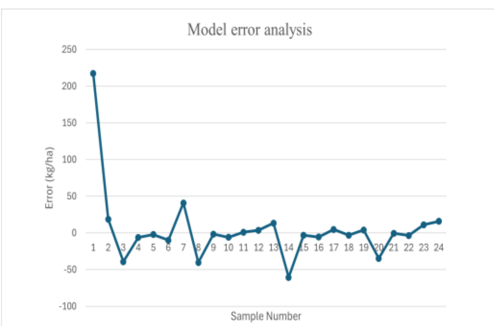
Climate parameters such as rainfall, temperature, and humidity were integrated with SAR backscatter values to improve prediction accuracy. These datasets were used as input variables for the Random Forest Regression model.

The model showed strong performance in predicting crop yield. The accuracy assessment resulted in an $R^2$ value of 0.9502 and RMSE value of 11.24 kg/ha, indicating high prediction accuracy and low error.

$$R^2 = 1-\frac{\sum (y_i-\hat{y_i})^2}{\sum (y_i-\bar{y})^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum (y_i-\hat{y_i})^2}$$

The spatial yield map generated using QGIS identified high, medium, and low-yield zones in Orathanadu Taluk, Thanjavur District, Tamil Nadu, India, which can help farmers improve crop management practices.




Field yield Vs Predicted yield

**Accuracy Assessment**

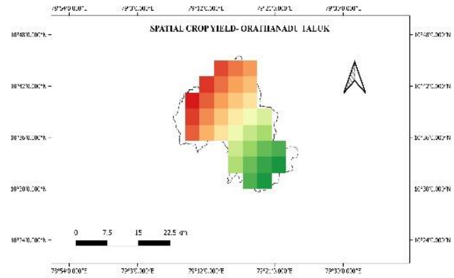
$$R^2 = 1-\frac{\sum (y_i-\hat{y_i})^2}{\sum (y_i-\bar{y})^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum (y_i-\hat{y_i})^2}$$

**Results:**

- $R^2 = 0.9502$
- RMSE = 11.24 kg/ha

The spatial crop yield map identified high, medium, and low yield regions in Orathanadu Taluk.




## V. CONCLUSION

This study focused on estimating rice crop yield using Sentinel-1 SAR data integrated with Land Use/Land Cover (LULC) and climatic parameters in Orathanadu Taluk, Thanjavur District, Tamil Nadu, India. The use of SAR data provided an effective solution for crop monitoring under all weather conditions, especially during cloudy seasons where optical satellite data becomes limited. The preprocessing techniques improved the quality of SAR images and enabled accurate extraction of VV and VH backscatter values.

The integration of rainfall, temperature, humidity, and ground truth yield data improved the performance of the Random Forest Regression model. The model achieved a high prediction accuracy with an $R^2$ value of 0.9502 and an RMSE value of 11.24 kg/ha, indicating that the developed model can effectively estimate crop yield with minimum error.

The spatial crop yield map generated in QGIS helped identify high, medium, and low-yield zones within the study area. This information can support farmers, agricultural planners, and policymakers in improving crop management, irrigation planning, and resource allocation. Overall, this study proves that integrating remote sensing and machine learning techniques offers a reliable and cost-effective approach for crop yield estimation.

## VI. FUTURE SCOPE

Future research can improve crop yield estimation by integrating multi-temporal satellite datasets such as Sentinel-2 optical imagery along with Sentinel-1 SAR data for better crop monitoring. Additional parameters such as soil moisture, soil nutrients, fertilizer usage, and crop health indices can be included to enhance model performance.

Advanced machine learning algorithms such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), and deep learning models can be applied to compare prediction accuracy with Random Forest Regression. The methodology can also be extended to different crops and larger geographical regions for regional-level yield forecasting.

Integration of real-time weather data, IoT sensors, and drone-based monitoring systems can further support precision agriculture practices. In future, this model can be developed as a decision support system to help farmers and agricultural departments make better decisions related to irrigation planning, fertilizer management, and crop productivity improvement.

## REFERENCES

[1] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[2] M. Drusch "Sentinel missions for ESA's Copernicus program," Remote Sensing of Environment, vol. 120, pp. 25–36, 2012.
[3] C. Atzberger, "Advances in remote sensing of agriculture," Remote Sensing, vol. 5, no. 2, pp. 949–981, 2013.
[4] D. Lobell "Remote sensing of regional crop production," Proceedings of the National Academy of Sciences, vol. 100, no. 18, pp. 10323–10328, 2015.
[5] J. Jeong "Random forests for global and regional crop yield predictions," PLOS ONE, vol. 11, no. 6, 2016.
[6] A. Veloso "Understanding crop phenology using Sentinel-1 and Sentinel-2 data," Remote Sensing, vol. 9, no. 6, 2017.
[7] K. Liakos et al., "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, 2018.
[8] L. Zhang et al., "Crop yield prediction using remote sensing and machine learning techniques," Remote Sensing, vol. 11, no. 3, 2019.
[9] E. Kamir"Crop yield prediction using satellite data and machine learning models," Agricultural Systems, vol. 173, 2020.
[10] Singha and Swain, "Sentinel-1 SAR based crop biomass estimation using machine learning," Remote Sensing Applications, 2023.