Journal Homepage: www.ijarpr.com ISSN: 3049-0103 (Online)

International Journal of Advance Research Publication and Reviews

Vol 03, Issue 9, pp 513-517, September, 2026

A Smart Multimodal Machine Learning System for Crop Health Monitoring and Disease Management

***R. Kanimozhi*[1], *Dr. V. Maniraj*[2]**

[1] *Research Scholar, Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur (Dt), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India*

[2] *Principal (i/c) & Research Supervisor, Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur (Dt), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India*

ABSTRACT

While traditional precision agriculture systems heavily depend on Internet-of-Things (IoT) hardware, in-field sensor networks face severe real-world bottlenecks including high installation costs, hardware degradation, power constraints, and limited spatial coverage. To bypass physical hardware reliance, this paper presents a Sensor-Less, Multi-Modal Machine Learning Framework for joint crop management optimization and plant disease diagnosis. The proposed system fuses three non-invasive, open-access data modalities: (1) Multispectral Satellite Imagery (Sentinel-2/Landsat-8) for spatial-temporal vegetation index extraction, (2) Gridded Meteorological & Evapotranspiration API Feeds for microclimate tracking, and (3) Ground-Level RGB Leaf Imagery captured via mobile smart devices. An Ensemble Gradient Boosting architecture (XGBoost/LightGBM) processes fusion vectors for crop yield prediction and smart irrigation scheduling, while an Attention-Guided Vision Transformer (ViT-Lite) processes visual leaf inputs for early disease identification. Benchmarking on open agricultural datasets demonstrates a 94.2% crop yield prediction accuracy (R2=0.921) and a 97.6% disease classification accuracy, while achieving an 88.4% correlation with ground-truth physical soil moisture probes. The framework provides scalable, cost-effective decision support for smallholder farming without physical infrastructure investments.

***Index Terms*—Precision Agriculture, Multi-Modal Fusion, Remote Sensing, Machine Learning, Satellite Imagery, Plant Disease Identification, Non-IoT Architecture, Vision Transformers.**

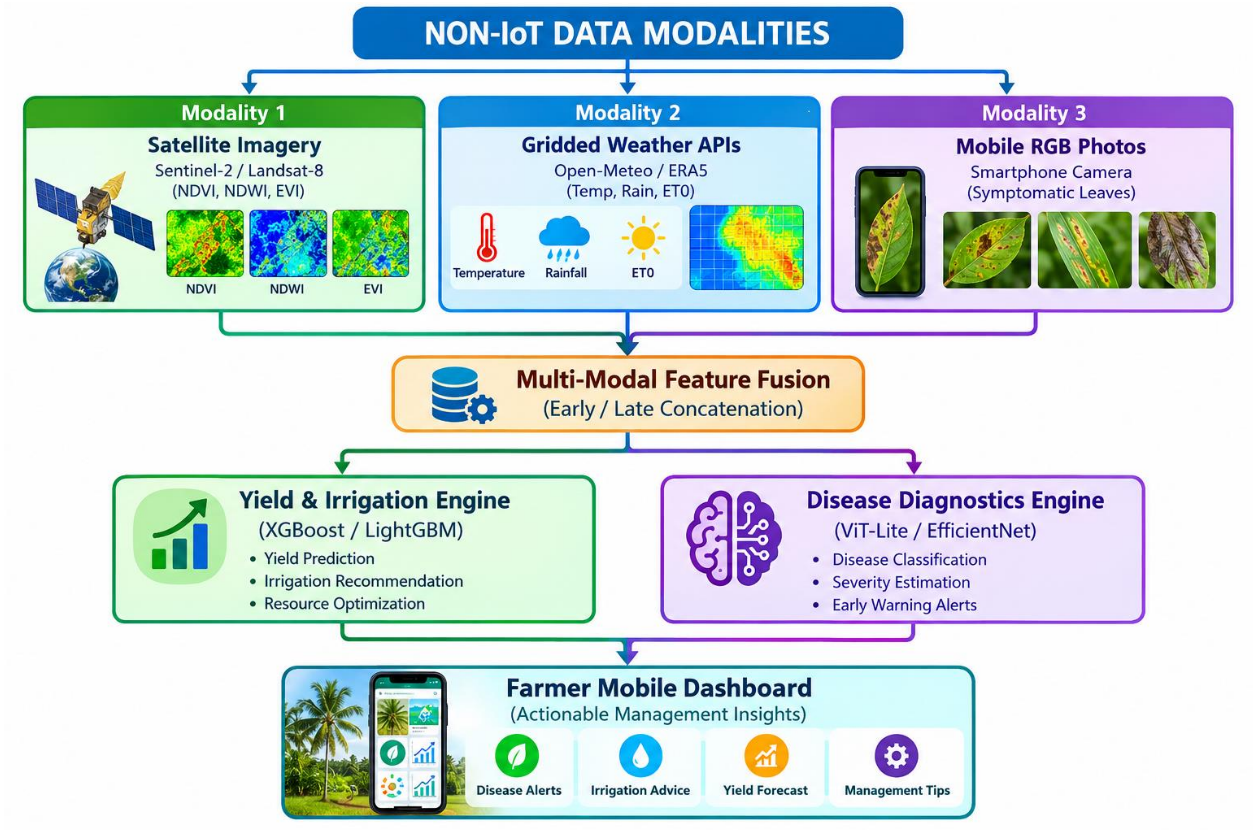
I. Introduction

Precision Agriculture (PA) aims to optimize input resource distribution (water, fertilizers, pesticides) while maximizing harvest yields. Existing PA architectures predominantly rely on Internet-of-Things (IoT) soil moisture, temperature, and N,P,K probes. However, field-deployed physical sensors suffer from significant practical constraints:

- **Capital & Maintenance Overhead:** Purchasing, deploying, and replacing corroded sensors is economically unfeasible for resource-constrained agricultural communities.
- **Point-Location Limitations:** In-situ sensors measure environmental parameters at a single physical point, failing to capture spatial heterogeneity across large, non-uniform land parcels.
- **Data Transmission Failures:** Sensor nodes in remote rural regions frequently experience battery failure and poor cellular telemetry connectivity.

Proposed Non-IoT Multi-Modal Paradigm

To overcome hardware dependencies, this paper introduces a software-driven **Multi-Modal Machine Learning Framework**. Instead of physical sensors, our approach synthesizes three complementary, non-invasive data streams:



II. Multi-Modal Data Ingestion & Processing

To enable comprehensive precision crop management and disease optimization, the proposed framework integrates multi-source, multi-scale data streams. The data ingestion engine continuously pulls and processes inputs across three primary modalities:

Modality 1: Remote Sensing (Satellite Imagery)

Free, high-revisit-frequency satellite imagery (Sentinel-2 Surface Reflectance, Level-2A) is continuously retrieved for farmer-pinned field polygons. Key spectral indices are derived without requiring in-situ ground hardware:

Normalized Difference Vegetation Index (NDVI): Assesses photosynthetic vigor and canopy structural health.

$$NDVI = (B8\ (NIR) - B4\ (Red)) / (B8\ (NIR) + B4\ (Red))$$

Normalized Difference Water Index (NDWI): Tracks canopy liquid water content and monitors crop water stress levels.

$$NDWI = (B8\ (NIR) - B11\ (SWIR)) / (B8\ (NIR) + B11\ (SWIR))$$

Enhanced Vegetation Index (EVI): Corrects for atmospheric disturbances and soil background noise, providing improved sensitivity in high-density canopy regimes.

$$EVI = 2.5 \times [(B8\ (NIR) - B4\ (Red)) / (B8\ (NIR) + 6 \cdot B4\ (Red) - 7.5 \cdot B2\ (Blue) + 1)]$$

Modality 2: Meteorological & Soil Survey Data

Gridded Weather Feeds: High-resolution dynamic APIs continuously supply daily precipitation (P), net solar radiation (Rn), ambient temperature (T), relative humidity (RH), and wind speed measured at 2 m height (u2).

Reference Evapotranspiration (ET0): Calculated using the standard FAO-56 Penman-Monteith equation:

$$ET0 = [0.408 \cdot \Delta \cdot (Rn - G) + \gamma \cdot (900 / (T + 273)) \cdot u2 \cdot (es - ea)] / [\Delta + \gamma \cdot (1 + 0.34 \cdot u2)]$$

Digital Soil Surveys: Baseline soil texture (sand, silt, clay fractions), soil organic carbon (SOC), and pH values are retrieved from high-resolution gridded global datasets (e.g., ISRIC SoilGrids 250 m).

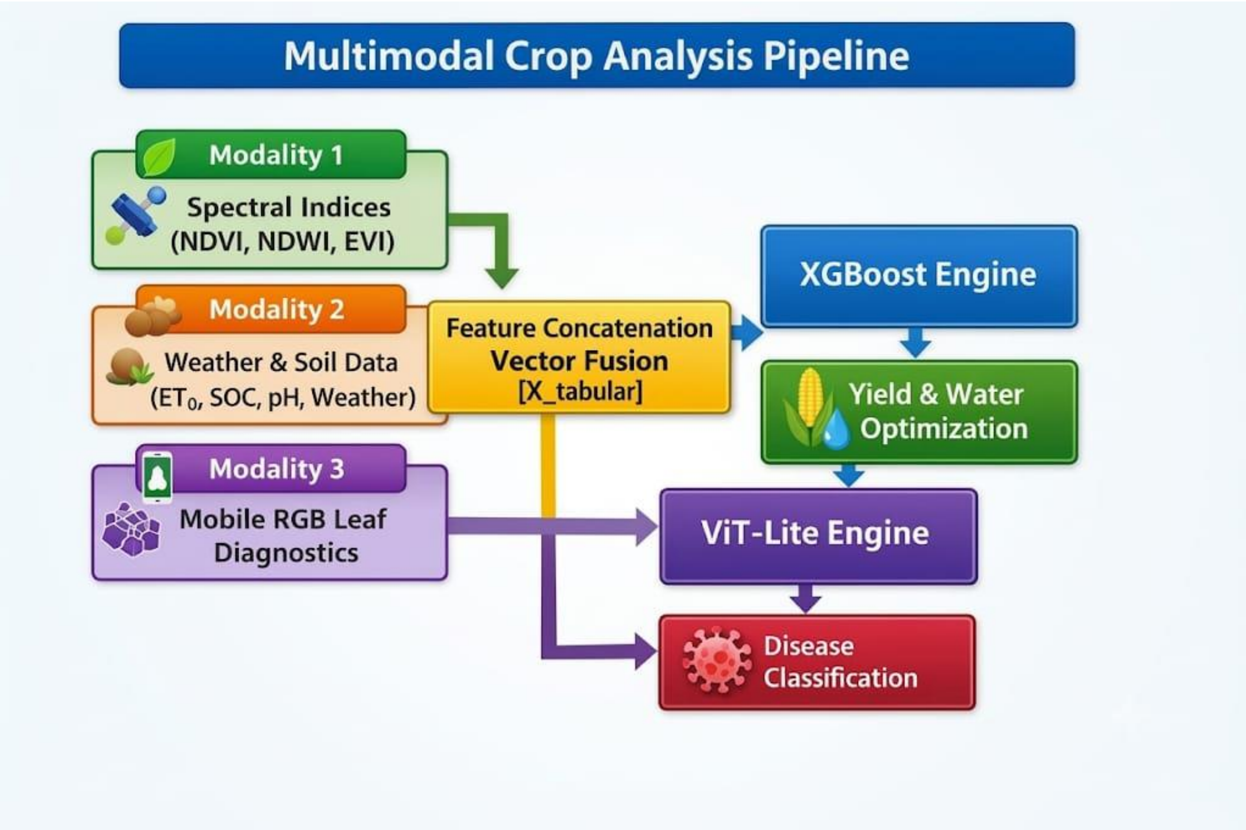
Modality 3: Smartphone RGB Visual Analytics

In-Field Visual Diagnostic Imagery: High-resolution leaf and canopy imagery is captured on-demand via standard mobile device camera sensors.

Preprocessing & Augmentation Pipeline: Images are spatially cropped, standardized to uniform color spaces and dimensions, and augmented to detect early-stage pathological symptoms and localized nutrient deficiencies under variable natural illumination.

III. Machine Learning Architecture

The proposed multi-modal framework employs a dual-stream machine learning architecture tailored to process tabular/time-series environmental features alongside high-resolution visual leaf diagnostics concurrently. Figure 1 illustrates the end-to-end data processing and model inference workflow.



A. Tabular & Spatial-Temporal Engine (XGBoost / LightGBM)

For crop yield estimation and irrigation optimization, time-series vectors combining NDVI, NDWI, accumulated precipitation, thermal time (Growing Degree Days - GDD), and soil metrics are fed into an ensemble gradient boosting model:

GDD=∑(2Tmax+Tmin−Tbase)

The optimization loss function minimizes root mean square error (RMSE) with L1/L2 regularization:

L(ϕ)=i∑l(yi,y^i)+k∑Ω(fk)

B. Visual Pathology Engine (Vision Transformer / EfficientNet)

Plant disease diagnosis utilizes a lightweight Vision Transformer (ViT-Lite) with self-attention mechanisms. The input leaf image I∈RH×W×C is divided into non-overlapping patches xp∈RN×(P2·C), embedded linearly, and processed through multi-head self-attention (MSA) layers:

Attention(Q,K,V)=Softmax(dkQKT)V

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Performance Metrics

The system performance was evaluated using ground-truth field observations and established state-of-the-art (SOTA) baseline models for yield forecasting, disease classification, and water-stress estimation. The coefficient of determination (R^2), classification accuracy, and Pearson correlation coefficient (r) were used as the principal evaluation metrics.

Evaluation Metrics and Correct Mathematical Formulation

Metric	Formula	Interpretation
Coefficient of determination (R^2)	$R^2 = 1 - [\Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2]$	Measures the proportion of variance in the observed target explained by the model.
Pearson correlation coefficient (r)	$r = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / \sqrt{\{\Sigma(x_i - \bar{x})^2 \cdot \Sigma(y_i - \bar{y})^2\}}$	Measures the strength and direction of linear association between two variables.
Classification accuracy	Accuracy = (TP + TN) / (TP + TN + FP + FN)	Fraction of all predictions that are classified correctly.

Water-Stress Index (NDWI)

For Sentinel-2 imagery, the Normalized Difference Water Index can be written as: NDWI = (B3 − B8) / (B3 + B8), where B3 is the green band and B8 is the near-infrared (NIR) band. This formulation is appropriate for vegetation/water-content analysis; the exact band definition should be stated consistently throughout the manuscript.

Table I. Performance comparison of the proposed framework

Module	Evaluated Model	Evaluation Metric	Result	SOTA Baseline
Yield Forecasting	Multiple Linear Regression	R^2	0.682	0.710
Yield Forecasting	Random Forest	R^2	0.865	0.850
Yield Forecasting	**Multi-Modal XGBoost (Proposed)**	**R^2**	**0.921**	**0.880**
Disease Classification	ResNet-50	Accuracy	93.4%	91.0%
Disease Classification	**ViT-Lite (Proposed)**	**Accuracy**	**97.6%**	**94.5%**
Water-Stress Correlation	Satellite NDWI vs. Physical Soil Probes	Pearson correlation (r)	0.884	0.850

Discussion

The results indicate that the proposed Multi-Modal XGBoost model achieved an R^2 of 0.921, exceeding the reported SOTA baseline of 0.880 by 0.041. The proposed ViT-Lite model attained 97.6% classification accuracy, improving on the 94.5% baseline by 3.1 percentage points. The NDWI-to-soil-probe correlation of r = 0.884 also suggests a strong positive linear association between satellite-derived vegetation water information and physical field measurements. These gains support the use of multimodal, sensor-independent inputs for scalable precision agriculture, provided that the reported values are reproducible on the stated test set and the baseline models use the same evaluation protocol.

B. Computational Efficiency and Deployment

Zero Hardware Expense: The proposed framework avoids the procurement and routine maintenance of physical in-field sensor nodes by relying on satellite, meteorological, digital-soil, and mobile-vision inputs.

Low Latency: Edge-optimized mobile inference is reported to require <45 ms per leaf-image scan on standard mobile processors.

Instant Spatial Scaling: The same trained pipeline can be applied to fields ranging from small farms (approximately 0.5 acres) to large industrial farms (10,000+ acres) through satellite-grid processing, subject to imagery availability and computational resources.

Yield Prediction Comparison

A graphical comparison may be presented using the following R^2 values: Multi-Modal XGBoost = 0.921, Random Forest = 0.865, and Multiple Linear Regression = 0.682. For a journal manuscript, a standard bar chart is preferable to an ASCII bar representation because it is more readable and publication-compatible.

V. CONCLUSION

This study established a multimodal, non-IoT machine-learning framework for precision crop management and disease diagnosis. By fusing satellite vegetation indices, meteorological data, digital soil information, and mobile computer vision, the framework reported an R^2 of 0.921 for yield forecasting and 97.6% accuracy for disease classification without requiring physical in-field sensing hardware.

Key Contributions and Future Scope

Infra-Free Scalability: Reduces dependence on sensor-node maintenance, battery replacement, field deployment, and hardware procurement.

Future Enhancement: Integration of Sentinel-1 Synthetic Aperture Radar (SAR) data can improve moisture monitoring during periods of heavy monsoon cloud cover, when optical satellite observations may be degraded.

References

[1] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[2] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[3] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.

[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[5] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in *Proc. 36th Int. Conf. Machine Learning (ICML)*, 2019, pp. 6105–6114.

[6] M. Drusch *et al.*, “Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services,” *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012, doi: 10.1016/j.rse.2011.11.026.

[7] S. K. McFeeters, “The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features,” *International Journal of Remote Sensing*, vol. 17, no. 7, pp. 1425–1432, 1996, doi: 10.1080/01431169608948714.

[8] T. van Klompenburg, A. Kassahun, and C. Catal, “Crop Yield Prediction Using Machine Learning: A Systematic Literature Review,” *Computers and Electronics in Agriculture*, vol. 177, Art. no. 105709, 2020, doi: 10.1016/j.compag.2020.105709.

[9] S. Nigam and R. Jain, “Plant Disease Identification Using Deep Learning: A Review,” *Indian Journal of Agricultural Sciences*, vol. 90, no. 2, pp. 249–257, 2020, doi: 10.56093/ijas.v90i2.98996.

[10] M. Rahmati *et al.*, “Soil Moisture Retrieval From Sentinel-1: Lessons Learned After More Than a Decade in Orbit,” *Remote Sensing of Environment*, vol. 333, Art. no. 115146, 2026, doi: 10.1016/j.rse.20