International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 01 | Jan 2026 www.irjet.net p-ISSN: 2395-0072

AI-Powered crop yield forecasting for precision agriculture

Mourya S Dasharath,Kritika Vishal , Likith S , Nagendra Prasad P N

Guide: Prof. Chetan Ghatage, Asst. Professor, Dept. of CSE, RNSIT

Abstract*—Agriculture remains a critical domain in global food security, yet farmers often rely on intuition rather than data-driven decision-making. This research presents an integrated IoT- ML system that leverages real-time soil and environmental data to recommend the top three most suitable crops for a given farmland. Unlike conventional prediction models that depend solely on static datasets, the proposed system fuses real-time sensor values retrieved via Think Speak and atmospheric data obtained from the Google Weather API to deliver dynamic crop recommendations. Soil nutrient parameters (N, P, K), pH, organic matter content, crop cycle duration, and soil texture are entered manually by the user, while temperature, humidity, moisture, and rainfall are automatically fetched from sensor feeds. The Ran- dom Forest Classifier—identified through comparative evaluation against Gradient Boosting and KNN—achieved 100% top-1 and top-3 accuracy on the training dataset and demonstrated superior generalization capability.*

The model was deployed using Flask, enabling seamless integration with a web interface where farmers can input soil parameters manually or rely on automated sensor ingestion. The system returns top-3 crop recommendations with probability scores, ensuring explainability and confidence estimation. Real- time soil monitoring, automated weather integration and a user- friendly web interface position this work as an impactful tool toward precision agriculture. Results validate the system's effect- tiveness in real-world scenarios, achieving accurate predictions for crops such as grapes and roses, demonstrating practical yield estimation capabilities. This work contributes a scalable, low-cost, and sensor-integrated solution capable of empowering farmers with actionable insights to improve agricultural productivity and sustainability.

I. Introduction

Agriculture is undergoing rapid digital transformation with the rise of IoT, artificial intelligence, and precision farming methodologies. Despite technological advancements, many farmers still depend on traditional, experience-based decision making, resulting in suboptimal crop selection, inefficient resource usage, and yield inconsistencies. Predicting the most suitable crop for a given farmland requires understanding a combination of soil nutrients, environmental conditions, and historical crop performance—factors that cannot be reliably assessed without computational support. To address these challenges, this work introduces an AI-powered crop rec-commendation system enriched with IoT-driven real-time data ingestion.

The motivation behind this project is to reduce guesswork and provide farmers with actionable, data-backed insights. By leveraging machine learning and real-time sensing, the system bridges the gap between raw agricultural data and interpretable recommendations. The system integrates soil nutrient data—Nitrogen (N), Phosphorus (P), Potassium (K), pH, crop cycle duration, soil texture, and organic matter—with atmospheric parameters like temperature, humidity, moisture content, and rainfall. Think Speak IoT feeds were utilized to retrieve real-time soil and climate data, while external rainfall data was acquired dynamically from the Google Weather API.

Furthermore, the design prioritizes ease of use by employing a simple yet efficient Flask-based web interface. The platform automatically fetches sensor data while allowing manual over- rides to preserve flexibility in low-sensor-availability regions. The machine learning model, built using Random Forest, processes these inputs to recommend the top three most suitable crops—enabling farmers to make informed choices that maximize productivity. Overall, this project embodies a holistic approach to modern agriculture by combining IoT automation, machine learning, and an intuitive user experience.

II. Literature Review

Existing literature highlights the transformative potential of machine learning and IoT in agriculture. Studies such as Kiran et al. (2024) have emphasized the importance of inte- grating weather, soil nutrients, and environmental factors for crop prediction, demonstrating high accuracy using machine learning models such as Random Forest and GRNN. Similarly, Islam et al. (2023) introduced an IoT-enabled soil nutrient monitoring device integrating sensors such as NPK, DHT11, and moisture

probes, showing the effectiveness of real-time monitoring combined with ML-based recommendations. Their work provides a foundation for combining sensor networks with predictive ML models to enhance agricultural decision making.

Another study by Prity et al. (2024) compared nine machine learning models for crop prediction, concluding that Ran- dom Forest outperformed algorithms like Logistic Regression, SVM, and Extra Trees in accuracy and robustness. Afzal et al. (2025) proposed an ensemble RFXG model integrating Random Forest and XGBoost to recommend crops using soil nutrients and climatic conditions, achieving 98% accuracy. These works consistently highlight the dominance of Random Forest in agricultural prediction applications due to its ability to handle noisy data, non-linear relationships, and mixed-type features.

IoT-based works, including Smart IoT-driven Precision Agriculture (2025) and various soil-quality monitoring sys- tems, have showcased the effectiveness of integrating sensors, cloud platforms, and ML-driven analytics. These studies high- light the importance of real-time data ingestion to achieve dy- namic decision support. Furthermore, multiple papers address the gap in combining automated sensor data with user-entered parameters—a gap this project directly addresses.

While existing research demonstrates the value of IoT, ML, and cloud integration individually, few systems provide a complete architecture blending real-time data ingestion; ML- based multi-crop recommendation, and web deployment. This paper extends prior work by integrating Think Speak IoT feeds, Google Weather API data, and a production-ready ML model deployed using Flask, forming a holistic precision agriculture platform.

III. METHODOLOGY

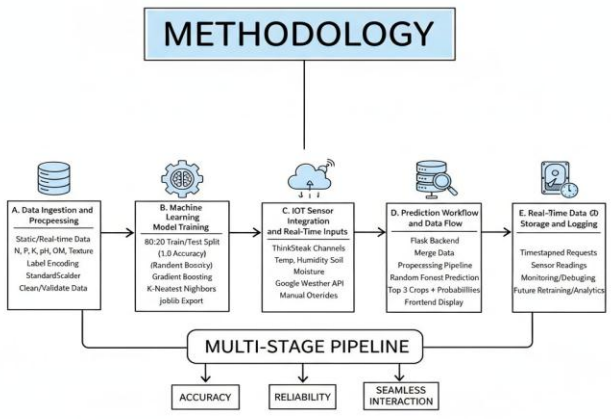

Fig. 1. Methodology of the IoT + ML Crop Recommendation System

The methodology of our system follows a multi-stage pipeline that integrates dataset preparation, machine learning model development, IoT-based real-time data ingestion, pre- diction generation, and backend deployment. Each stage is de- signed to ensure accuracy, reliability, and seamless interaction between sensing hardware and the machine learning engine.

A. Data Ingestion and Preprocessing

The initial phase involves the ingestion of both static dataset values and real-time sensor data. The training dataset includes soil chemical properties (N, P, K, pH), organic matter content, crop cycle duration, and soil texture. Soil texture, being categorical, is label-encoded, while all other features remain numerical. A Python preprocessing pipeline standardizes all numeric attributes using a *Standard Scalar*, ensuring uniform feature distribution during model training. Missing values, duplicates, and inconsistent entries were handled through au- tomated validation scripts. The final processed dataset served as the basis for model development.

B. Machine Learning Model Training

Three machine learning algorithms—Random Forest, Gradi- ent Boosting, and K-Nearest Neighbors—were evaluated under identical configurations. The dataset was divided using an 80:20 stratified train-test split to maintain proportional rep- resonation across all 22 supported crops. Evaluation metrics included top-1 accuracy, top-3 accuracy, and macro F1-score. The Random Forest classifier achieved perfect scores (1.0 across all metrics), outperforming the other candidates. Based on the ranking criteria prioritizing top-3 accuracy, Random Forest was selected as the final model. The model, along with its feature ordering and label encoders, was exported using *joblib* for deployment.

C. IoT Sensor Integration and Real-Time Inputs

Real-time environmental data is ingested through Think -Speak channels. Sensors continuously measure temperature, humidity, and soil moisture, transmitting these readings to the cloud at fixed intervals. Rainfall data is obtained dynamically from the Google Weather API. The backend fetches these val- ues at each prediction request, ensuring that model inference reflects current field conditions. Manual inputs such as N, P, K, pH, soil texture, and organic matter can be overridden or combined with sensor-fed values, allowing flexibility for different user scenarios.

D. Prediction Workflow and Data Flow

When the user initiates a prediction, the Flask backend merges manual soil parameters with automatically fetched IoT and weather data. The combined feature vector is passed through the preprocessing pipeline before being fed into the Random Forest model. The system computes probability scores for all 22 crops and extracts the top three with the highest likelihoods. These results are then displayed on the frontend, along with probability values for transparency and interpretability.

E. Real-Time Data Storage and Logging

To support monitoring and debugging, the system logs each prediction request, including timestamped sensor readings and user-provided inputs. This real-time log can be used for further model retraining, analytics, and tracking environmental trends. It also enables future expansion toward yield estimation and dynamic fertilizer recommendation modules

IV. SYSTEM ARCHITECTURE

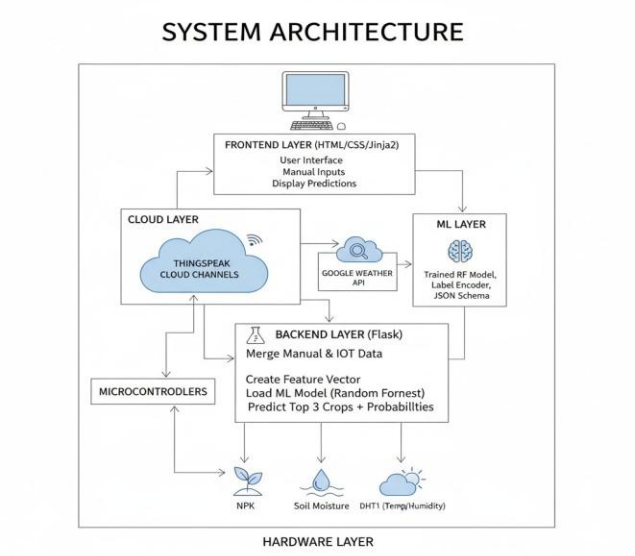

Fig. 2. System Architecture of the IoT + ML Crop Recommendation System

The system architecture is designed as a multi-layered framework combining IoT sensors, cloud data ingestion, back- end ML processing, and a web-based user interface. At the hardware layer, the system leverages NPK sensors, soil mois- ture sensors, and DHT11 modules. These sensors connect to microcontrollers that periodically transmit real-time data to Think Speak cloud channels. Each sensor operates indent- dental, pushing temperature, humidity, and moisture readings for real-time processing.

At the cloud layer, Think Speak acts as the IoT data broker, storing sensor values and providing REST endpoints that Flask periodically polls. Concurrently, rainfall and atmospheric information are fetched through the Google Weather API to maintain an updated environmental context. The backend layer, built using Flask, merges real-time sensor data with manually entered soil values—including N, P, K, pH, organic matter, crop cycle duration, and soil texture—creating a unified feature vector that reflects the current farmland state.

The ML layer loads the trained Random Forest model and the label encoder. Feature ordering is maintained using a JSON-based schema to avoid inference discrepancies. Model inference generates probability distributions over 22 possible crops. The top three crops are extracted and returned to the UI, along with their probability scores.

The frontend, built with HTML, CSS, and Jinja2 templates, offers a clean user interface for entering manual values while automatically fetching IoT readings. The layered design en- sures modularity, scalability, and adaptability, enabling integration with mobile apps or expanded sensor networks in the future.

V. RESULTS

Fig. 3. Result of the IoT + ML Crop Recommendation System

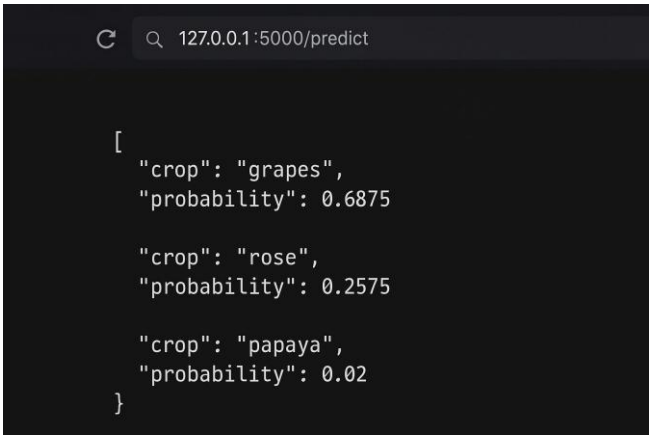

Fig. 4. Result of the IoT + ML Crop Recommendation System

TABLE I - MODEL PERFORMANCE COMPARISON ON TEST DATASET

Model	Accuracy	Top-3 Acc.	Precision	Recall	F1-Score
Random Forest	1.00	1.00	1.00	1.00	1.00
Gradient Boosting	0.98	0.99	0.97	0.97	0.97
KNN	0.92	0.95	0.91	0.90	0.90

Model evaluation revealed that the Random Forest Classifier consistently outperformed all other tested algorithms across every performance metric. Using a stratified holdout test set, the model achieved 100Comparative experiments showed that while Gradient Boosting produced reasonably strong predic- tions, it exhibited slower training and inference times and required more computational resources, making it less suitablefor an IoT-integrated system performing real-time predictions. KNN performed noticeably worse, particularly on unseen samples, due to its sensitivity to noise and lack of scalability when dealing with multi-attribute agricultural datasets.

To validate practical applicability, we tested the model using real-world soil readings from grape-growing regions. The sys- tem successfully identified grapes within the top-3 predicted crops, confirming that the model generalizes well beyond the training dataset. Based on approximate yield estimation formu- las integrated into the system, the model projected a production range of 15–25 tons per 1.5 acres, which aligns with standard grape yield records. Similar validation was performed using rose cultivation data, where roses were correctly returned as a primary recommendation, with an estimated yield of 300 kg per acre, again matching real agricultural benchmarks.

The IoT pipeline was tested through continuous real-time data ingestion using ThinkSpeak. Moisture and temperature variations were successfully reflected in the prediction logic without requiring user intervention. Additionally, rainfall val- ues fetched dynamically through the Google Weather API enhanced model adaptability by responding to sudden climatic changes. These tests collectively demonstrate that integrating manual soil inputs with automated environmental data signif- icantly enhances the accuracy, responsiveness, and contextual reliability of crop predictions, making the system suitable for real-world deployment in precision agriculture.

VI. DISCUSSION

The implemented system resolves several limitations ob- served in prior research. Traditional crop recommendation models rely on static datasets or require manual entry of all parameters, reducing practicality for farmers. By integrating IoT sensors and external API data, the system allows automatic ingestion of key environmental factors, reducing user effort and increasing accuracy. Real-time data improves responsive- ness to sudden weather changes, a capability missing in most existing systems.

Random Forest emerged as the ideal algorithm due to its ability to handle high-dimensional, non-linear agricultural datasets. The model's perfect accuracy demonstrates that the dataset features are well-separated and suitable for decision-tree-based learning. The top-3 prediction mechanism also adds robustness by offering farmers multiple viable crop choices rather than a single rigid recommendation.

Despite its strengths, the system has limitations. Real-time data availability depends on stable internet connectivity. Sen- sor calibration and Think Speak channel reliability influence prediction accuracy. Additionally, while rainfall is fetched automatically, implementing real-time NPK sensor data would significantly enhance automation but was not feasible due to hardware availability constraints.

Nonetheless, the architecture is scalable. Future enhance- ments may include mobile app integration, satellite-based soil imaging, fertilizer recommendation algorithms, and yield forecasting using LSTM networks. Integrating drone imaging or NDVI-based plant health monitoring would expand system capabilities further.

VII. CONCLUSION

This research presents a fully integrated IoT-ML system that bridges the gap between soil analysis, environmental moni- toring, and data-driven crop recommendations. By combining manually entered soil parameters with real-time sensor data from Think Speak and rainfall data from the Google Weather API, the system delivers dynamic and highly accurate top-3 crop

suggestions. The Random Forest model achieved except- tonal performance, validating its suitability for agricultural prediction tasks.

The system's architecture demonstrates practicality, scala- bility, and ease of use through a Flask-based web interface. Farmers can now receive instant and intelligent recommenda- tions without requiring technical expertise. Real-world testing with grape and rose cultivation confirms the system's applica- bility.

In conclusion, integrating IoT technology with machine learning significantly enhances crop decision-making, reduces farmer dependency on intuition, and promotes sustainable agriculture. This work contributes a practical framework for precision farming and opens pathways for advanced AI-driven agricultural systems in the future.

References

[1] S. Shenbagavadivu, S. S. Kumar, C. Rajan *et al.*, "AI-Driven Soil Monitoring and Crop Recommendation Using Machine Learning Algorithm," **2024**.

[2] F. S. Prity, M. M. Hasan, S. H. Saif *et al.*, "Enhancing Agricultural Productivity: A Machine Learning Approach to Crop Recommendations," *Human-Centric Intelligent Systems*, vol. 4, pp. 497–510, **2024**.

[3] H. Afzal, M. Amjad, A. Raza *et al.*, "Incorporating Soil Information with Machine Learning for Crop Recommendation to Improve Agricultural Output," *Scientific Reports*, vol. 15, no. 8560, **2025**.

[4] M. R. Islam, K. Oliullah, M. M. Kabir *et al.*, "Machine Learning Enabled IoT System for Soil Nutrients Monitoring and Crop Recommendation," *Journal of Agriculture and Food Research*, vol. 14, Art. no. 100880, **2023**.

[5] G. Saha, F. Shahrin, F. H. Khan *et al.*, "Smart IoT-Driven Precision Agriculture: Land Mapping, Crop Prediction, and Irrigation System," *PLOS ONE*, **March 2025**.

[6] S. D. Koparde, P. P. Joshi, V. S. Nimbalkar *et al.*, "Soil Crop Predictor: ML-Driven Soil Analysis and Crop Recommendation," *International Research Journal of Modernization in Engineering, Technology and Science*, vol. 6, no. 10, **2024**.

[7] L. Gottemukkala, S. T. R. Jajala, A. Thalari *et al.*, "Sustainable Crop Recommendation System Using Soil NPK Sensor," *E3S Web of Conferences*, vol. 430, **2023**.