Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

FarmGPT: An Explainable AI-Powered Smart Agriculture System for Crop Recommendation and Decision Support

Arun T[1], Poojasri M[2]

[1,2]Department of Computer Science and Engineering, Ramco Institute of Technology, Rajapalayam, Tamil Nadu, India

Abstract

Agriculture remains a critical pillar of the global economy, yet farmers face persistent challenges in optimal crop selection due to dynamic environmental conditions and a lack of interpretable guidance. While machine learning (ML) models offer high predictive accuracy, their black-box nature limits adoption among non-technical users. This paper presents FarmGPT, an explainable AI-based crop recommendation and agricultural decision support system. FarmGPT employs a Random Forest classifier achieving 98.2% accuracy, trained on agro-meteorological parameters including nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, soil pH, and rainfall. SHAP (SHapley Additive exPlanations) is integrated to provide transparent, feature-level justifications for each recommendation. The system further incorporates a CNN-based plant disease detection module trained on the PlantVillage dataset, a regression-based yield prediction engine, real-time weather integration via OpenWeatherMap API, and an INR profit estimation module. Built on a full-stack Flask–Node.js–React architecture, FarmGPT delivers comprehensive, real-time agricultural intelligence through a conversational interface. Evaluation results confirm 98.2% crop recommendation accuracy, 90% disease detection accuracy in user trials, and a System Usability Scale (SUS) score of 79.3.

Keywords — Explainable AI, Crop Recommendation, Random Forest, SHAP, CNN, Plant Disease Detection, Precision Agriculture, Smart Farming.

1. Introduction

Agriculture is a fundamental pillar of the global economy, providing food security and livelihoods for a significant portion of the world’s population. Farmers face considerable challenges in selecting suitable crops, as traditional methods rely on experience and intuition, often failing to account for complex soil-climate interactions.

Machine Learning (ML) has emerged as a powerful tool for predictive analytics in agriculture, capable of analyzing large volumes of agro-meteorological data. However, most models operate as ‘black boxes,’ offering predictions without explaining the reasoning behind them. This opacity creates a trust gap that hinders real-world adoption.

To address this, we present FarmGPT—an intelligent, explainable, and integrated agricultural decision support system that combines crop recommendation, plant disease detection, yield prediction, and financial analysis under a unified conversational interface. The key contributions of this paper are:

- A Random Forest-based crop recommendation engine achieving 98.2% accuracy.
- Integration of SHAP for transparent, per-prediction feature attribution.
- A CNN-based plant disease detection module using the PlantVillage dataset.
- Real-time weather API integration and INR-based profit estimation.
- Full-stack deployment with a React-based conversational user interface.

2. Related Work

Afzal et al. [1] demonstrated that integrating soil parameters (N, P, K) with supervised ML models significantly improves crop selection accuracy. Sam and D'Abreo [2] emphasized the importance of preprocessing and feature selection in achieving reliable crop recommendations. Sawant et al. [3] extended this by fusing IoT sensor data with ML models for real-time precision agriculture.

On explainability, Adadi and Berrada [4] surveyed XAI methods and highlighted interpretability as critical for non-technical user adoption. Lundberg and Lee [5] introduced SHAP based on cooperative game theory, providing consistent and locally accurate feature attributions. The TreeExplainer variant [6] enables efficient SHAP computation for ensemble models such as Random Forest.

For disease detection, Mohanty et al. [7] established deep CNNs as effective tools for classifying plant diseases from leaf images using the PlantVillage dataset, though noting degraded performance under real-world field conditions.

3. System Architecture

A. Full-Stack Architecture

FarmGPT is implemented using a three-tier full-stack architecture. Flask (Python) serves as the AI inference backend, hosting all ML and DL models. Node.js with Express.js manages API orchestration and routing between the frontend and AI services. React.js provides the frontend conversational interface. MongoDB handles user session and history storage.

B. AI Microservices Pipeline

Each AI task—crop recommendation, disease detection, and yield prediction—is deployed as an independent microservice. Input data flows through a preprocessing module (Min-Max normalization for soil/weather parameters; resize, normalize, and augment for images), then through the respective AI model, and finally through the XAI explanation layer before returning to the user.

4. Methodology

A. Dataset and Features

The crop recommendation model is trained on a structured agricultural dataset with 22 crop classes and over 40,000 samples. Seven features are used: N (0–140 mg/kg), P (5–145 mg/kg), K (5–205 mg/kg), temperature (8.8–43.6°C), humidity (14.2–99.9%), pH (3.5–9.9), and rainfall (20.2–298.5 mm). All features are normalized using Min-Max scaling prior to training.

B. Random Forest Classifier

A Random Forest classifier is employed for its robustness to non-linear relationships, noise tolerance, and built-in feature importance estimation. The ensemble is configured with n_estimators=100, criterion=gini, random_state=42, n_jobs=−1 (parallel), and oob_score=True. Bootstrap sampling and majority voting across trees reduce overfitting and improve prediction stability.

C. SHAP-Based Explainability

SHAP TreeExplainer computes exact Shapley values for each prediction, quantifying the signed contribution of every input feature to the model output. Feature contributions are rendered as force plots and bar charts, enabling farmers to understand why a specific crop was recommended (e.g., ‘High rainfall (+0.18) and nitrogen (+0.12) are the primary drivers for rice’).

D. CNN Disease Detection

A sequential CNN processes 224×224 RGB leaf images through three convolutional blocks (32, 64, 128 filters; 3×3 kernels; ReLU activation), interleaved with 2×2 max-pooling layers. A fully connected layer (128 units), 50% dropout, and a softmax output layer (6 disease classes) complete the architecture. The model is trained on the PlantVillage dataset with data augmentation (rotation, flipping, zoom, shift) to improve generalization.

E. Yield Prediction and Profit Analysis

Regression models estimate yield (tonnes/acre) from soil and weather inputs. The financial module computes gross revenue (yield × market price per kg), total cost (area × unit cost), net profit, and ROI using backend market data, providing per-crop financial projections in INR.

5. Results and Discussion

A. Crop Recommendation Performance

The Random Forest model was evaluated on an 80-20 train-test split and validated using 10-fold cross-validation (mean accuracy: 98.1%, std: 0.4%). Bootstrap confidence intervals (1,000 samples) yield a 95% CI of [97.8%, 98.5%]. McNemar’s test confirms statistical significance over Naive Bayes, KNN, and Logistic Regression ($p < 0.05$ for all). Performance metrics are summarized in Table I.

Table I. Crop Recommendation Model Performance Metrics

Metric	Value
Accuracy	98.2%
Precision	97.8%
Recall	97.5%
F1-Score	97.6%
Training Accuracy	98.6%
Validation Loss	0.048

B. SHAP Interpretation Analysis

SHAP analysis confirms that rainfall and nitrogen are the top contributors for water-intensive crops (e.g., rice), while pH and temperature dominate recommendations for drier crops (e.g., wheat). The SHAP force plots provide per-prediction transparency aligned with agronomic domain knowledge.

C. Disease Detection Evaluation

In controlled user trials with 20 labeled leaf images, the CNN model achieved 90% accuracy (18/20). The two misclassifications occurred under poor lighting, consistent with known limitations of image-based classifiers under real-world variability. Real-time inference latency was under 4 seconds, compared to 5–10 day turnaround for laboratory sample analysis.

D. Yield Prediction

The yield regression model achieved MAE of 2.1 t/ha, RMSE of 2.8 t/ha, and R^2 of 0.96 on 850 test samples, demonstrating strong predictive alignment with actual yield values.

E. Comparative Benchmarking

FarmGPT was benchmarked against AgroAdvisor (Decision Tree, 91.5%), SmartFarm AI (SVM, 94.8%), and CropPredict Pro (Neural Network, 96.1%) on the identical dataset and train-test split. FarmGPT achieved the highest accuracy (98.2%) and the highest user trust score (4.3/5 Likert scale), compared to 2.8/5 for the system with no explainability.

F. System Performance and Scalability

Load testing with Apache JMeter at 100 concurrent users maintained crop recommendation response times below 3 seconds. Redis caching for weather data and frequent SHAP queries reduced average response times by 40%. Horizontal scaling across two Flask instances with Nginx load balancing reduced peak response times by 48%. A TensorFlow Lite compressed model (2.3 MB) enables offline mobile deployment for rural areas.

G. User Acceptance Study

A pilot study with 30 participants (farmers, agricultural students, and extension officers) achieved a SUS score of 79.3 (‘Good’ usability). 87% of participants found SHAP-based explanations helpful in understanding recommendations, and farmers specifically noted increased confidence when specific contributing factors were displayed alongside recommendations.

6. Conclusion

This paper presents FarmGPT, a transparent and accurate AI-powered agricultural decision support system. By combining a 98.2%-accurate Random Forest crop recommendation engine with SHAP-based interpretability, CNN-based disease detection, yield prediction, and profit estimation, FarmGPT bridges the gap between high-performance ML and practical farmer usability. The system’s explainability layer demonstrably improves user trust, while its full-stack architecture ensures real-time, scalable deployment.

Future work will incorporate IoT sensor integration, multilingual support, real-time mandi price feeds, and enhanced offline capabilities for low-connectivity rural environments.

7. Acknowledgment

The authors gratefully acknowledge the support of the Department of Computer Science and Engineering, Ramco Institute of Technology, Rajapalayam, and thank the farming community participants who contributed to the user acceptance study.

References

1. H. Afzal et al., “Incorporating soil information with machine learning for crop recommendation,” Scientific Reports, 2025.
2. S. Sam and S. D’Abreo, “Crop recommendation using machine learning,” arXiv preprint, 2025.
3. N. R. Sawant et al., “IoT-driven ML system for smart crop recommendation,” Springer, 2026.
4. A. Adadi and M. Berrada, “Peeking inside the black box: A survey on explainable artificial intelligence (XAI),” IEEE Access, vol. 6, pp. 52138–52160, 2018.
5. S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” Proc. NeurIPS, pp. 4765–4774, 2017.
6. S. M. Lundberg et al., “From local explanations to global understanding with explainable AI for trees,” Nature Machine Intelligence, vol. 2, no. 1, pp. 56–67, 2020.
7. S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Frontiers in Plant Science, vol. 7, p. 1419, 2016.