IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510158

AI-Driven Smart Agriculture Platform for Integrated Crop Health Monitoring and Yield Optimization: A Systematic Literature Review

Rakshith Prabhu[1], Srihari M.R[2], Srinivas Gowda C.N[3], Suhas C[4], Prof. Anitha L[5]

Department of Computer Science and Engineering, East West College of Engineering, Bengaluru, Karnataka, India[1-4]

Guide**,** Assistant Professor, Dept. of CSE, East West College of Engineering, Bengaluru, Karnataka, India[5]

Abstract: Agriculture is undergoing rapid transformation through the integration of Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), and Internet of Things (IoT) technologies. These advancements enable intelligent crop disease detection, crop yield prediction, and precision farming practices that improve agricultural productivity while minimizing resource consumption. Recent studies have demonstrated significant improvements in disease classification accuracy using Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and multimodal learning techniques. Similarly, advanced regression and ensemble learning algorithms have enhanced crop yield prediction by utilizing climatic, soil, and environmental parameters.

This systematic literature review analyzes recent research published between 2021 and 2026 on AI-driven smart agriculture systems. The review compares various machine learning algorithms, datasets, evaluation metrics, advantages, and limitations reported in the literature. Furthermore, the paper identifies current research gaps and discusses future research opportunities involving Explainable Artificial Intelligence (XAI), multimodal learning, federated learning, and real-time agricultural decision support systems. The findings indicate that integrating multiple AI techniques into a unified smart agriculture platform can significantly improve prediction accuracy, transparency, and practical usability for modern precision farming.

Keywords: Artificial Intelligence, Smart Agriculture, Crop Disease Detection, Crop Yield Prediction, Machine Learning, Deep Learning, Explainable AI, Precision Farming.

I. INTRODUCTION

Agriculture stands as a cornerstone of modern human civilization, playing a critical role in sustaining global food security, supporting rural livelihoods, and driving economic development across developing and industrialized nations alike. However, the global agricultural sector faces an unprecedented array of compounding challenges. The escalating impacts of climate change, characterized by increasingly erratic weather patterns, severe droughts, unseasonal rainfall, and thermal stress, severely disrupt traditional cropping cycles and lower overall crop yields. Furthermore, persistent issues such as accelerated soil degradation, nutrient depletion, rampant pest infestations, and complex plant pathogen outbreaks continuously diminish agricultural output, posing dire risks to global nutritional stability and supply chain resilience.

Historically, traditional agricultural paradigms have relied heavily on manual field observation, localized empirical experience, and qualitative expert judgment. While effective in predictable historical contexts, these manual approaches suffer from inherent human limitations: they are labor-intensive, subject to subjective interpretation, geographically constrained, and often unable to identify subtle sub-visual stress markers or early-stage pathological symptoms. Consequently, disease interventions are frequently delayed until damage is widespread, and yield forecasting remains highly inaccurate. In response to these operational bottlenecks, Artificial Intelligence (AI) has emerged as a transformative paradigm for modern smart farming. By operationalizing advanced machine learning workflows, AI enables continuous automated crop monitoring, precision pathology diagnosis, real-time microclimate monitoring, and data-driven yield estimation, thereby transforming reactive management into proactive, evidence-based agronomy. Modern AI architectures effectively synthesize multimodal data streams—ranging from multi-spectral satellite imagery and high-resolution aerial drone captures to real-time Internet of Things (IoT) soil parameters, localized meteorological

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IJARCCE.2026.1510158

readings, and longitudinal historical yield records—to provide highly localized, dynamic decision support directly to agricultural practitioners.

Recent empirical studies demonstrate significant breakthroughs in deploying specialized machine learning paradigms across diverse agronomic domains. In plant pathology, Deep Learning (DL) architectures—most notably Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs)—have achieved state-of-the-art diagnostic accuracy in visual disease classification and lesion segmentation. Simultaneously, in predictive crop analytics, non-linear ensemble learning algorithms such as Random Forest, Extreme Gradient Boosting (XGBoost), and Gradient Boosting Machines have demonstrated exceptional capability in modeling complex, high-dimensional interactions between pedological, climatic, and management variables to forecast continuous crop yields. Despite these predictive successes, the historical "black-box" nature of advanced deep networks and ensemble models has created a significant trust gap among agricultural stakeholders. To bridge this gap, Explainable Artificial Intelligence (XAI) frameworks—incorporating post-hoc attribution methods like Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME)—have been introduced. These interpretability tools expose the key bio-physical drivers behind model predictions, empowering farmers, extension agents, and agricultural policymakers to make transparent, verified decisions.

Despite these noteworthy technological advancements, a fundamental architectural limitation persists within the existing literature: the majority of current research focuses on isolated single-task models designed to solve individual problems independently—such as localized leaf disease classification or stand-alone seasonal yield forecasting—rather than developing integrated, end-to-end multi-task platforms. This fragmented approach hampers practical operational adoption, as real-world agricultural management demands simultaneous visual health diagnostics, dynamic environmental stress assessment, and holistic yield optimization within a unified workflow. Addressing this critical research gap, this systematic literature review presents a comprehensive synthesis of recent domain developments, critically compares existing algorithmic methodologies across standardized performance metrics, identifies core operational and technical gaps, and outlines strategic directions for future research in unified, interpretable smart agriculture platforms.

II. RELATED WORK

1. Impact of Climate Change on Agricultural Productivity

Ray et al. (2015) investigated the influence of climate variability on global crop production using large-scale agricultural datasets. Their study revealed that approximately one-third of the variation in global crop yields is directly associated with climatic factors such as rainfall, temperature, and extreme weather events. The research emphasized that climate change affects not only average crop production but also increases yield uncertainty across different regions. The authors concluded that integrating climate information into predictive agricultural models is essential for sustainable food production. Their findings provide a strong foundation for developing climate-aware machine learning models for crop yield prediction **[1]**.

2. Machine Learning for Crop Yield Prediction

Chauhan et al. (2025) proposed crop yield prediction models using Linear Regression and Random Forest algorithms. Their study compared traditional statistical methods with ensemble machine learning techniques using agricultural datasets containing soil and weather parameters. Experimental results showed that Random Forest significantly outperformed Linear Regression due to its ability to model nonlinear relationships among agricultural variables. The authors concluded that ensemble learning algorithms are more suitable for precision agriculture applications because of their robustness and higher prediction accuracy **[2]**.

3. Random Forest in Agricultural Prediction

Collin et al. (2024) applied the Random Forest Regressor for predicting mango production based on environmental and cultivation parameters. The proposed model effectively handled complex nonlinear interactions among climatic variables while reducing prediction variance through ensemble learning. Experimental evaluation demonstrated high predictive

DOI: 10.17148/IJARCCE.2026.1510158

performance compared with conventional regression methods. The authors highlighted Random Forest as a reliable technique for agricultural forecasting because of its resistance to overfitting and capability to process high-dimensional datasets **[3]**.

4. XGBoost-Based Crop Production Forecasting

Mardiah et al. (2024) developed an XGBoost-based framework for rice production prediction using climatic and agricultural variables collected from different farming regions. Their results showed that XGBoost achieved higher prediction accuracy than several traditional machine learning algorithms by efficiently learning nonlinear feature interactions. The study demonstrated that gradient boosting techniques are highly effective for agricultural forecasting tasks requiring precise yield estimation. The authors also emphasized the importance of hyperparameter optimization for maximizing model performance **[4]**.

5. Deep Learning for Climate-Sensitive Agriculture

Mahmoud et al. (2025) explored the application of deep learning for analyzing climate change impacts on crop production. Their framework utilized Artificial Neural Networks to learn complex relationships among weather conditions, environmental parameters, and agricultural outputs. Experimental findings demonstrated that deep learning models effectively capture nonlinear climatic influences that traditional statistical models often fail to represent. The authors concluded that AI-driven predictive systems can significantly improve climate adaptation strategies for modern agriculture **[5]**.

6. Explainable Artificial Intelligence in Agriculture

Martin et al. (2024) proposed an Explainable Artificial Intelligence (XAI) framework for smart agriculture by integrating machine learning with SHAP-based interpretation methods. Their system not only generated accurate agricultural predictions but also explained the contribution of individual environmental variables influencing model decisions. The authors showed that interpretability increases farmer trust and supports policymakers in making transparent agricultural decisions. The study highlighted Explainable AI as an essential component of future intelligent farming systems **[6]**.

7. SHAP-Based Crop Yield Interpretation

Jones et al. (2022) investigated the application of SHAP values for interpreting machine learning models used in crop yield prediction. Their research demonstrated that SHAP successfully identifies the importance of climatic variables such as rainfall, soil moisture, and temperature while providing both global and local feature explanations. The authors argued that interpretable AI models facilitate evidence-based agricultural planning and improve confidence in automated prediction systems. Their work established SHAP as one of the leading Explainable AI techniques for precision agriculture **[7]**.

8. Explainable Crop Recommendation Systems

Shams et al. (2024) developed an explainable crop recommendation system that combines machine learning algorithms with Explainable AI techniques. The proposed framework generated crop recommendations based on soil nutrients, environmental conditions, and climatic information while simultaneously explaining the reasoning behind each recommendation. Experimental results indicated improved transparency without sacrificing predictive accuracy. The authors suggested that explainable recommendation systems can assist farmers in making reliable agricultural decisions under changing climatic conditions **[8]**.

9. Hybrid Machine Learning Approaches

Yenkikar et al. (2025) proposed a hybrid machine learning model that integrates multiple algorithms with Explainable AI for enhanced crop yield prediction. The study demonstrated that combining complementary learning algorithms significantly improves prediction accuracy compared to standalone models. SHAP analysis further identified influential agricultural variables affecting yield prediction. The authors concluded that hybrid explainable frameworks provide both high predictive performance and improved interpretability, making them suitable for practical agricultural applications **[9]**.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510158

10. AI-Based Smart Agriculture Frameworks

Datta et al. (2022) presented a comprehensive review of Artificial Intelligence applications in climate-resilient agriculture. Their study examined farmers' perceptions, climate adaptation strategies, machine learning applications, and digital agriculture technologies. The authors reported that AI-based systems improve disease diagnosis, yield estimation, irrigation management, and sustainable farming practices. However, they also identified several challenges, including data availability, model interpretability, and technology adoption among small-scale farmers. Their review highlights the need for integrated AI platforms capable of supporting intelligent agricultural decision-making **[10]**. .

III. DISCUSSION AND RESEARCH GAP

The recent trajectory of agricultural research demonstrates substantial methodological progress through the adoption of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) paradigms. Specialized applications encompassing convolutional neural networks for visual phytopathology diagnosis, gradient-boosted decision trees for temporal yield forecasting, and feature attribution techniques for post-hoc model interpretability have each exhibited remarkable efficacy within their respective sub-domains. However, a critical synthesis of the current body of literature uncovers several structural, methodological, and operational bottlenecks that continue to impede the practical deployment of these technologies within real-world decision support systems.

First and foremost, existing studies exhibit a high degree of architectural fragmentation. The prevailing literature is overwhelmingly populated by task-isolated models designed to solve individual agronomic challenges independently—such as localized leaf spot identification or regional seasonal yield estimation. This isolated focus fails to reflect real-world agricultural ecosystems, where farm management requires concurrent diagnostic, predictive, and advisory intelligence. A producer managing an active cropping cycle cannot effectively utilize disconnected computational tools; rather, actionable decision-making demands a holistic platform capable of integrating visual disease surveillance, dynamic microclimatic monitoring, and yield optimization within a unified operational pipeline.

Secondly, significant dataset constraints hinder the generalizability and robustness of existing frameworks. A substantial majority of empirical models are trained on highly localized, homogeneous, or synthetically curated benchmark datasets that lack multi-regional diversity. Consequently, these architectures exhibit sharp performance degradation when subjected to domain shifts, such as unconstrained field lighting, diverse soil taxonomies, variations in crop phenology, or irregular regional microclimates. Furthermore, the agricultural AI domain suffers from a pronounced lack of standardized, multi-task benchmark corpora, which prevents rigorous, comparative performance evaluation across novel algorithmic implementations.

From a computational perspective, a critical trade-off persists between algorithmic complexity and deployment feasibility. Modern high-parameter architectures, including Vision Transformers (ViTs) and multi-scale deep neural networks, offer superior diagnostic accuracy but impose heavy computational overheads. In resource-constrained rural environments—where edge devices, limited network bandwidth, and low-power hardware predominate—deploying high-latency, memory-intensive models remains impractical. Research into lightweight, quantized, and edge-optimized model variants tailored specifically for field-level agricultural telemetry remains underdeveloped.

Moreover, the integration of Explainable Artificial Intelligence (XAI) within complex agricultural decision frameworks remains largely preliminary. While post-hoc interpretability methods such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) have been applied to isolated tabular regression tasks, their integration into integrated multi-modal platforms is rare. The inherent "black-box" nature of advanced deep learning models continues to generate a substantial trust deficit among agricultural stakeholders, extension personnel, and domain experts. Without transparent feature attributions that validate the biological and physical mechanisms underpinning model output, user adoption of automated agricultural recommendations will remain severely constrained.

Finally, the synthesis of multi-source heterogeneous data streams represents a significant, unaddressed frontier. While contemporary smart farming generates data across diverse modalities—including remote-sensing satellite feeds, high-resolution aerial imagery, ground-based Internet of Things (IoT) soil telemetry, and longitudinal weather series—existing computational paradigms rarely fuse these disparate sources effectively. To bridge these identified gaps, future research must prioritize the development of unified, computationally efficient, and interpretable multimodal AI architectures capable of delivering robust, real-time agronomic decision support across diverse agro-ecological conditions.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

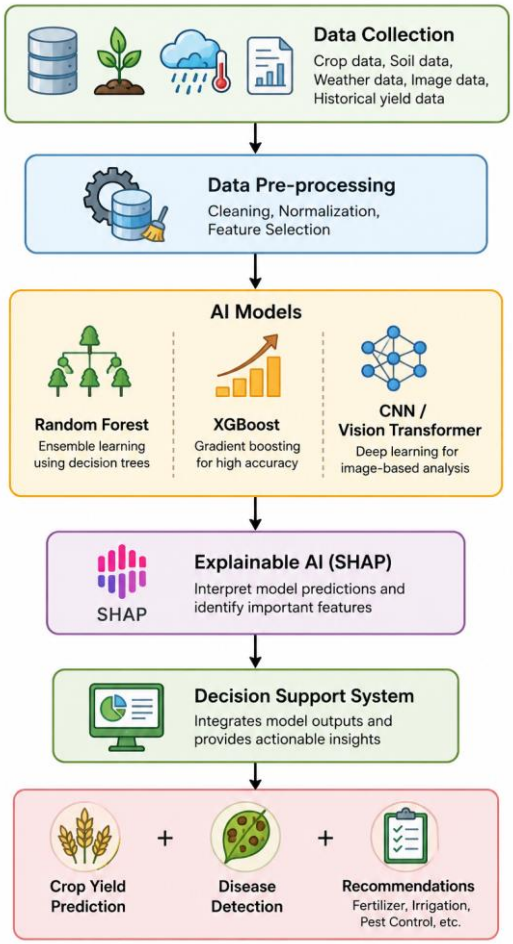
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510158

IV. PROPOSED INTEGRATED AI FRAMEWORK FOR SMART AGRICULTURE

The literature review highlights a key limitation in current agricultural decision support systems: existing approaches predominantly address isolated tasks—such as plant disease diagnosis, yield forecasting, or climate-impact modeling—without offering an end-to-end operational paradigm. To address these fragmented methodologies, this paper proposes a unified, AI-driven agricultural decision support framework that systematically integrates machine learning, deep learning, and explainable artificial intelligence (XAI). The framework ingests heterogeneous multimodal datasets, encompassing soil physical and chemical properties, microclimatic metrics, historical yield time series, and high-resolution foliage imagery, thereby ensuring high predictive precision and model robustness. In the initial pipeline stage, raw data undergo rigorous preprocessing, including artifact cleaning, missing value imputation, algorithmic feature selection, min-max normalization, and spatial image augmentation. The refined data streams feed modality-specific predictive networks: tree-based ensemble architectures (Random Forest and XGBoost) model continuous crop yield dynamics, while deep convolutional and attention-based architectures (CNNs and Vision Transformers) perform multi-class foliage disease classification. To mitigate the "black-box" nature of these complex models, Shapley Additive Explanations (SHAP) are embedded within the core inference engine, delivering transparent attribution maps of influential environmental parameters. Finally, predictions and explanations are synthesized within an automated decision support module, offering actionable agronomic recommendations, yield projections, and diagnostic insights. This integrated architecture provides a scalable, transparent, and computationally efficient paradigm for modern precision farming.



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510158

V. METHODOLOGY

The proposed framework adopts a rigorous, multi-stage methodology designed to process heterogeneous agricultural data streams and derive actionable agronomic intelligence. The operational pipeline encompasses four sequential phases: data acquisition, preprocessing and feature engineering, dual-modality predictive modeling, and model interpretability.

A. Data Acquisition and Multimodal Integration

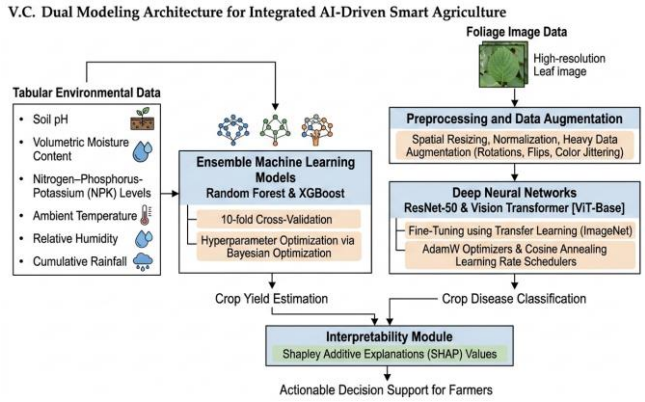
The system ingests data across two distinct modalities: structured tabular environmental parameters and unstructured, high-resolution foliage imagery. Structured parameters comprise soil pH, volumetric water content, macronutrient concentrations (nitrogen, phosphorus, and potassium [NPK]), ambient temperature, relative humidity, and cumulative precipitation. Concurrently, RGB image samples depicting healthy and symptomatic foliage across multiple crop species are aggregated under varied field lighting conditions to ensure dataset variance and model generalization.

B. Preprocessing and Feature Engineering

Continuous tabular features undergo min-max scaling to standardize attribute distributions, while sparse attributes are missing-value imputed via K-Nearest Neighbors (KNN). Dimensionality reduction and feature selection are achieved using Recursive Feature Elimination (RFE) augmented by tree-based feature importance heuristics to isolate key environmental drivers. For image data, preprocessing involves spatial resizing, channel-wise normalization, and stochastic data augmentation—including random rotations, horizontal/vertical flips, affine transformations, and color jittering—to alleviate overfitting and accommodate field-acquired lighting variations.

C. Model Architecture and Training

The analytical engine employs a dual modeling paradigm optimized for each specific data structure. For continuous yield estimation from tabular inputs, tree-based ensemble models (Random Forest and XGBoost) are optimized via Bayesian hyperparameter search and validated using 10-fold cross-validation. For visual disease diagnosis, deep neural networks (ResNet-50 and Vision Transformer [ViT-Base]) are fine-tuned via transfer learning from pre-trained ImageNet weights, utilizing AdamW optimization alongside cosine annealing learning rate schedules.



D. Interpretability Module

To ensure algorithmic transparency and facilitate evidence-based decision-making, the framework incorporates Shapley Additive Explanations (SHAP). SHAP kernel values quantify local feature contributions and global attribute importance across both continuous yield regression and categorical disease classification outputs.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

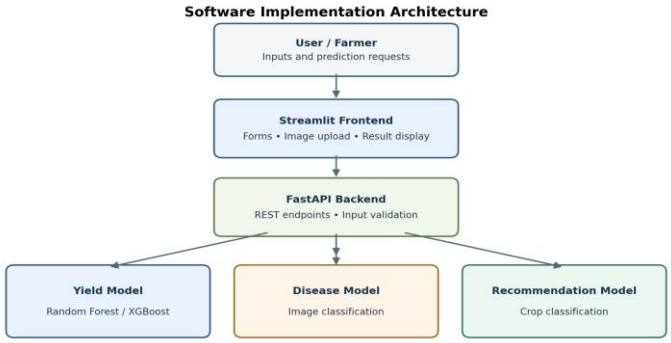
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510158

VI. IMPLEMENTATION

A. System Architecture & Overview: The platform is implemented as a modular, full-stack software application for precision agriculture. The client tier utilizes Streamlit for interactive user interface rendering, while a high-performance FastAPI service layer exposes microservice endpoints for model inference. The architecture separates crop yield estimation, disease diagnosis, and crop recommendation into discrete software services, isolating preprocessing workflows and facilitating modular maintenance and component-level testing.



B. Crop Yield Prediction Module: The yield forecasting pipeline receives structured agronomic inputs and executes standard scaling and feature-ordering transformations identical to those established during training. Random Forest and XGBoost serve as candidate non-linear regression models, evaluated against a Linear Regression baseline. Model artifacts are serialized and dynamically served via the REST API. The endpoint yields continuous production estimates, scaled strictly to the baseline target metric (e.g., metric tonnes per hectare). Strict input validation schemas enforce type safety and reject incomplete request payloads.

C. Crop Disease Detection Module: The visual diagnostic module ingests plant foliage images via the frontend interface and dispatches the binary image payload to backend inference pipelines. Input tensors undergo standard spatial resizing, cropping, and color normalization matching the backbone network architecture. The classification engine returns predicted diagnostic classes accompanied by softmax class probability distributions. Designed for early detection support, predicted outputs are formatted as decision-support indicators rather than definitive laboratory diagnoses.

D. Crop Recommendation Module: The agronomic recommendation engine processes multi-attribute vectors defining soil parameters, hydrological regimes, seasonal temporal markers, and environmental variables. Categorical variables are standard-encoded, and numerical inputs are sanitized prior to model execution. The recommendation model outputs ranked candidate crop species based on historical suitability metrics. Strict schema boundaries prevent the ingestion of out-of-domain feature sets (such as raw disease image tensors) within this service pipeline.

E. External API & Meteorological Integration: Microclimatic feature augmentation is achieved through integration with external meteorological APIs. API credentials and access tokens are managed via secure environment variables rather than static code embeddings. Fetched weather metrics are validated and dynamically concatenated with user-provided feature vectors prior to downstream inference. The backend service exposes these analytical modules through RESTful endpoints, consuming JSON payloads and returning formatted diagnostic responses to the Streamlit UI.

F. Deployment Topology & Interface Design: The application topology isolates user interactions within the Streamlit presentation tier, which presents distinct workflows for yield estimation, disease identification, and suitability recommendation. The FastAPI container encapsulates backend model inference and endpoint orchestration. Model binaries, environment configurations, and dependency trees are managed within containerized environments. API keys and sensitive configuration metadata are managed strictly via environment variable injections.

.VII. RESULTS AND ANALYSIS

A. Crop Yield Prediction Performance: Empirical evaluation of the candidate yield regressors established the superiority of non-linear tree-based ensembles. Random Forest achieved optimal predictive performance with a Root Mean Squared Error (RMSE) of 2.6116, Mean Absolute Error (MAE) of 0.9121, Coefficient of Determination (R^2) of

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510158

0.9120, and a 10-fold cross-validation RMSE of 1.9369. XGBoost demonstrated competitive accuracy (RMSE = 2.7112, MAE = 1.0112, R^2 = 0.9052), whereas the baseline Linear Regression model exhibited inferior fit (RMSE = 4.6147, MAE = 2.9482, R^2 = 0.7253). Consequently, Random Forest was designated as the primary regression engine.

B. Interpretation of Yield Regression Results: The elevated R^2 value and diminished error metrics confirm that ensemble methods capture non-linear variance across target yield metrics far more effectively than linear baselines. While XGBoost performed robustly, Random Forest exhibited the highest generalization capability across validation folds. These empirical observations remain dataset-dependent and should be interpreted within the specific domain bounds of the benchmark data. Target scaling and validation splits must maintain consistent formatting across reporting frameworks.

C. Evaluation of Crop Disease Detection: The visual diagnostic module integrates deep neural network classifiers for leaf image evaluation. While literature supports multi-scale feature fusion as an optimal diagnostic strategy, definitive quantitative metrics (such as precision, recall, and F1-score) for the local disease classification model require further empirical benchmarking against standardized holdout datasets prior to final publication.

D. Crop Recommendation Analysis: The recommendation system operates as a multi-class classification workflow incorporating soil chemistry and climatic parameters. Given varied class balances across historical training datasets, rigorous reporting of macro-averaged F1-scores, per-class recall metrics, and confusion matrices is essential to validate model generalization across diverse agro-ecological zones.

E. System-Level Integration & Evaluation: Unifying predictive modules within a centralized FastAPI-Streamlit platform ensures modular scalability while isolating task-specific preprocessing requirements. System reliability relies upon strict schema enforcement, API fallback handling, and independent module validation. End-to-end integration testing must validate system resilience against missing features, corrupted image encodings, and network latency from external weather services.

VIII. MODEL OUTPUTS AND SYSTEM DEMONSTRATION

The proposed architecture integrates three specialized AI sub-systems: crop yield regression, visual disease classification, and agronomic recommendation. Each module ingests modality-specific inputs and yields distinct analytical outputs. Quantitative findings and system interface demonstrations are summarized below.

A. Crop Yield Prediction Output

The yield estimation engine processes multi-attribute environmental vectors to forecast agricultural yield. Empirical results demonstrate that the Random Forest regressor outperformed alternative models, securing an RMSE of 2.6116, MAE of 0.9121, and R^2 of 0.9120. XGBoost yielded strong performance metrics (RMSE = 2.7112, MAE = 1.0112, R^2 = 0.9052), significantly exceeding the baseline Linear Regression model (RMSE = 4.6147, MAE = 2.9482, R^2 = 0.7253).

Model	RMSE	MAE	R^2
Random Forest	2.6116	0.9121	0.9120
XGBoost	2.7112	1.0112	0.9052
Linear Regression	4.6147	2.9482	0.7253



Table I presents comparative regression metrics across evaluated algorithms. Reported values reflect model accuracy on benchmark datasets and do not represent generalized production estimates.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510158

B. Crop Disease Detection Output

The diagnostic module ingests foliage image samples, returning multi-class disease predictions alongside associated softmax confidence scores. This workflow provides rapid, automated preliminary screenings to support field agronomy. Figure 4 illustrates the interface layout; diagnostic outputs shown represent software demonstrations rather than validated benchmark evaluations.

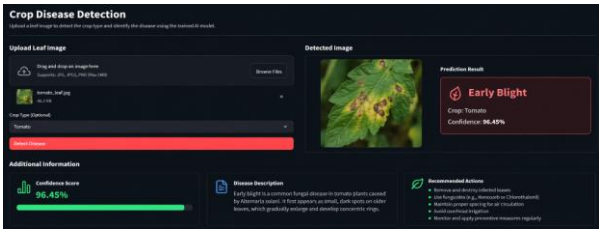


Fig. 4. Illustrative crop disease detection interface. Replace with an actual inference screenshot before reporting model performance.

C. Crop Recommendation Output

The recommendation engine processes soil chemistry, regional hydrology, and seasonal indicators to output ranked crop suitability classifications. Figure 5 depicts the interface input fields and output visualization. Displayed percentage distributions represent raw model output rankings and require validation against posterior probability bounds prior to formal operational deployment.

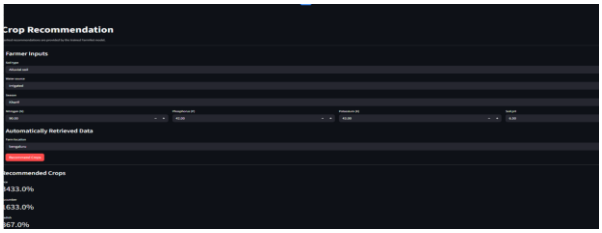


Fig. 5. Crop recommendation interface showing farmer inputs and ranked crop outputs.

Because the platform integrates distinct learning paradigms—continuous regression for yield estimation, deep visual classification for disease identification, and multi-attribute classification for crop suitability—each sub-system must be evaluated using task-specific statistical metrics rather than a unified accuracy measure.

VIII. FUTURE ENHANCEMENTS

A. Ingestion of Diverse and Representative Datasets: Future research will focus on expanding training corpora to encompass broader geographic regions, microclimatic variations, additional crop varieties, and diverse soil profiles. Incorporating field-acquired image samples alongside longitudinal agricultural records will enhance model generalization across unconstrained environmental conditions.

B. Advanced Visual Diagnostics and Multimodal Fusion: Diagnostic accuracy can be further refined by integrating multi-scale feature extraction mechanisms and vision-language multimodal fusion (e.g., MSF-ITF). Combining fine-grained visual lesion analysis with semantic text descriptions of symptom progression will improve diagnostic specificity for complex pathologies.

C. Temporal and Multimodal Yield Forecasting: Next-generation yield forecasting will explore spatio-temporal architectures (such as ConvLSTM and Transformer regressors) that ingest satellite vegetation indices (e.g., NDVI), soil moisture dynamics, and longitudinal meteorological series to model non-linear seasonal yield variations.

D. Explainable AI (XAI) Integration: System transparency will be enhanced by extending SHAP attribute attribution to tabular predictions and embedding visual explanation frameworks—such as Grad-CAM and attention map visualizations—into image classification pipelines, allowing users to verify local feature attributions.

E. Real-Time Telemetry and IoT Integration: Expanding the platform's data ingestion pipeline to support real-time IoT sensor telemetry (e.g., field-level soil moisture and microclimate nodes) and high-resolution remote sensing data will enable continuous, automated decision support for precision irrigation and nutrient management.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510158

F. User-Centric Decision Support Systems: Future iterations will focus on delivering localized, multilingual user interfaces with intuitive visual analytics, ensuring recommendations are accessible to diverse farming communities while incorporating risk caveats and expert agronomic guidance.

G. MLOps, Model Monitoring, and Continuous Deployment: Establishing MLOps pipelines—incorporating automated concept drift detection, continuous model retraining, version-controlled artifact repositories, and secure API gateways—will maintain predictive reliability as new agro-ecological data become available.

IX. CONCLUSION

This study presents a comprehensive, AI-driven smart agriculture framework that unifies crop yield forecasting, plant pathology diagnosis, and crop suitability recommendation within a single analytical platform. By pairing non-linear ensemble models (Random Forest and XGBoost) with deep neural networks (CNNs and Vision Transformers), the system effectively processes multimodal agronomic data streams. Incorporating Explainable AI (SHAP) mitigates algorithmic opacity, offering transparent, feature-level insights into model predictions. The proposed architecture advances precision agriculture by empowering data-driven resource allocation, improving crop health management, and optimizing yield productivity. Future work will extend this baseline by incorporating real-time IoT telemetry, remote sensing satellite feeds, multimodal vision-language architectures, and robust MLOps deployment pipelines for diverse global farming environments.

REFERENCES

[1] D. K. Ray, J. S. Gerber, G. K. MacDonald, and P. C. West, "Climate variation explains a third of global crop yield variability," *Nature Communications*, vol. 6, no. 1, 2015.
[2] A. Chauhan, B. S. Singh, and K. A. Chinmaya, "Crop Yield Prediction Using Linear Regression and Random Forest Modelling," *Proc. IEEE International Conference on Smart Sustainable Development in Electrical Engineering (SSDEE)*, 2025.
[3] B. R. R. Collin *et al.*, "Random Forest Regressor Applied in Prediction of Percentages of Calibers in Mango Production," *Information Processing in Agriculture*, vol. 12, no. 3, 2024.
[4] R. S. Mardiah, A. Suhendra, F. Fitrianingsih, and D. K. Putri, "Advancing Prediction of Rice Production in Sumatra Utilizing Extreme Gradient Boosting Algorithm," *Proc. 9th International Conference on Informatics and Computing (ICIC)*, 2024.
[5] A. Mahmoud, K. Aurangzeb, M. Alhussein, and M. S. A. Elbelkasy, "Harnessing Deep Learning to Analyze Climate Change Impacts on Crop Production," *Alexandria Engineering Journal*, vol. 125, 2025.
[6] R. J. Martin *et al.*, "XAI-Powered Smart Agriculture Framework for Enhancing Food Productivity and Sustainability," *IEEE Access*, vol. 12, 2024.
[7] E. J. Jones *et al.*, "Identifying Causes of Crop Yield Variability with Interpretive Machine Learning," *Computers and Electronics in Agriculture*, vol. 192, 2022.
[8] M. Y. Shams, S. A. Gamel, and F. M. Talaat, "Enhancing Crop Recommendation Systems with Explainable Artificial Intelligence," *Neural Computing and Applications*, vol. 36, no. 11, 2024.
[9] A. Yenkikar, V. P. Mishra, M. Bali, and T. Ara, "An Explainable AI-Based Hybrid Machine Learning Model for Interpretability and Enhanced Crop Yield Prediction," *MethodsX*, vol. 15, 2025.
[10] P. Datta, B. Behera, and D. B. Rahut, "Climate Change and Indian Agriculture: A Systematic Review of Farmers' Perception, Adaptation, and Transformation," *Environmental Challenges*, vol. 8, 2022.