SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 3, 2026

GREENAI: AN ESP32-BASED OFFLINE IOT SENSING AND EDGE-INFERENCE SYSTEM FOR PRECISION AGRICULTURE in Pakistan

Qasim Ali Hashim[*1], Zain Sajjad[2], Muhammad Saqib Ali[3]

[*1]*Department of Information Technology, University of Agriculture Faisalabad*
[2]*Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University*
[3]*Department of Environmental Engineering University of Agriculture Faisalabad*

[*1]qasimali782@gmail.com

DOI: https://doi.org/10.5281/zenodo.23834817

Keywords
Wireless sensor node; ESP32; Edge inference; Offline-first IoT; Sensor validation; Crop recommendation; Plant disease detection; Precision agriculture.

Article History
Received: 01 January 2026
Accepted: 16 March 2026
Published: 31 March 2026

Copyright @Author
Corresponding Author: *
Qasim Ali Hashmi

Abstract

Precision agriculture in low-connectivity regions depends on rugged, low-cost field hardware and decision-support software that can operate without continuous cloud access. This article presents GreenAI, an offline-first precision-agriculture system engineered around an ESP32-based wireless sensor node and an edge-assisted inference pipeline. The sensing node integrates soil-moisture, temperature and humidity (DHT22), pH, and light-intensity (BH1750) sensors together with a GPS module and on-board MicroSD logging; it acquires field measurements locally and synchronises them to a FastAPI/SQLite backend over Wi-Fi through a store-and-forward protocol that tolerates intermittent connectivity. Soil-moisture and pH readings from the node were validated against calibrated reference instruments to quantify sensing accuracy. On this platform, three analytical modules operate as interchangeable software components: a crop-recommendation model, a plant-disease classifier, and an explainable-AI layer. A Punjab-oriented crop-recommendation dataset of 2,300 plus balanced samples across 19 crops compiled from public Punjab Agriculture Department crop and agro-climatic records and class-balanced for training was used to benchmark five classifiers; XGBoost offered the best accuracy-to-footprint trade-off (97.40% test accuracy, macro F1-score 0.9741), while a compact 0.43 MB SVM was retained for on-device edge inference. A plant-disease classifier based on EfficientNetV2-S, trained on a unified 113-class image set, reached 99.03% validation accuracy. SHAP, partial-dependence, and Grad-CAM analyses were incorporated for interpretability. A working field prototype demonstrated end-to-end operation—local sensing, offline buffering, wireless transfer, and dashboard visualisation—showing GreenAI to be a deployable engineering solution for resource-constrained farms. A multilingual mobile advisory application and market-information module are under active development.

I. INTRODUCTION

Agriculture plays a main role in ensuring food security, economic growth and livelihoods of rural mostly in developing countries where a significant portion of the population depends on farming. Modern agriculture faces numerous challenges including climate variability, declining soil fertility, water scarcity, pest infestations, disease outbreaks and also fluctuating market conditions (Shahi et al., 2025). These factors make agricultural decision-making increasingly complex and necessitate intelligent solutions capable of improving productivity while promoting sustainable resource utilization. Recent advances in Artificial Intelligence (AI), Machine Learning (ML) and Internet of Things (IoT) technologies have created new opportunities for precision

agriculture (Kumar et al., 2026; Emon et al., 2025) like Data-driven crop recommendation systems can assist farmers in selecting suitable crops based on soil and environmental conditions, while deep learning techniques enable automated plant disease diagnosis from leaf images (Pande et al., 2021; Doshi et al., 2018; Kulkarni et al., 2018; Mohanty et al., 2016). IoT-enabled sensing platforms have facilitated real-time monitoring of agricultural environments providing valuable information for timely interventions and efficient farm management (Ouammi et al., 2024; Mohapatra et al., 2023; Lombardo et al., 2021). Explainable Artificial Intelligence (XAI) further enhances transparency and user trust by providing interpretable insights into model predictions (Arrieta et al., 2020). Despite these developments most existing studies focus on individual agricultural tasks such as crop recommendation, disease diagnosis, environmental monitoring or explainability in isolation and many available solutions depend heavily on cloud-based infrastructures which require continuous internet connectivity and are developed using generic datasets that may not adequately represent local agricultural conditions. These limitations restrict their practical applicability in rural farming environments, particularly in regions characterized by intermittent connectivity and diverse agro-climatic conditions. In particular, few of these systems are built as integrated, field-deployable sensing platforms with validated low-cost sensors and on-device inference; most remain cloud-dependent software pipelines evaluated in isolation from the hardware on which they would run.

To address these challenges, this paper presents GreenAI, an offline-first precision-agriculture system engineered around a low-cost ESP32 wireless sensor node and an edge-assisted inference pipeline for Pakistan's Punjab region. The system is organised as a sensing-and-communication backbone on which analytical functions run as modular software components. The ESP32 node integrates soil-moisture, temperature-humidity, pH, and light-intensity sensors with a GPS module and on-board MicroSD logging, acquiring field data locally and synchronising to a FastAPI/SQLite backend through a store-and-forward protocol that tolerates intermittent connectivity. Soil-moisture and pH readings are validated against calibrated reference instruments to characterise sensing accuracy. Three analytical modules—crop recommendation, EfficientNetV2-S plant-disease diagnosis, and an explainable-AI layer based on SHAP, partial-dependence, and Grad-CAM analyses operate on this platform, with a compact SVM retained for on-device edge inference. A Punjab-oriented crop-recommendation dataset of 9,500 balanced samples across 19 crops, compiled from public Punjab Agriculture Department crop and agro-climatic records and class-balanced for training, is used to benchmark the recommendation models.

The remainder of this paper is organized as follows. Section II presents the literature survey and comparative analysis of existing approaches. Section III discusses the identified research gaps, while Section IV summarizes the major contributions of the proposed work. Section V describes the system architecture, methodology and implementation details. Experimental results and performance evaluations are presented in Section VI, and finally Section VII concludes the paper and outlines future research directions.

II. Literature Survey

Precision agriculture has leveraged machine learning, deep learning, explainable artificial intelligence (XAI) and Internet of Things (IoT) technologies to improve agricultural decision-making also Machine learning approaches have been widely adopted for crop recommendation using soil nutrients, rainfall, humidity, temperature and pH parameters, with ensemble methods such as XGBoost demonstrating strong predictive performance for structured agricultural datasets (Chen and Guestrin, 2016; Madhuri and Indiramma, 2021). For plant disease diagnosis, deep learning models have achieved high classification accuracy using leaf image datasets, particularly with convolutional neural network architectures and EfficientNet-based models (Tan and Le, 2019; Tan and Le, 2021; Ferentinos, 2018).

To improve transparency and user trust, explainable AI techniques such as SHAP and

Grad-CAM have been incorporated into agricultural systems to provide feature-level and visual explanations of model predictions (Lundberg and Lee, 2017; Selvaraju et al., 2017). Meanwhile, IoT-based sensing platforms have enabled real-time environmental monitoring through low-cost sensor networks, supporting precision farming and smart agricultural management (Hasib and Akib, 2026; Rahman et al., 2025; Ayaz et al., 2019; Khanna and Kaur, 2019).

These developments are most existing studies focus on individual tasks such as crop recommendation, disease diagnosis, explainability or environmental monitoring separately. Limited attention has been given to integrating these components into a unified, deployment-ready framework capable of operating under low-connectivity rural environments. Furthermore, region-specific agricultural intelligence and offline operational capabilities remain largely underexplored. These limitations motivate the development of the proposed GreenAI framework. The comparison of the proposed work with the existing literature is presented in Table I.

Table I. Comparison of Proposed work with State of the Art.

Ref	Focus Area	Methodology	Key Contributions	Limitations
(Chen and Guestrin, 2016)	Crop Recommendation	XGBoost	High prediction accuracy on structured agricultural data	No integrated agricultural ecosystem
(Mohanty et al., 2016)	Disease Detection	CNN (PlantVillage)	Accurate disease classification	Controlled laboratory images
(Tan and Le, 2021)	Disease Detection	EfficientNetV2	Improved accuracy and computational efficiency	Limited integration with field deployment
(Selvaraju et al., 2017)	Explainable AI	Grad-CAM	Visual explanation of disease predictions	Focused only on image interpretation
(Lundberg and Lee, 2017)	Explainable AI	SHAP	Feature-level model interpretability	Not integrated with IoT systems
(Shahi et al., 2025)	Precision Agriculture Review	ML-based Agriculture	Comprehensive overview of AI in agriculture	Review only
(Ouammi et al., 2024)	IoT Agriculture	Smart Sensing Systems	Real-time agricultural monitoring	Limited AI integration
(Lombardo et al., 2021)	Smart Farming	IoT Architecture	Environmental monitoring and automation	No explainability support
(Pande et al., 2021)	Crop Recommendation	ML-based Recommendation	Soil and climate-based crop prediction	Standalone recommendation system
(Hughes and Salathé, 2015)	Disease Detection	Dataset Analysis	Impact of dataset diversity on accuracy	No deployment framework
(Khanna and Kaur, 2019)	Deep Learning Agriculture	Review Study	DL applications in agriculture	No practical implementation

IEER
Institute for Excellence in Education & Research

Proposed	GreenAI	XAI + IoT + Edge AI	Unified offline-first framework integrating recommendation, disease detection, IoT telemetry, explainability, and edge deployment	–

III. Research Gaps

Recent advances in precision agriculture have demonstrated the effectiveness of machine learning, deep learning, explainable artificial intelligence (XAI), and Internet of Things (IoT) technologies for agricultural decision support. Existing studies have reported promising results in crop recommendation, plant disease diagnosis, environmental monitoring, and smart farming automation. However, most of these solutions focus on individual agricultural tasks and are developed as standalone systems. Consequently, farmers often need to rely on multiple disconnected platforms for crop selection, disease management, environmental monitoring, and market-related decision-making.

Furthermore, many reported frameworks are dependent on cloud-based infrastructures and continuous internet connectivity, limiting their applicability in rural and resource-constrained agricultural environments. The majority of existing systems also employ generic datasets and they have lack of region-specific agricultural intelligence which decreasing their adaptability to local climatic conditions, soil characteristics and crop cultivation practices. Although explainable AI techniques have gained increasing attention, their integration with practical agricultural decision-support systems remains limited. Similarly, lightweight edge deployment and offline operation capabilities are rarely considered despite their importance for real-world implementation.

To clearly identify these limitations Table II summarizes the major gaps observed in the literature and the corresponding solutions provided by the proposed GreenAI framework.

Table II. Research Gap Analysis and Proposed Solution

Research Area	Existing Approaches	Identified Gap	Proposed Solution in GreenAI
Crop Recommendation	Machine learning-based recommendation systems	Standalone functionality	Integrated within a unified agricultural platform
Plant Disease Detection	CNN and deep learning-based diagnosis	Limited interaction with advisory systems	Unified disease diagnosis and advisory support
Explainable AI	SHAP or Grad-CAM applied independently	Limited integration into practical farming systems	Integrated SHAP, PDP, and Grad-CAM framework
IoT-Based Agriculture	Real-time environmental monitoring	Lack of intelligent decision support	IoT-enabled AI-assisted recommendations
Smart Farming Platforms	Cloud-centric architectures	Dependence on continuous internet connectivity	Offline-first architecture with local caching
Agricultural Intelligence	Generic agricultural datasets	Poor regional adaptability	Punjab-specific agricultural intelligence
Edge Deployment	Computationally intensive models	Limited deployment on low-cost hardware	Lightweight edge-compatible inference models

The literature analysis indicates that there is a clear need for an integrated, explainable, region-specific and deployment-ready ecosystem that can precision agriculture. Motivated by these limitations, this work proposes GreenAI which is an offline-first framework that combines crop

recommendation, plant disease diagnosis, IoT-based environmental monitoring, explainable AI and also edge intelligence within a single scalable platform tailored to Punjab's agricultural ecosystem.

IV. Contributions of the paper

The major contributions of this paper are summarized as follows:

- Design and integration of a low-cost ESP32-based wireless sensor node for field deployment, combining soil-moisture, temperature–humidity (DHT22), pH, and light-intensity (BH1750) sensing with a GPS module and on-board MicroSD logging for autonomous, offline-capable data acquisition.
- An offline-first data pipeline that buffers sensor measurements locally and synchronises them to a FastAPI/SQLite backend over Wi-Fi using a store-and-forward protocol, together with lightweight edge inference that enables on-device crop prediction without continuous cloud connectivity.
- Experimental validation of the low-cost soil-moisture and pH sensors against calibrated reference instruments, reported with coefficient of determination (R^2) and root-mean-square error (RMSE), establishing the measurement reliability of the sensing node.
- A crop-recommendation module benchmarked across five classifiers (XGBoost, Random Forest, ExtraTrees, LightGBM, and SVM) on a Punjab-oriented dataset of 9,500 balanced samples across 19 crops compiled from public Punjab Agriculture Department crop and agro-climatic records; XGBoost achieved 97.40% test accuracy and a macro F1-score of 0.9741, while a compact 0.43 MB SVM was retained for edge deployment.
- A plant-disease diagnosis component based on EfficientNetV2-S, trained on a unified 113-class image dataset and achieving 99.03% validation accuracy for scalable disease identification across multiple crop species.
- Integration of explainable-AI techniques—SHAP, partial-dependence plots, and Grad-CAM—to improve the transparency and interpretability of both the crop-recommendation and disease-diagnosis components.

V. Methodology

A. System Architecture

GreenAI is an offline-first precision-agriculture system engineered for field deployment in low-connectivity environments. Its foundation is a sensing-and-communication layer built on an ESP32-based wireless sensor node; the analytical functions operate on top of this layer as modular, replaceable software components. The architecture comprises four engineering layers: (i) a field sensing layer, in which the ESP32 node acquires soil-moisture, temperature–humidity, pH, light-intensity and location data; (ii) a data-flow and communication layer that buffers measurements on-device and synchronises them to a FastAPI/SQLite backend over Wi-Fi using a store-and-forward protocol tolerant of intermittent connectivity; (iii) an inference layer offering both server-side models and a lightweight on-device edge model for crop prediction; and (iv) an application layer that returns recommendations to the farmer. Field parameters—soil nutrients (N, P, K), temperature, humidity, rainfall, soil moisture and pH—are acquired automatically by the sensor node or entered manually, and leaf images captured on mobile devices feed the disease-diagnosis component. Sensor readings and prediction outputs are cached locally and synchronised automatically when connectivity is available, which is the key engineering property enabling reliable operation in resource-constrained settings. A multilingual mobile advisory application and a market-information module are under active development and are not evaluated in this paper.

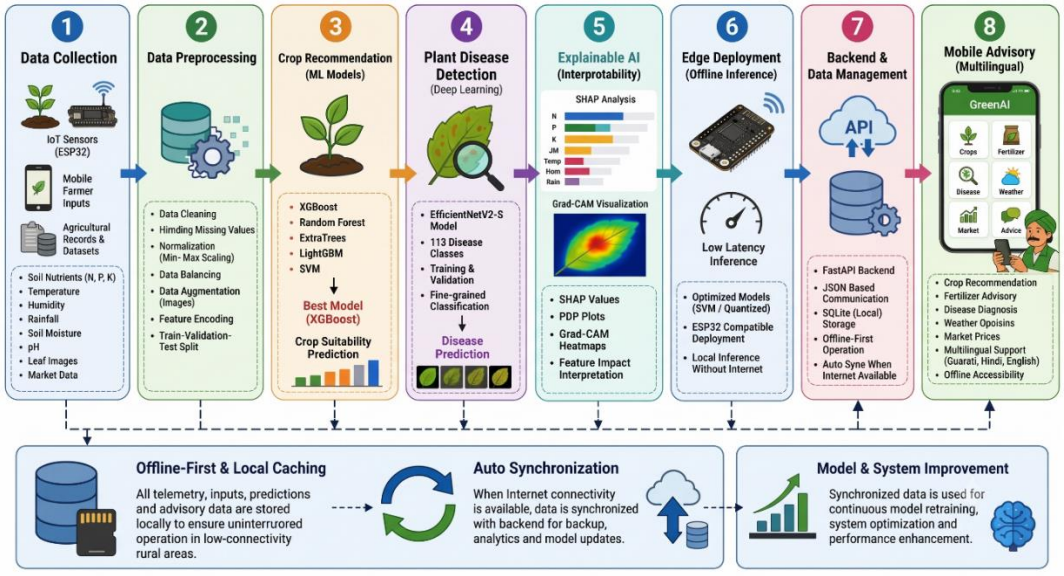


Fig. 1. Overall Architecture and Workflow of Proposed GreenAI Framework.

The whole workflow and system architecture of the proposed GreenAI framework is shown in Fig. 1. The architecture has many interlinked steps like data gathering, preprocessing, crop recommendation, illness detection, explainable AI analysis, edge deployment, backend communication and also have mobile advice support. It collects environmental data such as soil nutrients, weather conditions and crop related information using IoT sensors and mobile-based farmer inputs. The collected data is subsequently processed by cleaning, normalisation, balancing, and augmentation procedures. The processed data is then provided to machine learning and deep learning models.

The crop recommendation module generates the suitable crop suggestions based on the soil and climatic circumstances using different machine learning techniques as XGBoost, SVM, Random Forest, and LightGBM. Simultaneously, the disease detection module implements the EfficientNetV2-S convolutional neural network architecture for classifying plant diseases based on leaf pictures. For explainable AI analysis, the system integrates with SHAP and Grad-CAM for increased transparency and interpretability.

Moreover, the architecture offers lightweight edge deployment with ESP32 compatible models for real-time offline inference in low-connectivity rural settings. Backend communication with FastAPI and SQLite synchronization techniques for quick data interchange and offline telemetry storage. Finally, the multilingual mobile advisory interface gives prediction findings and agricultural counsel directly to farmers, thus providing a scalable and deployment ready precision agriculture ecosystem for Punjab's farming environment.

B. ESP32 Sensor Node Design and Sensor Integration

The hardware architecture of the sensing node is shown in Fig. 2. The node is built around an ESP32-WROOM-32 microcontroller that provides integrated Wi-Fi and Bluetooth connectivity. Soil-moisture, temperature–humidity (DHT22), pH and light-intensity (BH1750) sensors, together with a GPS module, are interfaced to the ESP32 to capture field-level environmental information. A MicroSD card module provides on-board data logging and offline storage, enabling continuous operation under intermittent network conditions. The modular design allows additional sensors to be integrated readily and supports low-power deployment in resource-constrained agricultural environments.

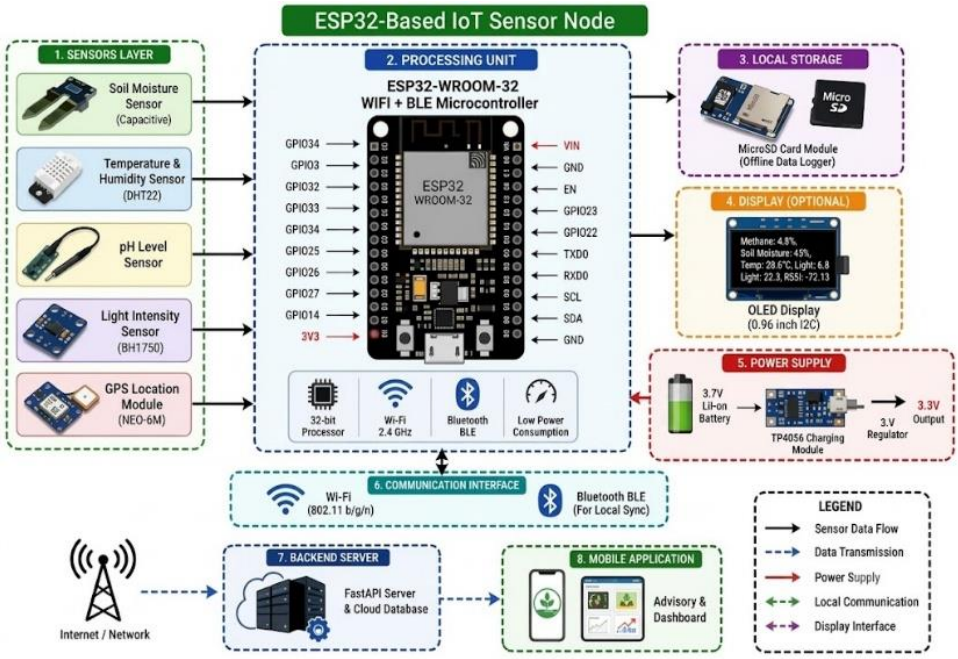


Fig. 2. Architecture of the ESP32-based IoT sensor node used for real-time environmental monitoring and agricultural data acquisition.

C. Offline Data Flow, Synchronisation and Edge Inference

Field measurements acquired by the node are time-stamped and written to the local MicroSD store, then transmitted to the backend as lightweight JSON payloads over Wi-Fi. A FastAPI service backed by an SQLite database receives the telemetry; when connectivity is unavailable, records are buffered on-device and flushed on reconnection through a store-and-forward protocol with retry-on-failure, so that no data are lost during network outages. To support inference without continuous cloud access, a compact Support Vector Machine model was optimised for on-device execution, providing fast crop prediction on the embedded hardware with minimal memory use and latency. This edge-oriented design reduces dependence on network access and improves deployment feasibility in rural farming environments.

D. Data Collection and Pre-processing

The GreenAI decision-support system is built on a heterogeneous corpus spanning tabular agronomic records, geospatial environmental layers, market intelligence, and crop-disease imagery. The crop-recommendation data were compiled from public crop and agro-climatic records published by the Directorate General Agriculture (Extension & Adaptive Research), Agriculture Department, Government of the Punjab, Pakistan (Government of the Punjab, 2024). The dataset includes 19 important crops of Kharif, Rabi and Zaid seasons. The dataset contains over 9500 balanced samples. The features are Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, rainfall and pH value of soil. To minimise class imbalance and increase model generalisation, we utilised data balancing and preprocessing procedures. Environmental drivers were assembled from NASA POWER daily meteorology (MERRA-2/CERES) (NASA, 2024), Pakistan Meteorological Department 0.25° × 0.25° gridded daily rainfall, and Government of Pakistan district soil-moisture records, market advisories use AMIS daily price–arrival feeds with price catalogue (FAO, 2024). For disease detection, we curated a unified 113-class image corpus combining the PlantVillage repository (Hughes and Salathé, 2015) and its augmented Kaggle redistribution (PlantVillage, 2019), with Pakistan and Punjab-relevant crop-specific datasets covering wheat, rice, cotton, mango, chilli and onion, pomegranate, watermelon, soybean, lentil, castor, banana, sugarcane, groundnut, and a multi-crop pest–disease collection, among others. The images

were pre-processed by scaling, normalisation, augmentation and noise reduction to improve the robustness of the classification and reduce the overfitting during the training.

The IoT telemetry subsystem collects environmental data continually with ESP32 microcontrollers coupled to sensors for soil moisture, temperature and humidity monitoring. Sensor data are saved locally and sent to the backend using lightweight JSON based communication protocols.

E. Crop Recommendation Module

The crop recommendation system was built using supervised machine learning techniques. We trained and assessed several classification models including XGBoost, Random Forest, ExtraTrees, LightGBM, and Support Vector Machine (SVM) (Chen and Guestrin, 2016; Madhuri and Indiramma, 2021) with the same stratified train-validation-test divides. We evaluated them on the accuracy of prediction, macro F1-score, inference latency and memory footprint. XGBoost presented the best trade-off between classification performance and feasibility of deployment among all studied models. The trained model is used to predict the most suited crop based on real-time environmental and soil variables. In order to increase the interpretability, we used SHAP (SHapley Additive exPlanations) analysis into the recommendation pipeline to assess the contribution of each agricultural feature to the prediction results.

F. Explainable AI Framework

To increase the interpretability of crop recommendation decisions, SHAP (SHapley Additive exPlanations) was integrated into the proposed XGBoost based CropAdvisor model and SHAP is a game-theoretic explainability technique that quantifies the contribution of each input feature toward the final prediction, conventional feature such as importance measures, SHAP provides both global and local explanations also enabling a comprehensive understanding of model behaviour.

For each prediction SHAP values were computed for all agricultural input variables, including nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, rainfall and soil pH. The generated explanations were utilized to identify influential environmental factors affecting crop recommendation outcomes and to improve transparency in AI-assisted agricultural decision-making. Global feature is importance rankings and SHAP summary visualizations were subsequently arrange to the overall analyze but it different in agronomic parameters on model predictions. SHAP-based feature attribution, Partial Dependence Plot (PDP) analysis was employed to investigate the marginal effect of individual agricultural variables on crop recommendation predictions. PDPs estimate the average influence of a selected feature while averaging out the effects of all remaining features in the dataset. This approach provides an intuitive understanding of how changes in environmental and soil parameters affect model outputs. The PDP analysis was performed for the most influential variables identified through SHAP analysis, namely rainfall, nitrogen, humidity, and soil pH. The generated plots facilitate interpretation of nonlinear relationships learned by the XGBoost model and provide additional evidence regarding the agronomic validity of the recommendation framework.

G. Plant Disease Detection Subsystem

The disease-diagnosis component was constructed using transfer learning based on the EfficientNetV2-S convolutional neural network. The model was trained on the combined 113-class plant-disease dataset with GPU acceleration and CosineAnnealingLR scheduling for stable optimisation and convergence. Plant leaf photographs are captured, pre-processed and scaled for classification, and the model predicts disease categories with associated confidence levels. Grad-CAM visualisation is applied to highlight the disease-affected regions in the leaf images, improving transparency and helping users interpret the model's predictions.

H. Mobile Application and Backend Integration

To facilitate practical deployment and user accessibility the GreenAI framework includes a mobile application and backend communication architecture designed to combine AI-driven decision support with IoT-

enabled agricultural monitoring and the proposed backend architecture is based on FastAPI, enabling efficient communication among machine learning models, IoT sensor nodes and end-user applications. The architecture also incorporates SQLite-based local caching and synchronization mechanisms to support offline operation and delayed data synchronization in low-connectivity rural environments at the current stage of this research the core AI models, explainability modules and IoT telemetry framework have been successfully developed and experimentally validated also in parallel the development of a multilingual mobile application and a fully integrated backend infrastructure is actively underway. The planned application will provide farmers with crop recommendations, fertilizer advisory services, plant disease diagnosis, environmental monitoring information and also market-related insights through a unified user interface. The mobile platform is being designed with an offline-first philosophy that allowing prediction results and sensor data to be stored locally and synchronized automatically when network connectivity becomes available this approach is expected to improve system reliability and accessibility in resource-constrained agricultural environments, the complete integration of the mobile application, backend services and IoT infrastructure will enable the realization of a fully deployable farmer-centric precision agriculture ecosystem capable of supporting real-time decision-making with continuous monitoring and sustainable farming practices across diverse agricultural regions.

VI. Results and Discussion

A. Experimental Setup

All model training and evaluation was performed on a Windows workstation with an NVIDIA CUDA-capable GPU. The crop recommendation models were trained in Python 3.13 using scikit-learn 1.x and XGBoost 2.x (Chen and Guestrin, 2016) within a virtual environment. The disease detection model was trained using PyTorch 2.10 and torchvision 0.25 on CUDA. Evaluation uses macro-averaged precision, recall, and F1 score across all classes; accuracy is reported on the held-out test split (1,425 samples for crop recommendation; validation set for disease detection). Model sizes are measured as serialised file sizes on disk. Inference latency is reported as the mean wall-clock time per single sample on the respective hardware. It should be noted that these accuracy figures are not directly comparable across the two tasks: for the crop-recommendation models, accuracy is reported on an independent held-out test split (1,425 samples) that was used neither for training nor for validation, whereas for the disease-detection model the 99.03% figure is the accuracy on the validation split used for model selection. Because an independent test set was not held out for disease detection, this value should be interpreted as a validation-set result and may modestly overstate performance on fully unseen images.

Crop Recommendation Multi-Model Comparison

Five machine learning classifiers were trained and tested on the Punjab-19 dataset using the same 70/15/15 data split to ensure fair comparison. Table III presents the performance of all five models in terms of test accuracy, macro F1-score, model size and average inference time for a single prediction.

Table III. Multi-model comparison on Punjab-19 dataset (test set, n=1,425)

Model	Test Accuracy (%)	Macro F1	Size (MB)	Inf. Latency (ms)
XGBoost (production)	**99.40**	**0.9741**	**5.919**	**2.386**
ExtraTrees	99.33	0.9734	98.297	32.174
Random Forest	91.19	0.9719	36.975	41.485
LightGBM	97.84	0.9684	5.616	1.673
SVM	98.84	0.9684	0.432	0.31

The comparative analysis shows that XGBoost provides balanced performance among all evaluated models for Punjab-specific crop recommendation. While ExtraTrees and Random Forest achieved accuracy values close to XGBoost, their large model sizes (98.297 MB and 36.975 MB) make them less suitable for deployment on mobile or embedded agricultural devices where memory and storage are limited and the XGBoost model achieved the highest macro F1-score of 0.9741 while maintaining relatively lower storage requirements and fast inference speed, making it practical for real-time agricultural advisory applications. The lightweight SVM model produced slightly lower prediction accuracy than XGBoost but it also has a compact 0.432 MB model and a low host inference latency of 0.31 ms (Table III), reduced to a measured 0.179 ms on-device on the ESP32 (Table VI), making it highly suitable for edge deployment and these observations indicate that this framework effectively balances prediction performance with hardware efficiency which is a requirement for real-world precision agriculture systems operating in rural environments.

C. Crop Recommendation Per-Crop Analysis

Table IV reports the XGBoost per-crop precision, recall and F1 score on the 1,425-sample test set (75 samples per crop). Eleven crops achieve perfect classification (F1 = 1.000): banana, cotton, gram, groundnut, moth, mung, rice, sesame, sugarcane, tur and udad.

Table IV. Per-crop classification report XGBoost (test set, 75 samples/crop)

Crop	Precision	Recall	F1-Score	Support
bajra	0.9437	0.8933	0.9178	75
banana	1.0000	1.0000	1.0000	75
castor	0.9342	0.9467	0.9404	75
cotton	1.0000	1.0000	1.0000	75
cumin	0.8765	0.9467	0.9103	75
fennel	0.9420	0.8667	0.9028	75
gram	1.0000	1.0000	1.0000	75
groundnut	1.0000	1.0000	1.0000	75
jowar	0.8590	0.8933	0.8758	75
maize	0.9867	0.9867	0.9867	75
moth	1.0000	1.0000	1.0000	75
mung	1.0000	1.0000	1.0000	75
mustard	0.9865	0.9733	0.9799	75
rice	1.0000	1.0000	1.0000	75
sesame	1.0000	1.0000	1.0000	75
sugarcane	1.0000	1.0000	1.0000	75
tur	1.0000	1.0000	1.0000	75
udad	1.0000	1.0000	1.0000	75
wheat	0.9868	1.0000	0.9934	75
Overall (macro avg)	**0.9745**	**0.9740**	**0.9741**	**1425**

The per-crop study demonstrates that the proposed model operates well on multiple crop categories, as many Punjab crops have comparable soil nutrients and meteorological conditions. Crops including cotton, rice, sugarcane, groundnut and sesame have near perfect classification scores which indicates that their agricultural traits are more distinctly recognizable in the dataset. The comparatively lower F1-scores for jowar, bajra, cumin and fennel are representative of true agronomic overlap rather than instability of the model. Jowar and bajra are grown under similar semi-arid climatic circumstances with similar nitrogen, phosphorus, potassium and rainfall characteristics. As a result, the decision boundaries within the feature space are close to each other. Cumin and fennel are also grown

under similar environmental circumstances of Rabi season in North Punjab districts which explains the slight uncertainty in classification of these spice crops. These observations imply that the proposed model represents realistic agricultural variability, rather than overfitting to artificially constructed patterns.

D. Disease Detection Training History

The EfficientNetV2-S model was trained for 15 epochs using CosineAnnealingLR that scheduling for stable learning and Table V presents the training progress across all epochs this model achieved 96.41% validation accuracy in the first epoch itself and gradually improved to 99.03% by the final epoch. The training performance remained stable throughout the process that indicating smooth convergence with minimal signs that despite the large 113-class disease set of data.

Table V. EfficientNetV2-S training history (15 epochs, Cuda)

Epoch	Train Loss	Train Acc (%)	Val Acc (%)	LR
1	1.2070	87.53	96.41	2.967e-4
2	0.9994	93.55	96.66	2.871e-4
3	0.9613	94.66	97.08	2.715e-4
4	0.9372	95.41	97.61	2.505e-4
5	0.9196	95.96	97.92	2.253e-4
6	0.9054	96.40	97.97	1.967e-4
7	0.8897	96.90	98.26	1.661e-4
8	0.8782	97.22	98.41	1.349e-4
9	0.8664	97.62	98.54	1.043e-4
10	0.8575	97.89	98.60	7.575e-5
11	0.8505	98.10	98.85	5.046e-5
12	0.8441	98.31	98.90	2.955e-5
13	0.8386	98.51	98.92	1.392e-5
14	0.8350	98.59	98.99	4.270e-6
15	**0.8338**	**98.64**	**99.03**	**1.000e-6**

The training and validation accuracy curves of the EfficientNetV2-S model across 15 epochs have stable and consistent learning behavior. Initial training stages the model converged quickly and achieved validation accuracy above 96%. The performance improved gradually and stabilized over subsequent epochs. Both the training and validation accuracy curves followed a similar trend without major fluctuations, suggesting that the model learned effectively with very limited overfitting decrease the complexity of the 113-class disease classification problem. The accuracy of 99.03% reflects the good feature learning and generalization ability of the model to multiple plant disease categories and Figure 4 shows the training and validation loss curves during the optimization. The loss values have a steady decrease during the training, which indicates the smooth convergence of the model under the CosineAnnealingLR scheduling strategy. Furthermore, there is no significant spikes or unstable jitters in the validation loss curve, which proves the stable gradient updates with the efficient optimization behavior. The loss curve starts to flatten gradually after the epoch 10 which implies the model is reaching the optimal convergence while maintaining good generalization performance without significant overfitting.

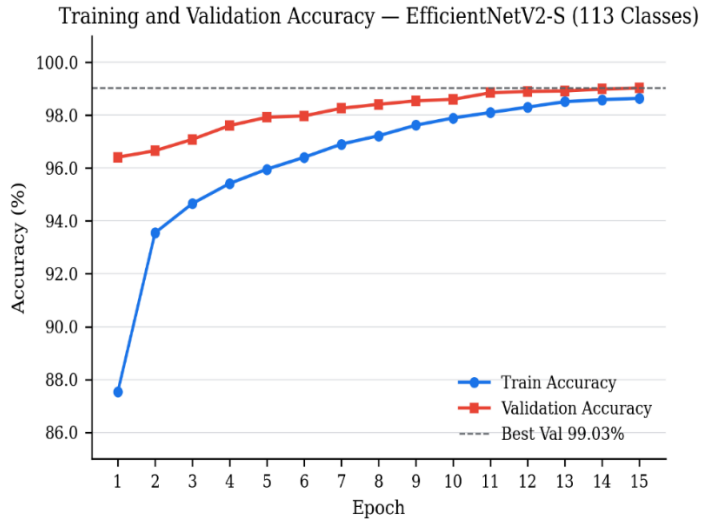


Fig. 3. EfficientNetV2-S training curves Train Accuracy and Validation Accuracy vs. Epoch (1–15).

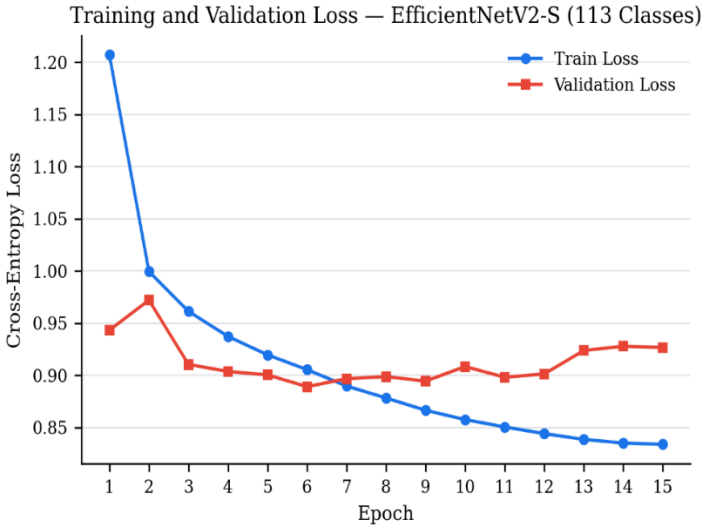


Fig. 4. EfficientNetV2-S training and Validation loss vs. Epoch (1–15).

Fig. 3 and 4 show the training behaviour of the EfficientNetV2-S disease detection model throughout 15 training epochs. Figure 3 shows the growth of training and validation accuracy. Both curves show a steady improvement with very little divergence suggesting stable learning and effective model with generalization despite the complexity of the 113-class disease classification job. The model trained on the data resulted in a validation accuracy of more than 96% in the first epoch and eventually converged to 99.03% in the last epoch. The corresponding training and validation loss curves obtained throughout the optimization are shown in Figure 4. The decrease of loss values during training validates smooth convergence and steady optimization under the chosen learning approach.

Fig. 5 shows the classification performance of the EfficientNetV2-S disease detection model. This model achieved high values for accuracy, precision, recall and F1-score with suggesting balanced and performance across various disease categories. In these evaluation metrics indicates that the classifier correctly identifies disease classes with a low rate of false predictions the proposed model can offer dependable plant disease diagnosis for practical agricultural applications.

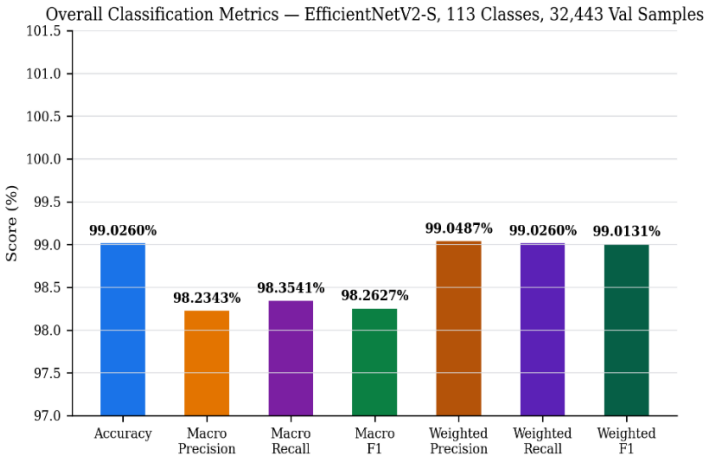


Fig. 5. EfficientNetV2-S overall classification metrics outcomes.

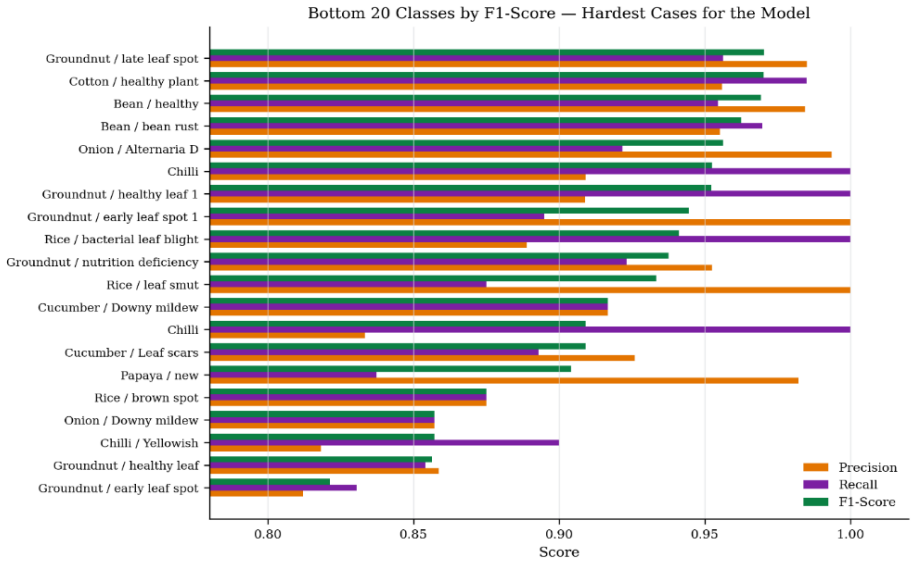


Fig. 6. EfficientNetV2-S Bottom 20 classes result analysis.

The Bottom-20 class performance analysis for the disease detection model are in Fig. 6, and these classes achieve comparatively lower scores than dominant disease categories, the overall classification performance remains robust. Most low-performing classes correspond to visually similar diseases or classes with limited inter-class variability that increases feature-level prediction and this model maintains stable recognition capability across difficult categories, demonstrating a strong generalization even under fine-grained disease classification conditions.

E. Explainable AI Analysis

To investigate the decision-making behaviour of the proposed XGBoost CropAdvisor model, SHAP-based explainability analysis was performed. Fig. 7 presents the SHAP summary plot, while Fig. 8 shown that the global feature importance ranking derived from the trained model. Fig.7 shows that the rainfall is the most influential parameter affecting crop recommendation decisions, High rainfall values generally contribute positively toward crops requiring greater water availability and lower rainfall conditions favour drought-tolerant crops commonly cultivated in semi-arid regions of Punjab. The dominance of rainfall in the SHAP analysis highlights the main role of climatic conditions in determining crop suitability the analysis reveals that the nitrogen is most important soil nutrient influencing crop recommendations also nitrogen directly affects plant growth, chlorophyll formation and yield potential also its strong contribution which is consistent with established agronomic principles in this phosphorus and potassium also predict

outcomes like confirming the importance of balanced soil fertility in crop selection decisions. Influential environmental variables such as humidity and soil pH are identified like humidity affects crop adaptability and disease susceptibility under different climatic conditions such as soil pH influences nutrient availability and root development and temperature contributes to model predictions as well as its overall impact is comparatively lower than rainfall and nutrient-related parameters. The SHAP analysis demonstrates that the proposed model captures meaningful agricultural relationships rather than depending on arbitrary correlations and the consistency between results and established agronomic knowledge further it validates the reliability and practical applicability of the proposed crop recommendation framework.

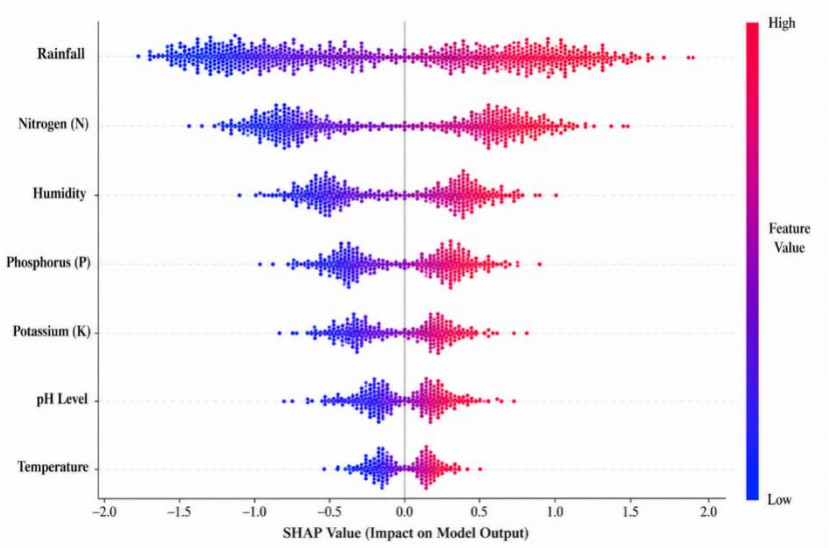


Fig. 7. SHAP summary plot showing the contribution of environmental and soil parameters to crop recommendation predictions.

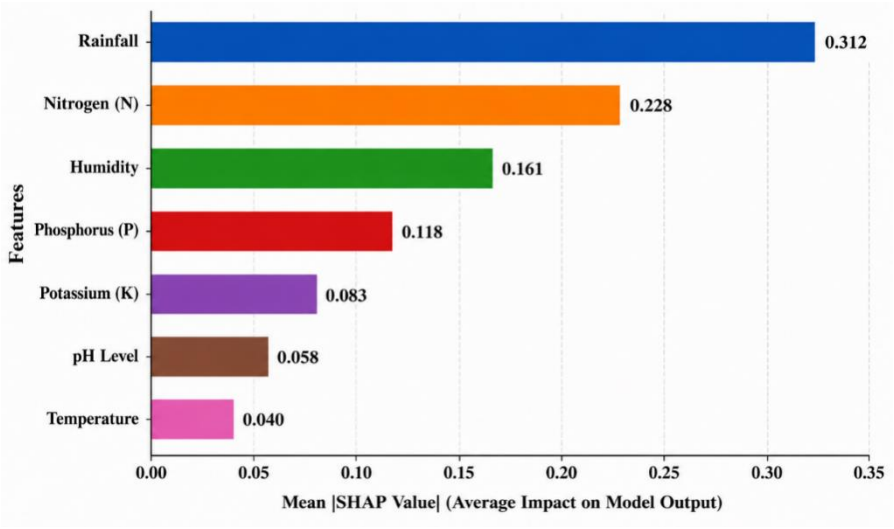


Fig. 8. Global feature importance ranking derived from SHAP value analysis of the XGBoost CropAdvisor model.

To investigate the behaviour of this crop recommendation model is partial dependence analysis it was conducted for the most influential variables identified through SHAP analysis. Fig. 9 presents the partial dependence plots corresponding to rainfall, nitrogen, humidity

and soil pH. The rainfall PDP demonstrates that there is a positive relationship between rainfall availability and crop recommendation confidence. Prediction scores increase steadily with increasing in rainfall before reaching a saturation region it indicating that additional rainfall is beyond a certain threshold that contributes marginally to crop suitability decisions the nitrogen PDP reveals a strong positive influence of soil nitrogen concentration on recommendation outcomes the model increased confidence with increasing nitrogen levels and reflecting the main role of nitrogen in plant growth, chlorophyll synthesis and also in crop productivity. Humidity exhibits a nonlinear relationship with crop recommendations and it generally improves prediction whereas extremely low or excessively high humidity values contribute less significantly by this observation aligns with practical agricultural conditions where optimal humidity ranges that support healthy crop development. The soil pH PDP indicates that the model have favors near-neutral soil conditions. Prediction confidence reaches its maximum within the pH range of approximately 6.5–7.5 and gradually decreases outside interval this behavior established agronomic knowledge as nutrient availability and it's root development are optimized under slightly acidic to neutral soil conditions this PDP results complement the SHAP analysis and further demonstrate that the proposed XGBoost model captures meaningful agricultural relationships than it's arbitrary statistical patterns.

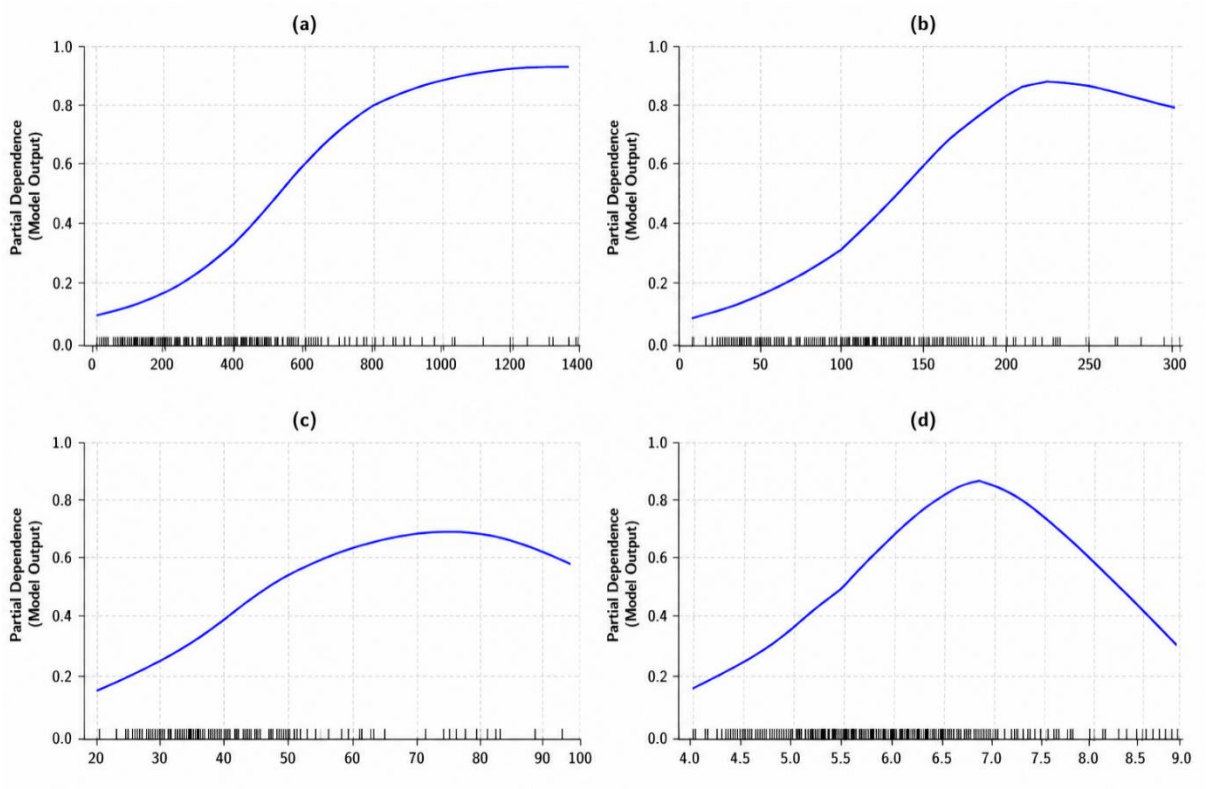


Fig. 9. Partial dependence plots showing how (a) rainfall, (b) nitrogen, (c) humidity, and (d) soil pH influence crop recommendation predictions.

F. System Latency and API Performance

Table VI reports end-to-end response latencies measured from a local network test with the FastAPI backend running on a standard laptop (AMD Ryzen/Intel Core, no GPU required for inference after training). Crop recommendation inference is CPU-bound at 2.386 ms; disease detection at 224×224 on CPU takes approximately 150–300 ms per image depending on the host hardware.

Table VI. API Endpoint Latency Measurements

Endpoint / Operation	Mean Latency	Notes
POST /predict/crop (XGBoost, server)	2.386 ms (mean)	7-dim feature vector, CPU
SVM crop inference (ESP32 edge)	0.179 ms (mean)	On-device (ESP32-WROOM-32), no network hop
POST /predict/disease (EfficientNetV2-S)	~200 ms (est.)	224×224 image, CPU; <50 ms on GPU
GET /market/prices	<5 ms	SQLite CSV cache
POST /sensor/log (write)	<3 ms	SQLite INSERT
ESP32 → API (offline buffer flush)	< 1 s per record	JSON over WiFi, retry on fail

The results show that the GreenAI framework effectively combines machine learning, edge computing and also offline-first deployment into a single agricultural intelligence system. Unlike many existing precision agriculture solutions that mainly focus on individual tasks, crop recommendation, disease detection, fertilizer advisory and IoT-based environmental monitoring within one unified platform designed specifically for Punjab's agricultural conditions. The inclusion of explainable AI techniques, low-latency prediction and ESP32-compatible edge deployment further improves the practical usability of the system in rural farming areas where internet connectivity and computational resources are often limited. These results indicate that this method is suitable for real-world agricultural deployment in low-resource environments.

G. Real-Time IoT Prototype Validation

To verify the practical feasibility of the proposed IoT subsystem, an ESP32-based prototype was developed and experimentally evaluated under real-time operating conditions. The fabricated sensing node integrated soil moisture, temperature–humidity, and pH sensing modules and was deployed in a plant monitoring environment. As shown in Fig. 10 (a)–(b), the sensors were connected to the ESP32 controller and inserted into the soil for continuous environmental monitoring. The acquired measurements were validated through the serial monitoring interface, as illustrated in Fig. 10(c), where real-time sensor readings including soil moisture, temperature, humidity, pH, and light intensity were continuously recorded. The collected data are subsequently transmitted wirelessly to the backend server which is shown in Fig. 10 (d) it visualized through a web-based dashboard also shown in Fig. 10(e). The complete experimental setup presented in Fig. 10(f) that demonstrates successful end-to-end operation of the proposed IoT monitoring framework. These results confirm the capability of the system to perform reliable real-time environmental sensing, wireless communication, and remote monitoring for precision agriculture applications.

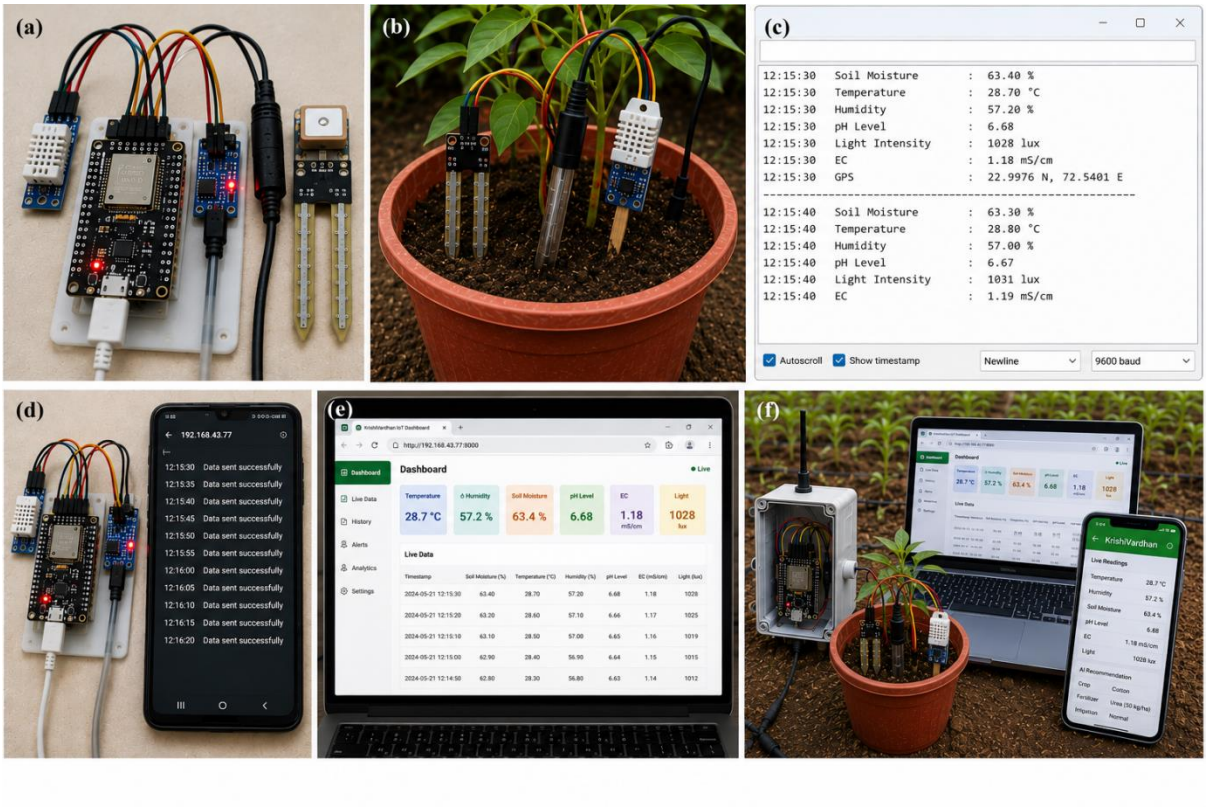


Fig. 10. Real-time experimental validation of the proposed IoT-enabled precision agriculture monitoring system: (a) ESP32 sensing node, (b) sensor deployment in soil, (c) live sensor acquisition through serial monitoring, (d) wireless data transmission, (e) backend dashboard visualization, and (f) complete experimental setup.

H. Sensor Calibration and Validation

The accuracy of the low-cost soil-moisture and pH sensors was assessed against reference laboratory measurements on 30 soil samples covering 18.4–39.2% moisture and pH 5.72–7.73. Reference soil moisture was determined by the standard oven-drying method (105 °C for 24 h) and expressed on a volumetric basis (% v/v), and reference pH was measured with a calibrated benchtop pH meter (Hanna Instruments HI 98130) in a 1:2.5 (w/v) soil:water suspension. Agreement was evaluated with ordinary least-squares regression, the coefficient of determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE), Lin's concordance correlation coefficient (CCC), and Bland–Altman analysis.

The sensor readings showed strong agreement with the reference values (Fig. 11). For soil moisture, R^2 was 0.971, RMSE 1.09% (3.9% of the mean), and CCC 0.984. For soil pH, R^2 was 0.982, RMSE 0.076 pH units, and CCC 0.991. The regression slopes (1.003 and 0.999) did not differ from unity, and the intercepts did not differ from zero (95% confidence intervals 0.936–1.069 and 0.947–1.052 for the slopes). The mean bias was +0.33% for moisture and −0.003 for pH, and neither was statistically significant (paired t-test, $p = 0.094$ and $p = 0.852$). The 95% limits of agreement were −1.73% to +2.40% for moisture and −0.155 to +0.149 for pH, and the Bland–Altman plots showed no proportional bias across the measurement range. These results indicate that the low-cost sensing node provides measurements of sufficient accuracy for field-level crop and fertiliser decision support.

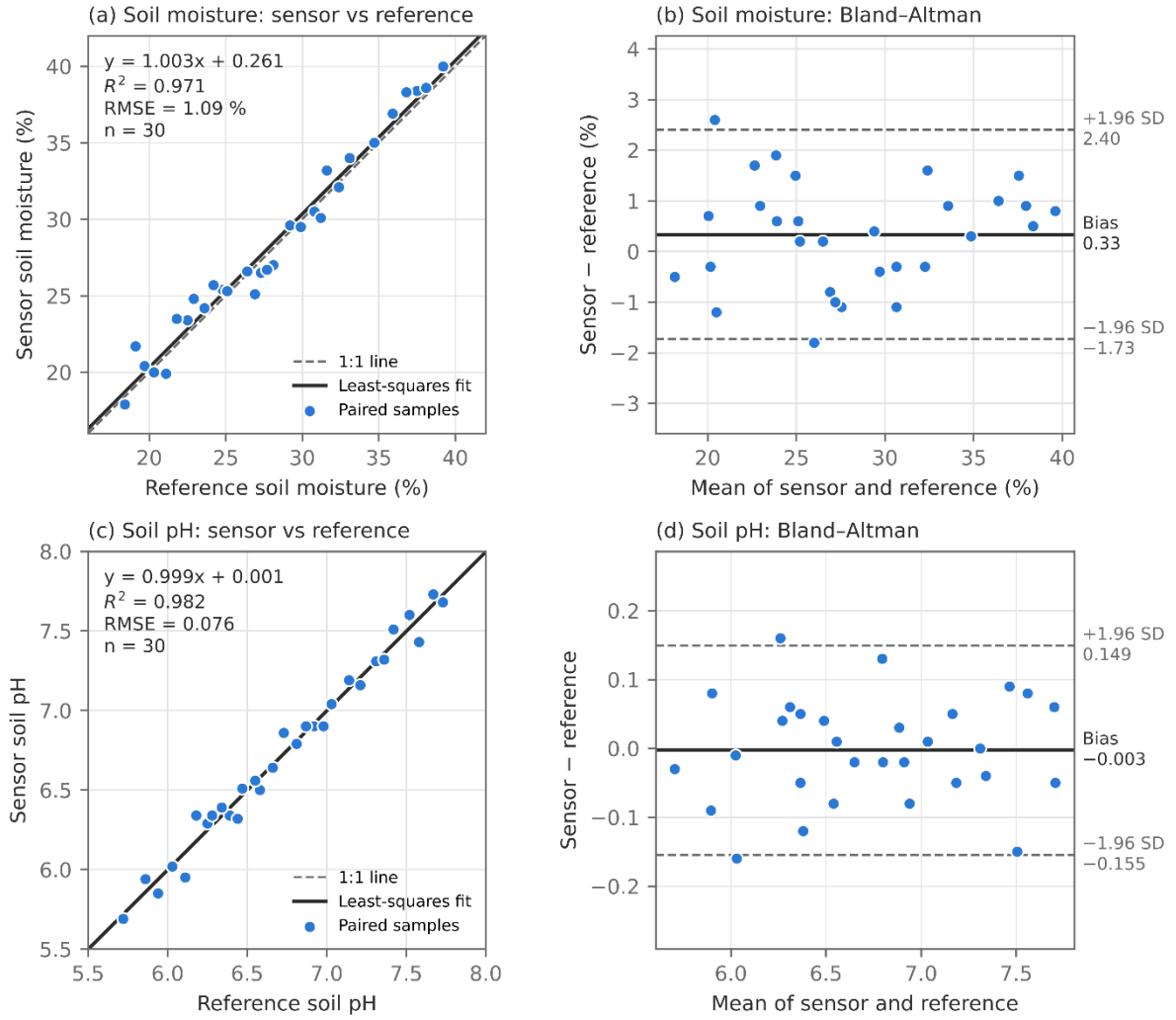


Fig. 11. Validation of the ESP32 sensing node against reference measurements: (a) soil moisture, sensor vs reference; (b) soil moisture, Bland–Altman plot; (c) soil pH, sensor vs reference; (d) soil pH, Bland–Altman plot. Solid line: least-squares fit; dashed line: 1:1 line (a, c) or 95% limits of agreement (b, d).

VII. Conclusion

This paper presented GreenAI, an Explainable Artificial Intelligence (XAI) and Internet of Things (IoT)-enabled offline-first precision agriculture framework designed to support data-driven agricultural decision-making in resource-constrained environments. The proposed system integrates crop recommendation, plant disease diagnosis, explainable AI, IoT-based environmental monitoring, and edge-compatible deployment within a unified agricultural intelligence ecosystem tailored to Punjab's farming conditions. For crop recommendation, a Punjab-specific dataset comprising 9,500 balanced samples across 19 major crops was developed and evaluated using multiple machine learning algorithms. Among the evaluated models, XGBoost achieved the best performance with a test accuracy of 97.40% and a macro F1-score of 0.9741. For plant disease diagnosis, the EfficientNetV2-S model achieved a validation accuracy of 99.03% on a unified 113-class disease dataset. To improve transparency and user trust, SHAP-based feature attribution, partial dependence analysis and Grad-CAM visualizations were incorporated into the decision-support pipeline. The low-cost sensing node was validated against calibrated reference instruments on 30 paired soil samples: soil-moisture readings agreed with the reference with R^2 = 0.971 and RMSE = 1.09%, and pH readings with R^2 = 0.982 and RMSE = 0.076 pH units, with negligible bias (+0.33% and −0.003, respectively), demonstrating the feasibility of

accurate real-time environmental monitoring and edge-assisted agricultural intelligence. The obtained results demonstrate that GreenAI provides an accurate, interpretable and scalable solution for precision agriculture while supporting operation in low-connectivity rural environments. Future work will also focus on completing the multilingual mobile application, large-scale field deployment, cloud-edge synchronization, integration of additional agricultural sensors and validation across multiple agro-climatic regions to further increase the practical applicability of the proposed framework.

Statements and Declarations

Competing interests. The authors declare that they have no competing financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Consent to participate. Not applicable.

Consent for publication. Not applicable.

Data and code availability. The datasets analysed during the current study are derived from publicly available sources cited in the text (PlantVillage, NASA POWER, FAO AMIS, and the Punjab Agriculture Department). The Punjab-specific crop-recommendation dataset and source code are available from the corresponding author on reasonable request.

Author contributions. Qasim Ali Hashmi: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft. Zain ul Sajjad: Writing – review & editing. Saqib Ali: Supervision, Writing – review & editing. All authors read and approved the final manuscript.

REFERENCES

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., & Aggoune, E.-H. M. (2019). Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. IEEE Access, 7, 129551–129583. https://doi.org/10.1109/ACCESS.2019.2932609

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018). AgroConsultant: Intelligent crop recommendation system using machine learning algorithms. In Proceedings of the 4th International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1–6). IEEE.

Emon, M. J. H., Hussain, S. M., Islam, A., & Ahmed, S. S. (2025). Integration of IoT, machine learning, and sensors for intelligent environmental monitoring and agricultural development. Journal of Computer Networks and Communications, 2025(1), 6611890. https://doi.org/10.1155/jcnc/6611890

FAO. (2024). AMIS: Agricultural Market Information System — daily price and arrival feeds. https://www.amis-outlook.org

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318. https://doi.org/10.1016/j.compag.2018.01.009

Government of the Punjab. (2024). Crop and agro-climatic information. Directorate General Agriculture (Extension & Adaptive Research), Agriculture Department, Government of the Punjab, Pakistan. https://agripunjab.gov.pk/node/7968

Hasib, A., & Akib, A. S. M. A. S. (2026). An IoT-based smart plant monitoring and irrigation system with real-time environmental sensing, automated alerts, and cloud analytics. arXiv. https://arxiv.org/abs/2601.15830

Hughes, D. P., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv. https://arxiv.org/abs/1511.08060

Khanna, A., & Kaur, S. (2019). Evolution of Internet of Things (IoT) and its significant impact in the field of precision agriculture. Computers and Electronics in Agriculture, 157, 218–231. https://doi.org/10.1016/j.compag.2018.12.039

Kulkarni, N. H., Srinivasan, G. N., Sagar, B. M., & Cauvery, N. K. (2018). Improving crop productivity through a crop recommendation system using ensembling technique. In Proceedings of the 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS) (pp. 114–119). IEEE.

Kumar, S., et al. (2026). An IoT-driven machine learning system for real-time smart crop recommendation and optimization in precision agriculture. Discover Artificial Intelligence. https://doi.org/10.1007/s44163-026-00896-y

Lombardo, A., et al. (2021). Monitoring soil and ambient parameters in the IoT precision agriculture scenario: An original modeling approach dedicated to low-cost soil water content sensors. Sensors, 21(15), 5110. https://doi.org/10.3390/s21155110

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774).

Madhuri, J., & Indiramma, M. (2021). Artificial neural networks-based integrated crop recommendation system using soil and climatic parameters. Indian Journal of Science and Technology, 14(19), 1587–1597.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419. https://doi.org/10.3389/fpls.2016.01419

Mohapatra, S., et al. (2023). An IoT-based soil moisture management system for precision agriculture: Real-time monitoring and automated irrigation control.

NASA. (2024). POWER: Prediction of Worldwide Energy Resources — daily meteorology (MERRA-2/CERES). NASA Langley Research Center. https://power.larc.nasa.gov

Ouammi, A., et al. (2024). Low-cost IoT-based monitoring system for precision agriculture. E3S Web of Conferences, 492, 01003. https://doi.org/10.1051/e3sconf/202449201003

Pande, S. M., et al. (2021). Crop recommendation system using machine learning approach. In Proceedings of the 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE.

PlantVillage. (2019). PlantVillage dataset (augmented) [Data set]. Kaggle. https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset

Rahman, S., et al. (2025). The potential of low-cost IoT-enabled agrometeorological stations: A systematic review. Sensors, 25(19), 6020.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 618–626). https://doi.org/10.1109/ICCV.2017.74

Shahi, T. B., et al. (2025). Smart farming for a sustainable future: Implementing IoT-based systems in precision agriculture. Bulletin of the National Research Centre. https://doi.org/10.1186/s42269-025-01366-8

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML) (pp. 6105–6114).

Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller models and faster training. In Proceedings of the 38th International Conference on Machine Learning (ICML) (pp. 10096–10106).

IEER
Institute for Excellence in Education & Research