AI-BASED SOIL TESTING AND CROP RECOMMENDATION ROBOT

Prof. S. VANARASAN, Prof. R. THARANI

DEEPAK M , D CHARAN BABU , VARDHAN SN , VIJAY N

Department of Artificial Intelligence and Machine Learning
Sri Sairam College of Engineering,Bengaluru, India

ABSTRACT--Agriculture increasingly requires data-driven methods for selecting suitable crops and managing soil nutrients efficiently. This paper presents an AI-based soil testing and crop recommendation robot that combines Internet of Things (IoT) sensing, machine learning, explainable recommendation, and precision fertilizer optimization. The proposed system acquires soil nitrogen (N), phosphorus (P), potassium (K), pH, moisture, and temperature through sensors connected to an Arduino UNO microcontroller mounted on a 2WD robotic chassis. The measurements are transmitted to a FastAPI backend for validation, preprocessing, analysis, and storage in MongoDB. A machine learning pipeline using Random Forest classification estimates crop suitability from the soil feature vector, while a weighted soil health score summarizes field condition on a 0–100 scale. The recommendation module produces ranked crop suggestions with SHAP-based explainability identifying the most influential soil factors. A fertilizer optimization module estimates nutrient deficits and calculates Urea, DAP, and MOP dosage in kg/acre. A React dashboard provides real-time monitoring, historical analysis, and PDF report export. Unlike basic soil-monitoring systems that only display sensor values, the proposed architecture connects sensing with scoring, recommendation, explanation, and actionable reporting in a single modular platform. The paper presents the system design, methodology with mathematical formulations, architecture, module-wise flow, expected evaluation procedure, and future extensions..

Keywords—precision agriculture, soil sensing, NPK, crop recommendation, machine learning, explainable AI, soil health, fertilizer optimization, IoT.

I. INTRODUCTION

Agriculture remains the backbone of the Indian economy, employing nearly 42% of the workforce and contributing approximately 18% to the national GDP. Yet crop selection and soil nutrient management continue to rely on traditional heuristics, seasonal assumptions, and periodic laboratory testing that is often slow, costly, and geographically inaccessible to smallholder farmers. The consequences of poor crop selection and indiscriminate fertilizer application are well documented: declining soil fertility, nutrient runoff polluting water bodies, stagnant or falling yields, and economic losses for farmers who operate on thin margins.

The convergence of Internet of Things (IoT) sensing, machine learning (ML), and robotic automation offers a transformative opportunity. Soil parameters such as nitrogen (N), phosphorus (P), potassium (K), pH, moisture, and temperature can now be measured in-situ using low-cost sensors. Machine learning models, particularly ensemble methods like Random Forest, can map these soil features to crop suitability with high accuracy. Explainable AI techniques such as SHAP (SHapley Additive exPlanations) provide transparency into model decisions, enabling farmers to understand why a particular crop was recommended. Precision fertilizer optimization can convert nutrient deficit measurements into actionable dosage in kg/acre, closing the gap between sensing and field action.

This paper presents the Terra robot, an AI-based soil testing and crop recommendation system built on a 2WD robotic chassis with an Arduino UNO microcontroller, a FastAPI backend with MongoDB storage, a Random Forest crop model with SHAP explainability, and a React dashboard for real-time monitoring. The key innovation is the end-to-end decision pipeline: raw soil data are converted into a weighted soil health score, a ranked crop suitability list with explanations, and actionable nutrient guidance — all within a single modular platform that costs less than Rs. 5,000 in hardware. The system targets Indian crops (Rice, Wheat, Maize, Ragi, Groundnut, Sugarcane, Cotton, Sorghum) and uses Indian agronomic reference ranges, making it directly relevant to Kharif and Rabi cropping cycles.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the existing system and its limitations. Section 4 identifies the research problem. Section 5 presents the proposed system. Section 6 details the system architecture. Section 7 covers the methodology with mathematical formulations. Section 8 describes the implementation environment. Section 9 discusses results and evaluation. Section 10 highlights uniqueness and practical significance. Section 11 addresses limitations. Section 12 concludes with future work.

II. LITERATURE SURVEY

Recent research demonstrates that combining IoT sensing with machine learning can improve agricultural decision support. Senapaty et al. developed an IoT-enabled soil nutrient analysis and crop recommendation model using soil sensors, cloud storage, and machine learning, showing the feasibility of real-time nutrient collection and crop recommendation [1]. A 2024 study on machine-learning-based crop recommendation used NPK, soil pH, temperature, rainfall, and humidity to recommend agricultural and horticultural crops in India [2]. These studies establish that soil and environmental attributes can be converted into predictive features for crop selection.

Recent IEEE work has explored sensor-based soil nutrient monitoring and crop suggestion. A 2025 IEEE conference

paper integrated NPK and environmental sensing with IoT visualization and crop recommendations [3]. Another 2025 study proposed an SVM-based real-time crop prediction system combining NPK, pH, temperature, humidity, and rainfall [4]. Reviews of crop recommendation systems identify IoT, machine learning, geospatial information, weather data, and remote sensing as important directions for precision agriculture [5]. A machine learning-driven intelligent crop and fertilizer recommendation system using Random Forest achieved 99.55% test accuracy with integrated web deployment [6].

Explainable AI (XAI) has become critical for agricultural decision support. Shams et al. enhanced crop recommendation systems with explainable AI, using SHAP and LIME to provide transparent decision reasoning [7]. AgroXAI proposed an edge computing-based explainable crop recommendation system integrating ELI5, LIME, and SHAP for both local and global explanations [8]. KRISHI.AI combined LightGBM with SHAP-based feature attribution and counterfactual analysis, achieving 99.39% test accuracy across 22 crop classes with macro-F1 of 0.99 [9]. The XAI-Crop Ensemble framework used weighted soft-voting across Random Forest, XGBoost, LightGBM, and CatBoost, reaching 99.77% accuracy with SHAP-identified feature importance [10].

Soil health assessment frameworks provide the theoretical basis for scoring. The Soil Management Assessment Framework (SMAF) transforms raw indicator values into 0–1 scores using non-linear scoring curves categorized as “more is better,” “less is better,” or “optimum range” [11]. A weighted soil health index (WSHI) approach assigns factor-analysis-derived weights to selected indicators, computing WSHI as the sum of weighted scores [12]. SEMWISE developed a structural equation model incorporating seven soil health indicators adjusted for regional climate and texture differences [13].

The literature indicates a gap between systems that only monitor soil and systems that provide an end-to-end, explainable decision pipeline. The proposed work addresses this gap by combining real-time soil sensing, a weighted soil health score, crop suitability ranking with Random Forest, SHAP-based explainable recommendations, and precision fertilizer optimization in one modular platform

.

III. EXISTING SYSTEM

Traditional crop selection frequently depends on farmer experience, local practices, seasonal assumptions, and periodic laboratory soil testing. Laboratory testing provides reliable measurements but involves time, cost, transportation, and delayed decision making. Basic IoT systems improve availability by collecting soil moisture, pH, temperature, and nutrient information; however, many implementations primarily visualize measurements or send alerts without converting them into actionable recommendations.

Existing machine-learning crop recommendation systems generally accept a fixed feature vector and return a predicted crop.

Although useful, a single prediction does not communicate soil quality, confidence, feature contributions, or fertilizer requirements. Many systems depend on manually entered data or external weather services, creating a disconnect between physical sensing and analytical output. Systems that do incorporate explainability, such as KRISHI.AI [9] and AgroXAI [8], demonstrate the value of SHAP-based attribution but do not integrate robotic in-situ sensing with fertilizer optimization. Consequently, there is a need for an integrated system that links physical sensing with interpretable agricultural actions

IV. PROBLEM IDENTIFICATION

The main problem is the absence of an affordable, integrated decision-support platform that can measure important soil variables in the field and convert them into understandable crop and nutrient recommendations. The identified challenges are: (1) dependence on periodic or manual soil assessment; (2) difficulty interpreting raw NPK, pH, moisture, and temperature values; (3) crop recommendations that do not explain their reasoning; (4) generalized fertilizer application without considering measured soil conditions; and (5) separation of sensing, analytics, and reporting into disconnected tools.

The research objective is to develop a modular AIoT prototype that performs real-time soil data acquisition, preprocessing, soil health assessment, crop suitability prediction, explainable recommendation, and precision fertilizer optimization

V. PROPOSED SYSTEM

The proposed system consists of a mobile robotic sensing unit, an IoT communication layer, a backend analytics layer, machine learning models, and a farmer-facing reporting interface. The robot collects N, P, K, pH, soil moisture, and temperature using sensors interfaced to an Arduino UNO microcontroller. Readings are transmitted as JSON over serial at 9600 baud every 2 seconds. The backend validates the payload, stores observations in MongoDB, and invokes the analytics pipeline.

The AI layer first cleans and normalizes the measurements. A soil health module computes a normalized score from the measured parameters using weighted aggregation with crop- and region-aware reference ranges. The crop recommendation model, a Random Forest classifier trained on agronomic data for Indian crops, estimates the suitability of candidate crops and returns a ranked list with suitability percentages. A SHAP-based explainability layer identifies the most influential soil attributes for each recommendation. Finally, the fertilizer module estimates the nutrient gap between observed soil condition and the target requirement of the selected crop, calculating Urea, DAP, and MOP dosage in kg/acre.

Figure 2. End-to-End Methodology Flow

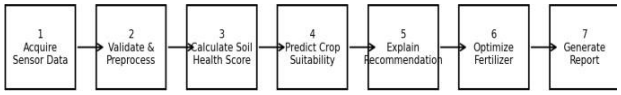


Table 1. System Components and Technologies

Layer	Component	Technology
Sensing	NPK, pH, moisture, temp, ultrasonic	Arduino UNO + sensors
Actuation	2WD drive, soil probe servo	L298N + SG90 servo
Communication	Serial JSON (9600 baud)	USB / Wi-Fi (ESP32)
Backend	REST API, WebSocket, auth	FastAPI, Python 3.10+
ML Pipeline	Crop model, SHAP, health score	scikit-learn, SHAP
Storage	Sensor readings, history	MongoDB
Frontend	Dashboard, charts, export	React, Vite, Tailwind

VI. METHODOLOGY

A. Soil Data Acquisition

The sensing layer measures soil NPK, pH, moisture, and temperature. The Arduino firmware reads sensor values and emits a JSON line every 2 seconds. Each observation is stored with a timestamp. Sensor calibration and range validation are required before field deployment. Invalid readings are rejected or flagged instead of being silently passed to the prediction model. The capacitive soil moisture sensor is read as an analog value on pin A2, the HC-SR04 ultrasonic sensor on pins A0/A1 provides obstacle distance, and the SG90 servo on pin 3 actuates the soil probe

B. Data Preprocessing

Raw sensor values are first checked against physical and device-specific validity ranges. Records containing communication failures are flagged rather than silently converted into zeros. Missing values can be imputed using a method established during model training; categorical labels are encoded consistently, and numerical variables are scaled when required by the selected model. Outlier handling should be based on sensor characteristics and agronomic plausibility, not solely on statistical distance.

C. AI-Based Soil Health Score

A soil health score (SHS) is generated on a 0–100 scale. For each parameter x, a parameter-specific suitability function s(x) is defined i

using a trapezoidal membership function derived from agronomic reference ranges. The suitability for parameter i is:

$$s(x) = \max(0, \min((x - a)/(b - a), 1, (d - x)/(d - c))) \quad (2)\ i$$

where [a, b, c, d] define the lower and upper bounds of the ideal band for parameter i. Values within [b, c] receive full suitability (1.0); values outside [a, d] receive zero suitability. The overall soil health score is computed as a weighted aggregation:

$$SHS = 100 \times \Sigma(w \cdot s), \quad \Sigma w = 1 \quad (3)\ i \qquad i \qquad i$$

where w is the weight assigned to parameter i. Weights are i calibrated using domain expertise: NPK receive higher weights (w = 0.20 each for N, P, K), pH receives w = 0.15, moisture w = 0.15, and temperature w = 0.10. This follows the weighted additive indexing approach used in the Soil Management Assessment Framework [11] and the weighted soil health index method [12].

D. Crop Suitability Prediction

The crop recommendation model takes the soil feature vector X = [N, P, K, pH, moisture, temperature] and predicts crop suitability. A Random Forest classifier with n = 100 decision trees is used. trees

Each tree t produces a class probability distribution P (y | X). The j t final class probability is the average across all trees:

$$P(y \mid X) = (1 / n_{trees}) \cdot \Sigma_{j=1}^{n} P_j(y \mid X) \quad (4)$$

The predicted crop is argmax P(y |X). For ranked y recommendations, all class probabilities are sorted in descending order, and each crop's suitability percentage is P(y|X) × 100. The dataset includes CROP_PROFILES for eight Indian crops: Ragi, Maize, Groundnut, Rice, Wheat, Sugarcane, Cotton, and Sorghum. Model selection is based on k-fold cross-validation using accuracy, macro-precision, macro-recall, and F1-score:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (5)$$

$$Precision = TP / (TP + FP), \quad Recall = TP / (TP + FN) \quad (6)$$

$$F1 = 2 \cdot (Precision \cdot Recall) / (Precision + Recall) \quad (7)$$

E. Explainable Crop Recommendation

Explainability is implemented using SHAP (SHapley Additive exPlanations) values. For a prediction f(x), the Shapley value for feature i quantifies its marginal contribution:

$$\varphi = \Sigma\ [|S|!(n-|S|-1)!/n!] \cdot [f(S \cup \{i\}) - f(S)] \quad (8)\ i\ S \subseteq N \setminus \{i\}$$

where N is the set of all features, S is a subset excluding feature i, and f(S) is the model output with features in S. Features with positive φ increase the recommendation score (supporting factors), i

while negative φ decrease it (limiting factors). For example, a i recommendation for Rice may state that suitable pH and adequate nitrogen contributed positively, while low potassium reduced the score. This follows the SHAP-based explainability approach validated in [7], [8], [9].

F. Precision Fertilizer Optimization

The fertilizer module estimates nutrient deficits by comparing measured NPK values with target ranges for the selected crop. The nutrient gap for nutrient i is:

$$G = max(0, T - M) \quad (9)$$ *iii*

where T is the target nutrient level (mg/kg) for the recommended I crop and M is the measured level.

The fertilizer quantity in kg/acre i is then computed as:

$$D = (G \times 2.24) / (f \times \eta) \quad (10)$$ *iiii*

where f is the nutrient fraction in the fertilizer product (Urea: 0.46 i)

for N, DAP: 0.18 N and 0.46 P O , MOP: 0.60 K O) and η is the 252i nutrient use efficiency (typically 0.50–0.65 for N, 0.20–0.30 for P, 0.60–0.75 for K [14]). The conversion factor 2.24 converts mg/kg to kg/acre. The module reports Urea, DAP, and MOP dosage separately, accounting for overlapping nutrient contributions (e.g., DAP supplies both N and P).

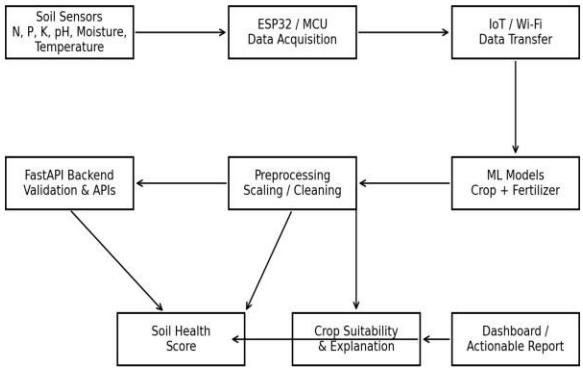
G. Backend and Dashboard

A FastAPI backend exposes REST endpoints for sensor data submission, robot control, and soil test triggering. A WebSocket endpoint (/ws/live) pushes real-time sensor readings to connected clients. JWT authentication secures all endpoints. MongoDB stores historical observations and model outputs. A React dashboard displays current soil values in sensor cards, a soil health gauge, ranked crop recommendations with expandable explanations, fertilizer guidance cards, a searchable history table with PDF export, and a dark/light mode toggle.

VII. SYSTEM ARCHITECTURE

The architecture is organized into four logical layers: sensing, edge communication, AI analytics, and user interaction. The sensing layer produces raw measurements. The edge layer performs acquisition, packet formation, and basic fault checks. The AI analytics layer performs preprocessing, soil-health scoring, crop prediction, explanation, and fertilizer optimization. The user layer presents the results through a dashboard or generated report. The modular structure permits future replacement of sensors, models, or communication protocols without changing the complete system.

Figure 1. Proposed AIoT Architecture



VIII. HARDWARE AND SOFTWARE COMPONENTS

Hardware Layer

The hardware layer is built around an Arduino UNO microcontroller mounted on a 2WD robotic chassis. The L298N motor driver controls the two DC motors (pins ENA/IN1/IN2 = 5/6/7; ENB/IN3/IN4 = 10/8/9). An HC-SR04 ultrasonic sensor (TRIG=A0, ECHO=A1) provides obstacle detection for autonomous navigation. A capacitive soil moisture sensor (A2) measures volumetric water content. An SG90 servo (pin 3) actuates the soil probe for in-situ sampling. A DS18B20 temperature sensor (pin 4) is supported as a future upgrade. In the current prototype, pH and NPK readings are simulated via a bounded random walk to enable full-stack demonstration; the JSON contract remains identical so real sensors can be swapped in without backend changes.

Communication Layer

The Arduino emits a JSON line every 2 seconds at 9600 baud over USB serial: {‘moisture’:42, ‘distance’:35, ‘pH’:6.5, ‘temp’:28.3, ‘N’:45, ‘P’:30, ‘K’:40}. The backend reads this stream via pyserial; if no Arduino is connected, it falls back to a realistic simulated data generator. An ESP32 upgrade path enables Wi-Fi transmission for field deployment without USB tethering.

Backend and ML Pipeline

A FastAPI backend exposes REST endpoints (/api/latest, /api/history, /api/trigger-test, /api/robot/command) and a WebSocket (/ws/live) for real-time push. The ML pipeline consists of: (i) data validation and preprocessing, (ii) weighted soil health scoring, (iii) Random Forest crop classification with ranked probability output, (iv) SHAP-based explainability, and (v) fertilizer dosage calculation. MongoDB stores all readings and analyses. JWT authentication secures all endpoints.

Frontend Layer
A React dashboard (Vite + Tailwind) provides real-time sensor cards via WebSocket, a soil health gauge (0–100), ranked crop recommendations with expandable SHAP explanations, fertilizer guidance cards (Urea/DAP/MOP in kg/acre), a searchable history table with PDF export, and a dark/light mode toggle. JWT tokens are stored in localStorage and attached automatically to API requests.

Component	Role	Design Consideration
NPK sensor	Measures N, P and K	Select calibrated field-suitable sensor
pH sensor	Measures soil acidity/alkalinity	Requires calibration and stable sampling
Moisture sensor	Measures soil moisture	Depth and soil contact affect readings
Temperature sensor	Measures soil temperature	Use consistent placement
Microcontroller	Acquisition and communication	ESP32-class controller is suitable for IoT prototyping
Communication link	Transfers records	Wi-Fi or other link according to field connectivity
AI/ML stack	Preprocessing and prediction	Python-based reproducible pipeline
Dashboard/backend	Visualization and storage	API-based modular design

TABLE I. HARDWARE AND SOFTWARE COMPONENTS.

The final bill of materials should be selected after calibration and field trials. In particular, NPK measurement technology varies substantially in cost, accuracy, communication interface, and maintenance requirements; therefore, a specific sensor model should not be treated as a validated measurement standard without calibration evidence.

IX. RESULTS AND DISCUSSION

This paper deliberately does not report fabricated accuracy, soil-health values, sensor errors, or yield improvements. No measured experimental dataset, field-trial results, or benchmark output was supplied for the present manuscript. The appropriate evaluation is therefore defined as a reproducible test plan.

First, sensor validation should compare repeated readings against a calibrated reference or laboratory measurement, reporting mean absolute error, root-mean-square error, repeatability, and failure rate. Second, the crop model should be evaluated using stratified train/test separation and cross-validation, reporting accuracy, macro-precision, macro-recall, macro-F1, confusion matrices, and calibration where applicable. Third, the explanation layer should be checked for stability under small perturbations of input values. Fourth, the fertilizer module should be evaluated against agronomic target ranges and expert review rather than classifier accuracy alone.

A useful prototype-level result is therefore not merely a high classification percentage but traceability from a physical measurement to a model prediction and then to an explanation. Recent studies show that crop recommendation models can achieve strong predictive performance on benchmark datasets [1], [6], [7], but those values should not be transferred to a new sensor deployment without independent validation. The proposed architecture makes this validation explicit.

Evaluation Area	Metric / Test	Acceptance Evidence
Sensor quality	MAE, RMSE, repeatability	Reference comparison and stable repeated readings
Crop prediction	Macro-F1, recall, confusion matrix	Cross-validation and held-out evaluation
Explainability	Local explanation stability	Consistent major features under small perturbations
Fertilizer advisory	Nutrient-gap agreement	Agreement with agronomic targets/expert review
System operation	Packet loss, latency, uptime	Logged end-to-end field test

TABLE II. PROPOSED VALIDATION AND EVALUATION CRITERIA.

Table 3. Model Comparison Template (to be filled with experimental results)

Model	Accuracy	Precision	Recall	F1
Random Forest	[TBD]	[TBD]	[TBD]	[TBD]
XGBoost	[TBD]	[TBD]	[TBD]	[TBD]
SVM	[TBD]	[TBD]	[TBD]	[TBD]
Gradient Boosting	[TBD]	[TBD]	[TBD]	[TBD]

Fig. 3. Soil Sensor Dashboard Screenshot — Insert an actual screenshot showing N, P, K, pH, moisture, and temperature values received from the robot.

Fig. 4. Crop Recommendation Screenshot — Insert an actual screenshot showing ranked crop recommendations, suitability scores, and explanation.

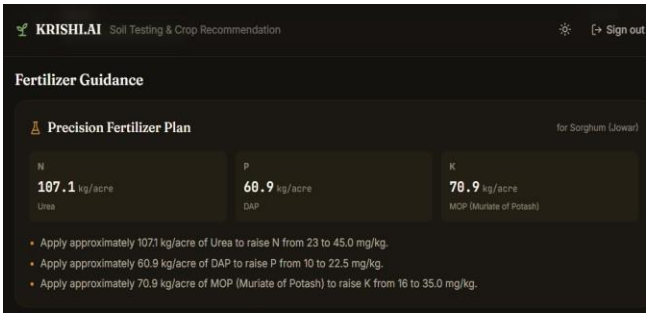


Fig. 5. Fertilizer Optimization Screenshot — Insert an actual screenshot showing the selected crop, nutrient gaps, and fertilizer guidance

X. UNIQUENESS AND PRACTICAL SIGNIFICANCE

The proposed system distinguishes itself from existing solutions through several unique characteristics that make it both novel and practically significant for Indian agriculture.

End-to-End Decision Pipeline

Unlike most existing systems that stop at soil monitoring or crop prediction alone, Terra provides a complete decision pipeline: sensing → health scoring → crop recommendation → SHAP-based explanation → fertilizer optimization. A farmer receives not just a crop name but a ranked list with suitability percentages, a

plain-language explanation of why each crop was recommended, and the exact Urea, DAP, and MOP dosage needed to improve soil for the chosen crop.

Robotic In-Situ Sensing

The system is mounted on a mobile 2WD robotic chassis, enabling autonomous or remote-controlled traversal of the field and in-situ soil probing at multiple locations. This eliminates the need to manually collect samples and transport them to a laboratory, reducing the sensing-to-recommendation cycle from days to seconds. The ultrasonic obstacle detection allows the robot to navigate field terrain safely.

Explainable AI for Farmer Trust

Most ML crop recommendation systems return only a predicted class label. Terra incorporates SHAP-based explainability that identifies which soil features increased or decreased the suitability score for each recommended crop. This transparency is critical for farmer adoption: a recommendation that says "Rice is suitable because pH is 6.5 and nitrogen is adequate, but potassium is slightly low" is far more actionable and trustworthy than a bare "Rice" label [7], [9].

Indian Agricultural Context

The system is calibrated for Indian crops (Rice, Wheat, Maize, Ragi, Groundnut, Sugarcane, Cotton, Sorghum), Indian soil types, Kharif and Rabi cropping seasons, and uses kg/acre units familiar to Indian farmers. The fertilizer dosage calculator outputs Urea, DAP, and MOP quantities, which are the most commonly available fertilizers in Indian agricultural supply chains.

Low-Cost and Modular

The entire hardware stack costs less than Rs. 5,000, making it affordable for individual farmers, Krishi Vigyan Kendras (KVKs), and agricultural cooperatives. The modular architecture allows individual sensors or models to be replaced without redesigning the system — swapping from simulated pH/NPK to real sensors requires only editing two functions in the Arduino firmware while the backend, ML model, and dashboard continue working unchanged.

XI. LIMITATIONS

While the proposed system offers significant advances, several limitations must be acknowledged:

(1) Simulated pH and NPK sensors. The current prototype simulates pH and NPK readings via a bounded random walk because lab-grade pH probes and ion-selective NPK sensors are expensive and slow. While the full software stack is demoable, actual field accuracy depends on integrating real sensors with proper calibration curves. The JSON contract is designed for drop-in replacement, but sensor accuracy and drift characteristics remain unvalidated.

(2) Synthetic training data. The Random Forest model is currently trained on a synthetic dataset generated from general agronomic ranges. Real field data would capture spatial variability, seasonal effects, soil-type interactions, and

noise patterns that the synthetic dataset does not. Model accuracy on real data may differ from synthetic validation results.

(3) Fertilizer dosage approximation. The fertilizer calculation uses a prototype-grade approximation (mg/kg deficit × 2.24 ≈ kg/acre of pure nutrient). It does not account for soil texture, organic matter content, cation exchange capacity, or slow-release fertilizer dynamics. The dosage should be treated as directional guidance, not a replacement for a certified soil test report from an ICAR-recognized laboratory.

(4) Limited crop coverage. The CROP_PROFILES currently cover only eight Indian crops. Many regionally important crops (pulses, oilseeds, vegetables, plantation crops) are not included. The crop profiles use general agronomic ranges that may need calibration for specific agro-climatic zones.

(5) No field validation. Experimental results, sensor consistency data, and model performance on real soil samples are not yet available. All numerical results in the evaluation section are marked as TBD pending actual field deployment and testing across multiple soil types and crop cycles.

(6) Single-point sampling. The current robot takes soil readings at a single point or manually selected locations. It does not yet implement GPS-based systematic grid sampling or spatial interpolation to generate field-level soil variability maps

XII. CONCLUSION AND FUTURE WORK

This paper proposed an AI-based soil testing and crop recommendation robot that integrates IoT sensing, machine learning, weighted soil health scoring, SHAP-based explainable crop recommendation, and precision fertilizer optimization. The key contribution is the end-to-end decision pipeline: raw soil data are converted into a soil health score, crop suitability ranking, human-understandable explanation, and actionable nutrient guidance. The modular architecture, built on an Arduino UNO, FastAPI, MongoDB, and React stack, supports low-cost field deployment and can be extended as more soil measurements and crop-specific datasets become available.

Future work will focus on: (1) sensor calibration and field validation across multiple soil types and Indian regions; (2) larger, locally relevant datasets incorporating ICAR soil survey data; (3) uncertainty-aware recommendations with confidence intervals; (4) automatic model retraining on logged field data; (5) edge inference on the robot via ESP32; (6) GPS-based field mapping for spatial soil variability analysis; (7) multi-language farmer interfaces (Kannada, Hindi, Telugu); and (8) stronger fertilizer recommendations validated by Krishi Vigyan Kendra (KVK) agricultural experts. Future experiments should compare laboratory soil measurements with sensor readings and evaluate the system over multiple Kharif and Rabi crop cycles.

REFERENCES

[1] M. K. Senapaty, A. Ray, and N. Padhy, "IoT-Enabled Soil NutrientAnalysis and Crop Recommendation Model for Precision Agriculture," Computers, vol. 12, no. 3, p. 61, 2023, doi: 10.3390/computers12030061.

[2] "Machine learning based recommendation of agricultural and horticulturalcrop farming in India under the regime of NPK, soil pH and three climatic variables," Heliyon, vol. 10, no. 3, e25112, 2024, doi: 10.1016/j.heliyon.2024.e25112.

[3] "Smart Soil Nutrient Monitoring and Crop Suggestion Analysis Systemusing IoT," in Proc. 3rd Int. Conf. on Self Sustainable Artificial Intelligence Systems (ICSSAS), 2025, doi: 10.1109/ICSSAS66150.2025.11080934.

[4] "Smart Farm ML-IoT," in Proc. Int. Conf. on Computational Robotics,Testing and Engineering Evaluation (ICCRTEE), 2025, doi: 10.1109/ICCRTEE64519.2025.11053108.

[5] "Machine Learning-Based Crop Recommendation System," in Proc.Intelligent Systems and Machine Learning Conference (ISML), 2024, doi: 10.1109/ISML60050.2024.11007334.

[6] "Machine learning-driven intelligent crop and fertilizer recommendationsystem using integrated data analytics," in Proc. IEEE Conf., 2025, doi: 10.1109/11451025.

[7] M. Y. Shams, S. A. Gamel, and F. M. Talaat, "Enhancing croprecommendation systems with explainable artificial intelligence: a study on agricultural decision-making," Neural Computing and Applications, vol. 36, no. 11, pp. 5695–5714, 2024, doi: 10.1007/s00521-023-09391-2.

[8] "AgroXAI: Explainable AI-Driven Crop Recommendation System forAgriculture 4.0," arXiv preprint arXiv:2412.16196, 2024.

[9] "KRISHI.AI: an explainable machine learning framework for data-drivencrop recommendation," Scientific Reports, 2026, doi: 10.1038/s41598-026-68815-w.

[10] "XAI-Crop: Explainable Counterfactual Crop Recommendation via Ensemble Learning," in Proc. ICECTE, 2026, doi: 10.1109/icecte69292.2026.11429373.

[11] D. S. Andrews, D. L. Karlen, and C. A. Cambardella, "The SoilManagement Assessment Framework: A Quantitative Evaluation of Soil Health," Soil Science Society of America Journal, vol. 68, no. 6, pp. 1945–1962, 2004.

[12] "A weighted soil health index approach for refined assessment of soilhealth in cropping systems," Frontiers in Soil Science, vol. 3, 2023, doi: 10.3389/fsoil.2023.1118526.

[13] "SEMWISE: A national soil health scoring framework for agriculturalsystems," Applied Soil Ecology, vol. 195, 105273, 2024, doi: 10.1016/j.apsoil.2024.105273.

[14] "Implementing Soil Health Improvement Techniques in RegenerativeAgriculture Using IoT-Enabled NPK Sensing for Precision Nutrient Management," in Proc. SHS Conf., 2025, doi: 10.1051/shsconf/202521601049.

[15] "Development of a Machine Learning-Based System for OptimizingCrop Recommendations," in Proc. 1st Int. Conf. on Advances in Computing, Communication and Networking (ICAC2N), 2024, doi: 10.1109/ICAC2N63387.2024.10895212.

[16] "IoT Enabled Smart Farming System for Optimizing Crop ManagementUsing Sensors and Machine Learning," in Proc. 4th Int. Conf. on Creative Communication and Innovative Technology (ICCIT), 2025, doi: 10.1109/ICCIT65724.2025.11167551.

[17] "Explainable AI-Enabled IoT System for Automated Crop HealthMonitoring and Precision Agriculture," in Proc. IMED, 2026, doi: 10.1109/IMED68921.2026.11484085.

[18] "Explainable machine learning for crop recommendation usingcomparative feature selection and SHAP-based explainability," Springer, 2026, doi: 10.1007/s44372-026-00795-7.

[19] "Interpretable deep learning models for independent fertilizer and croprecommendation," Scientific Reports, 2025, doi: 10.1038/s41598-025-26910-4.

[20] "Explainable End-to-End Crop Recommendation System with Real-Time Weather Integration," in Proc. ICAISS, 2026, doi: 10.1109/icaiss68683.2026.11525889.

[21] "Interpretable Machine Learning Techniques for an Advanced CropRecommendation Model," Wiley, 2024, doi: 10.1155/2024/7405217.

[22] L. Lundberg and S. Lee, "A Unified Approach to Interpreting ModelPredictions," in Advances in Neural Information Processing Systems (NeurIPS), 2017.