AgriFieldNet: A Multi-Source Remote Sensing and Machine Learning Framework for Soil Moisture Estimation and Irrigation Advisory

Yash Dhore[1] Prathamesh Nibandhe[2] Aashi Jaithly[3] Jayashri Phatangare[4] Mr. Shripad Bhide[5]
[1,2,3,4]Student [5]Guide
[1,2,3,4,5]Master of Computer Applications (MCA)
[1,2,3,4,5]Progressive Education Society's Modern College of Engineering, Pune, India

Abstract — Agricultural irrigation decisions depend strongly on soil water availability, yet conventional field monitoring is spatially sparse and difficult to scale. This paper presents AgriFieldNet, a web-based framework that estimates plot-level soil moisture from multi-source remote sensing and environmental data and converts the estimate into a water deficit class and an irrigation advisory. The system fuses Sentinel-1 radar backscatter, Sentinel-2 optical bands and vegetation indices, ERA5-Land meteorological variables, and MODIS leaf area index in a single Random Forest regression model with eighteen features. A historical pipeline operates on a prepared Berambadi dataset, while a live pipeline retrieves the latest suitable observations through Google Earth Engine. Evaluated with a plot-wise train and test split, the radar-only model gave a negative coefficient of determination, the radar plus optical model reached 0.34, and the full fusion model reached 0.53 with a mean absolute error of 6.78 and a root mean square error of 8.45. Humidity was the most influential feature. A successful live run confirmed end-to-end operation. The results are prototype-scale and require field validation before operational use.

Keywords: Soil Moisture, Sentinel-1, Sentinel-2, Random Forest, Google Earth Engine, ERA5-Land, Sensor Fusion, Irrigation Advisory

I. Introduction

Soil moisture is a key state variable for agricultural water management because it links rainfall, irrigation, evapotranspiration, vegetation condition, and soil-water availability. Field sensors give accurate local readings, but a small number of sampling points rarely represents the spatial variability of a cultivated landscape. Remote sensing offers a complementary way to observe field conditions over wider areas. Sentinel-1 provides C-band synthetic aperture radar (SAR) observations that are independent of daylight and comparatively less affected by cloud cover [1], while Sentinel-2 supplies multispectral information useful for vegetation and surface characterization [2].

This paper presents AgriFieldNet, a research-oriented web application that combines radar, optical, vegetation, and meteorological information in a machine-learning pipeline to estimate soil moisture and translate it into a water-deficit category and advisory. The application has two modes. *Historical Monitoring* operates on a prepared research dataset and supports analysis and demonstration. *Live Satellite Monitoring* uses plot coordinates to retrieve the latest suitable observations through Google Earth Engine (GEE), a cloud-based geospatial data catalog and computation platform accessible through Python and JavaScript APIs [3], [4].

The contributions of this work are: (1) a comparison of radar-only, radar-plus-optical, and, radar-plus-optical-plus-environment feature configurations under a plot-wise train/test split; (2) a live pipeline that aligns Sentinel-1, Sentinel-2, ERA5-Land, and MODIS observations through GEE; (3) conversion of the continuous estimate into a water-deficit class and advisory; and (4) a reproducible web implementation that preserves the trained model and its feature metadata. The contribution is an integrated engineering and research workflow rather than a new physical soil-moisture retrieval equation.

The remainder of the paper is organized as follows. Section II reviews related work, Section III states the research gap and problem, Section IV describes the methodology, Section V presents the system implementation, Section VI reports results and testing, and Sections VII and VIII discuss limitations and conclude.

II. Related Work

Research on soil-moisture estimation has moved from single-source retrieval toward multi-source remote sensing combined with machine learning. Breiman's Random Forest (RF) formulation [5] established an ensemble decision-tree method that covers regression and provides feature-importance measures. ERA5-Land [6] supplies a global hourly land reanalysis that can serve as environmental context. Zhou et al. [7] showed that spatial-temporal constraints and multi-polarization data can improve agricultural surface soil-moisture mapping from Sentinel-1 and Sentinel-2. Machine-learning estimation of surface soil moisture from Sentinel-1 SAR imagery has been demonstrated in [8], and synergistic use of Sentinel-1/2 data with meteorological variables, including RF feature-importance analysis, has been reported in [9]. A study of a heterogeneous tropical landscape combined active and passive Earth observations with RF [10], which is methodologically close to the present work. MODIS leaf area index (LAI) [11] provides additional vegetation context.

Three observations follow from this literature. First, SAR backscatter is complementary to optical vegetation information. Second, combining sources can improve accuracy when features are temporally and spatially compatible. Third, operational systems need more than a predictive model: they require data acquisition, quality control, temporal alignment, interpretation, and a user-facing interface.

III. Research Gap and Problem Statement

A. Research Gap

A practical gap remains between experimental models and small, usable field-monitoring tools. Many studies address a single dataset or region, or a sophisticated mapping method, without covering the whole chain from acquisition to prediction, deficit interpretation, and a simple interface. Specifically: radar-only approaches may miss vegetation and

All rights reserved by www.ijsrd.com

environmental context; Sentinel-1 and Sentinel-2 acquisitions occur on different dates and need explicit alignment; a trained model alone is not a monitoring workflow; a soil-moisture number is not actionable without a reference and deficit interpretation; and reproducibility requires saved feature order, metadata, and a repeatable live-acquisition procedure.

B. Problem Statement

Design and implement a web-based decision-support system that estimates plot-level soil moisture from historical and latest-available Earth-observation and environmental data, compares the estimate against a data-derived reference moisture level, and produces an interpretable water-deficit status and irrigation advisory. The inputs are the plot identifier and coordinates, historical observations, Sentinel-1 VV/VH backscatter, Sentinel-2 bands and derived indices, ERA5-Land variables, and MODIS LAI. The outputs are the estimated soil moisture (project target mv), the water-deficit percentage, the deficit class, the advisory, and the acquisition dates of the live data sources.

IV. Proposed Methodology

A. Overall Methodology

The methodology comprises data acquisition, preprocessing, feature engineering, model training, validation, live inference, and decision support, as shown in Fig. 1.

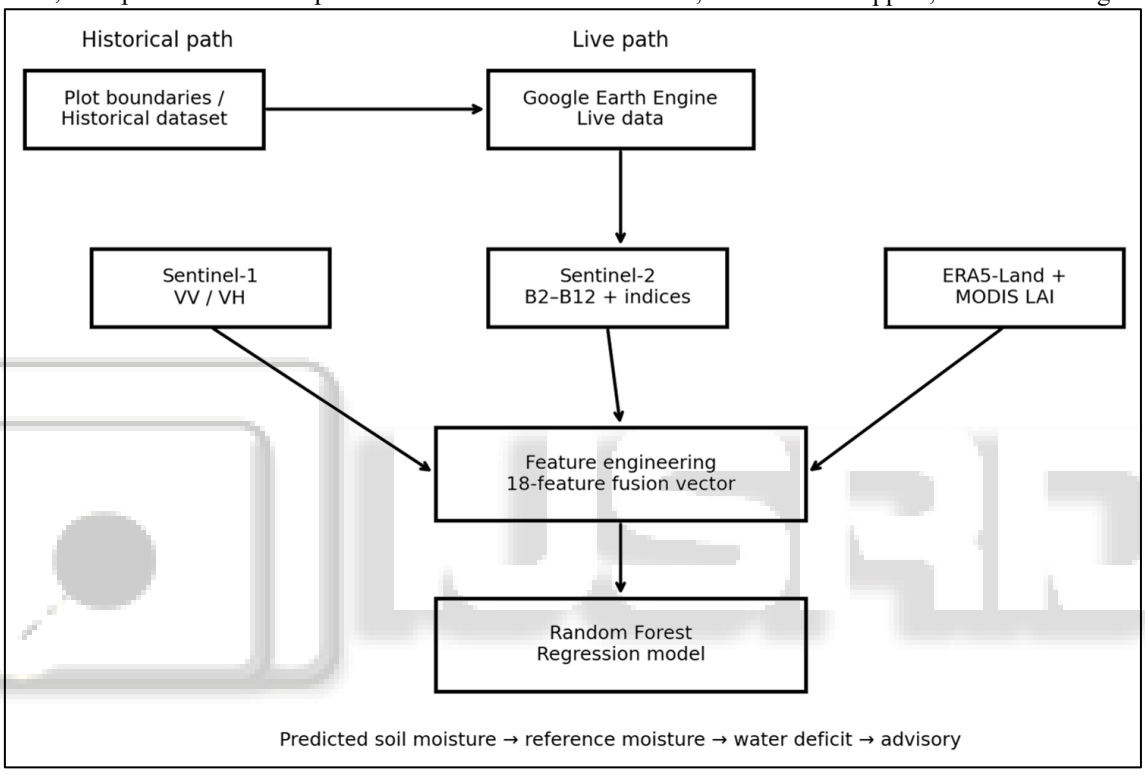


Fig. 1. AgriFieldNet system architecture.

B. Historical Data Pipeline

The prepared Berambadi observations are loaded and filtered to retain records with the required target and radar features, after which the project's optical preprocessing rules are applied. The data are split by plot rather than by individual observation. RF regression models are trained with increasing feature sets, compared using MAE, RMSE, and R^2, and the selected fusion model is saved together with its metadata for use in the historical dashboard.

C. Live Data Pipeline

For a selected plot, the pipeline queries recent Sentinel-1 observations and extracts VV and VH. It then searches Sentinel-2 within a ±7-day window around the radar acquisition and extracts bands B2, B3, B4, B8, B11, and B12. The latest suitable ERA5-Land hourly data are aggregated into the required environmental features, and the latest suitable MODIS LAI observation is retrieved. The values are assembled into the same 18-feature order used in training, the saved fusion model predicts soil moisture, and the water deficit and advisory are computed. The result is returned to the dashboard with the acquisition metadata. If an essential feature is unavailable, no prediction is produced.

D. Feature Engineering

Three spectral indices are computed from Sentinel-2 bands:

$$NDVI = (B8 - B4) / (B8 + B4) \quad (4.1)$$

$$GNDVI = (B8 - B3) / (B8 + B3) \quad (4.2)$$

$$NDWI = (B3 - B8) / (B3 + B8) \quad (4.3)$$

The feature vector contains 18 values in fixed order: VV, VH, B2, B3, B4, B8, B11, B12, NDVI, GNDVI, NDWI, LAI, air temperature, atmospheric pressure, global radiation, precipitation, humidity, and wind direction.

E. Model and Validation

RF was selected because it models nonlinear relationships, handles heterogeneous numeric features, needs little feature scaling, and yields feature-importance measures [5]. A plot-wise split keeps observations from the same plot out of both training and testing sets; because neighboring observations are correlated, a random split would make the model appear more accurate than it is. Performance is measured by

All rights reserved by www.ijsrd.com

$$MAE = (1/n)\ \Sigma\ |y_i - \hat{y}_i| \quad (4.4)$$

$$RMSE = \mathrm{sqrt}[(1/n)\ \Sigma\ (y_i - \hat{y}_i)^2] \quad (4.5)$$

$$R^2 = 1 - \Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2 \quad (4.6)$$

where y_i is the observed value, $\hat{y}_i$ the prediction, $\bar{y}$ the mean of observations, and n the number of test samples. MAE is the average absolute error, RMSE penalizes large errors more strongly, and R^2 is the fraction of variance explained relative to a mean-prediction baseline.

F. Water-Deficit Interpretation

The project uses a data-derived reference moisture level M_{ref}. The water deficit is

$$WD = \max[0, (M_{ref} - M_{pred}) / M_{ref} \times 100] \quad (4.7)$$

where M_{pred} is the predicted moisture and WD is expressed in percent. Table I gives the prototype thresholds. These are decision-support rules, not a crop-specific irrigation-dose model, and they do not specify irrigation volume or duration.

Deficit	Status	Interpretation
< 10%	Adequate	Close to or above reference
10–25%	Mild	Monitor; irrigation may be considered
25–40%	Moderate	Elevated stress risk; consider irrigation
≥ 40%	Severe	High risk; verify in field and plan irrigation

Table I: Water-Deficit Classification

V. System Implementation

A. Processing Flow

Fig. 2 shows the end-to-end flow. In historical mode the user selects a plot, stored features are read from the dataset, and the historical RF produces the estimate. In live mode the plot coordinates are sent to GEE, which supplies Sentinel-1, Sentinel-2, ERA5-Land, and MODIS data. Both modes then compute the reference comparison, water deficit, advisory, and dashboard output.

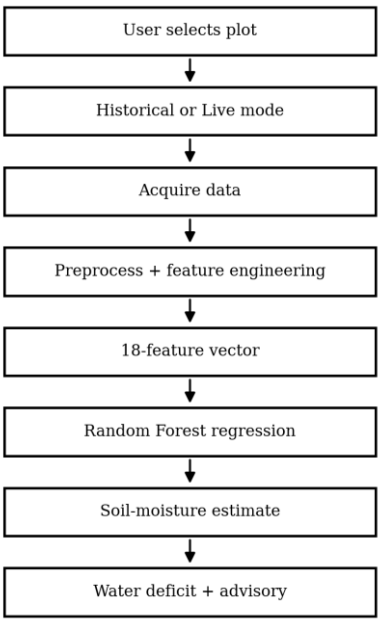


Fig. 2: End-to-end AgriFieldNet processing flowchart.

B. Components

The system is organized in layers. The presentation layer uses HTML, CSS, and JavaScript for plot and mode selection and result display. The application layer is a Flask/Python backend that handles routing, validation, orchestration, and response generation. The data layer combines a historical CSV with live Earth Engine access. The ML layer uses scikit-learn RF regression, and the persistence layer stores the model with joblib and the feature names and order in a JSON metadata file. A Python rule module handles reference comparison, deficit classification, and advisory text. The remaining tooling comprises pandas, NumPy, earthengine-api, Matplotlib, and Git, running under Python 3.12.

Earth Engine credentials are kept on the backend and never embedded in frontend code. The RF implementation needs no GPU; a dual- or quad-core CPU with 8 GB of RAM is sufficient, and an internet connection is required only for live prediction.

VI. Results and Discussion

A. Dataset

The Sentinel-1 extraction stage reported 707 observations, of which 664 had valid VV/VH values and the target. Observed VV ranged from approximately −16.81 to −2.44 dB and VH from −26.57 to −9.07 dB. Table II summarizes the split.

Metric	Value
Original observations	707
Usable observations (VV/VH and target)	664
Training observations	525
Testing observations	139
Training plots	52
Testing (held-out) plots	14

Table II: Dataset Summary

B. Model Comparison

Table III and Fig. 3 compare the three feature configurations on the held-out plots.

Model	Features	MAE	RMSE	R^2
VV/VH only	2	10.3993	13.2626	−0.1451
VV/VH + optical	11	8.0922	10.0948	0.3366
VV/VH + optical + environment	18	6.7792	8.4532	0.5348

Table III: Model Comparison

All rights reserved by www.ijsrd.com

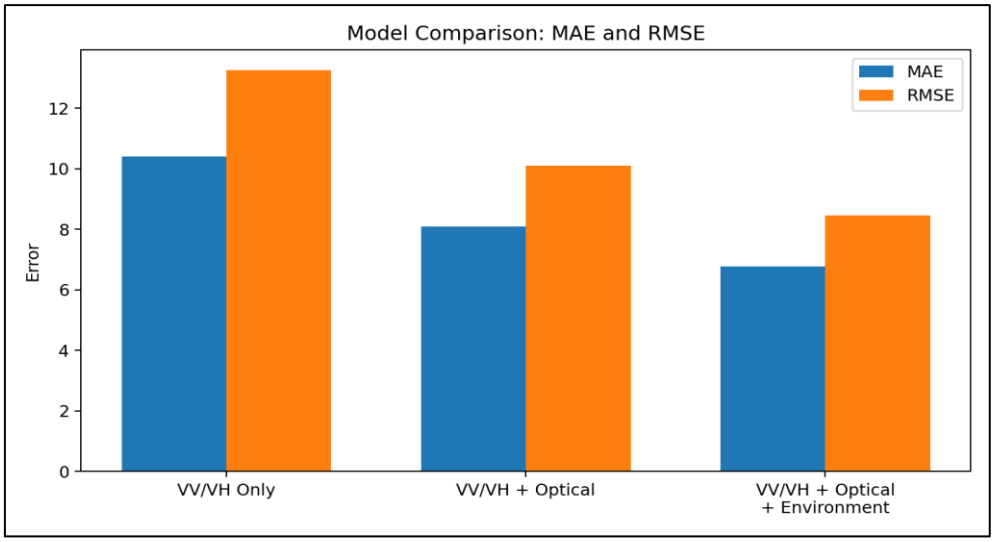


Fig. 3: Comparison of MAE and RMSE across model configurations.

Performance improves as complementary information is added. The radar-only model produced a negative R^2 on held-out plots, meaning it performed worse than a mean-based baseline under this evaluation. Adding optical features raised R^2 to 0.3366, and adding environmental and vegetation-context features raised it to 0.5348 while reducing MAE and RMSE.

C. *Feature Importance*

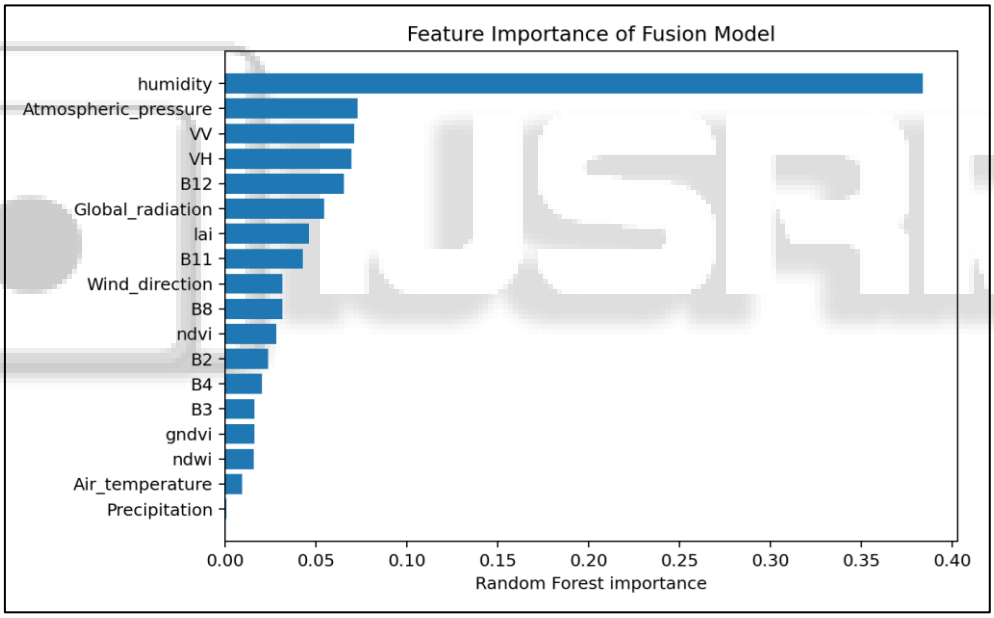


Fig. 4: Random Forest feature importance for the fusion model.

The built-in RF importance ranked humidity highest (0.3839), followed by atmospheric pressure (0.0729), VV (0.0712), VH (0.0696), B12 (0.0653), global radiation (0.0544), LAI (0.0461), and B11 (0.0430). Permutation importance on the held-out data also identified humidity as the strongest contributor. Some features had near-zero or negative permutation importance, indicating that they may be redundant, weak, or unstable for this dataset and split.

D. *Live Pipeline Demonstration*

A live run for Plot 15 retrieved all four data sources and returned an estimate, as summarized in Table IV. The Sentinel-1 and Sentinel-2 acquisitions differ by approximately 5.8 days, within the ±7-day window.

Component	Observation
Plot / location	Plot 15; 11.78101, 76.62736
Sentinel-1	15 Sep 2026; VV −5.6556, VH −15.1687
Sentinel-2	09 Sep 2026 (within ±7 days)
ERA5-Land	10 Sep 2026 22:00; 20.5 °C, 90.94 %, 900.36 hPa
Precipitation / radiation	0.082 mm; 11.295 MJ/m² (aggregated)
Wind direction	242.85°
MODIS LAI	1.2 on 06 Sep 2026
Live prediction	18.63 (soil-moisture target units)

Table IV: Live Pipeline Demonstration

The run confirms that the pipeline can authenticate with Earth Engine, retrieve the required sources, construct the

All rights reserved by www.ijsrd.com

fusion feature vector, load the saved model, and return an estimate.

E. System Testing

Ten functional test cases covered application launch, plot selection, historical prediction, live prediction, Sentinel-1 retrieval, Sentinel-2 alignment, ERA5-Land retrieval, LAI retrieval, missing-data handling, and model loading. All passed. In the missing-data case the system reports a failure rather than producing an invalid prediction, as intended. Model loading passed after an environment correction: on one Windows machine, Python 3.14 with scikit-learn triggered an Application Control DLL-blocking error, so the project was moved to a dedicated Python 3.12 virtual environment with the scikit-learn version matched to the training environment. An SSL certificate-chain issue during Earth Engine authentication was resolved by using the system certificate store through pip-system-certs. Non-functional checks confirmed a simple interaction path, backend-side credentials, and clear separation of acquisition, model, and interface roles; live response time depends on Earth Engine and network conditions.

VII. Limitations and Future Work

A. Limitations

The evaluation uses a finite Berambadi dataset and may not generalize to other soils, crops, climates, or management practices. The fusion model explains a moderate share of variance ($R^2 = 0.5348$), so errors remain significant, and no calibrated uncertainty interval is given per prediction. Sentinel-1 and Sentinel-2 acquisitions are not simultaneous, and ERA5-Land and MODIS LAI have coarser spatial resolution and different timing from the Sentinel data. The water-deficit reference is data-derived rather than based on crop evapotranspiration or a soil-water balance, and the advisory does not compute irrigation volume, duration, or application efficiency. Live monitoring also depends on cloud, network, and API availability.

B. Future Work

Planned extensions include field-sensor ground truth, more seasons and plots, crop-specific models with growth-stage features, uncertainty estimation, and comparison with XGBoost, CatBoost, and temporal models such as LSTM or Temporal Fusion Transformer. Additional static predictors (soil texture, elevation, slope, clay content) and a true irrigation-volume model based on evapotranspiration, root-zone water balance, and crop coefficients would improve decision value. Map-based water-stress visualization, secure cloud deployment, scheduled data refresh with alerts, and validation in independent geographic regions would support operational use.

VIII. Conclusion

AgriFieldNet demonstrates an end-to-end approach to soil-moisture monitoring that combines remote sensing, environmental information, machine learning, and a web-based decision-support interface, progressing from historical model development to a working live pipeline on Google Earth Engine. Feature fusion improved performance: R^2 rose from −0.1451 for radar only to 0.3366 with optical features and 0.5348 with environmental and vegetation context, with MAE falling from 10.3993 to 6.7792. The work also shows that a useful ML system needs data acquisition, temporal alignment, feature consistency, model persistence, validation, error handling, and a usable interface in addition to a model. These are prototype results for the available dataset and split; field validation, broader seasons and plots, crop-specific calibration, and independent geographic validation are required before agricultural decisions rely on the system.

Acknowledgment

The authors thank their project guide, Prof. Shripad Bhide, the faculty of the MCA Department, and P.E.S.'s Modern College of Engineering, Pune, for their guidance and support. The authors also acknowledge the open data ecosystems that made this work possible, including the Copernicus Sentinel missions, Google Earth Engine, ERA5-Land, and MODIS products.

References

[1] European Space Agency, "Sentinel-1: Radar vision for Copernicus," ESA Earth Observation mission information.
[2] European Space Agency, "Sentinel-2: Colour vision for Copernicus," ESA Earth Observation mission information.
[3] Google for Developers, "Google Earth Engine," Earth Engine documentation and Data Catalog.
[4] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," Remote Sens. Environ., vol. 202, pp. 18–27, 2017.
[5] L. Breiman, "Random forests," Mach. Learn., vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[6] J. Muñoz-Sabater et al., "ERA5-Land: A state-of-the-art global reanalysis dataset for land applications," Earth Syst. Sci. Data, vol. 13, pp. 4349–4383, 2021, doi: 10.5194/essd-13-4349-2021.
[7] Y. Zhou et al., "Spatial-temporal constraints for surface soil moisture mapping using Sentinel-1 and Sentinel-2 data over agricultural regions," Comput. Electron. Agric., vol. 219, Art. no. 108835, 2024, doi: 10.1016/j.compag.2024.108835.
[8] "Estimation of surface soil moisture from Sentinel-1 synthetic aperture radar imagery using machine learning method," Remote Sens. Appl.: Soc. Environ., vol. 36, Art. no. 101369, 2024, doi: 10.1016/j.rsase.2024.101369.
[9] "Machine learning-based inversion and sensitivity analysis of soil moisture in Hemerocallis cultivation," Agric. Water Manag., vol. 322, Art. no. 110011, 2025, doi: 10.1016/j.agwat.2025.110011.
[10] "Soil moisture dynamics in heterogeneous tropical landscapes using active–passive Earth observations and machine learning," Soil Adv., vol. 4, Art. no. 100077, 2025, doi: 10.1016/j.soilad.2025.100077.
[11] NASA/USGS LP DAAC, "MODIS LAI/FPAR product user's guide," MODIS LAI/FPAR documentation.

All rights reserved by www.ijsrd.com