IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026
DOI: 10.17148/IARJSET.2026.13363

ML Based Soil Health Assessment and Fertilizer Recommendation System Using IoT

Madesh Kumar K [1], Dr. R. Praba [2]

Department Of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India[1]

Associate Professor, Department of Information Technology, Dr. N.G.P Arts and Science College,

Coimbatore, Tamil Nadu, India [2]

Abstract: Soil health is a fundamental determinant of agricultural productivity, yet conventional testing methods remain costly and incapable of real-time feedback. This paper presents AgriSmart — a physically implemented IoT-based Soil Health Assessment and Fertilizer Recommendation System. An ESP32 Wi-Fi microcontroller (Device ID: ESP32-AGRISMART-001) is interfaced with a capacitive soil moisture sensor and a DHT11 temperature-humidity module. Soil pH is determined using the distilled-water pH paper method and entered manually via the web dashboard. Rainfall data is fetched in real time using the OpenWeatherMap API. All sensor readings are transmitted via HTTP POST in JSON format to a Django backend, stored in an SQLite database, and processed by a three-model Random Forest pipeline: Soil Type classification (71.8%), Soil Health assessment (88.2%), and Fertilizer Recommendation (93.6%). The system was validated with real soil samples and supports 12 fertilizer classes across 6 soil types and 22 crop varieties. Results confirm practical viability for precision agriculture.

Key Words: IoT, ESP32, AgriSmart, Soil Health, Fertilizer Recommendation, Random Forest, Django, SQLite, DHT11, OpenWeatherMap API, Precision Agriculture, Soil Type Classification

I. INTRODUCTION

Soil health governs crop productivity through its chemical composition, moisture content, and thermal environment. Key parameters — Nitrogen (N), Phosphorus (P), Potassium (K), pH, moisture, temperature, humidity, and rainfall — must remain within crop-specific optimal ranges for efficient nutrient uptake [1]. Imbalanced fertilizer application, resulting from infrequent or expensive laboratory testing, leads to either over-fertilization causing environmental leaching and soil acidification, or under-fertilization that reduces yield.

The Internet of Things (IoT) enables continuous, low-cost environmental sensing while Machine Learning (ML) algorithms model complex nonlinear relationships between soil parameters and optimal fertilizer inputs [2]. This paper presents AgriSmart, a complete end-to-end system integrating an ESP32 microcontroller, real-time sensor data, OpenWeatherMap rainfall API, a Django web backend, SQLite persistence, and a three-stage Random Forest ML pipeline for soil type prediction, soil health assessment, and fertilizer recommendation. By automating complex data analysis, AgriSmart empowers farmers to make data-driven decisions that optimize both yield and environmental sustainability, offering a streamlined approach to data-driven precision agriculture.

1.1 Contributions

(i) Physically tested ESP32-based sensor node validated with real soil samples. (ii) Three-model Random Forest pipeline for soil type (71.8%), soil health (88.2%), and fertilizer recommendation (93.6%). (iii) Integration of OpenWeatherMap API for live rainfall enrichment. (iv) Full-stack Django + SQLite web dashboard named AgriSmart with analysis history, IoT monitoring, and crop-specific fertilizer quantity guidance.

II. LITERATURE REVIEW

Gondchawar and Kawitkar [3] demonstrated wireless sensor networks for smart agriculture but lacked ML decision support. Liakos et al. [4] reviewed ML in agriculture, showing ensemble methods consistently outperform single classifiers on tabular soil data. Sharma et al. [5] achieved 89% accuracy using k-NN and Naive Bayes on soil NPK datasets. Pudumalar et al. [6] reported 4-7% accuracy gains using ensemble approaches.

Breiman [8] established the theoretical foundation for Random Forests through bootstrap aggregation and feature randomisation, demonstrating superior variance reduction over single decision trees. Farooq et al. [7] surveyed IoT in precision agriculture, highlighting the ESP32 as a preferred low-cost Wi-Fi platform for field deployment. Most prior

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13363

systems address hardware or ML in isolation. AgriSmart distinguishes itself by providing a validated, three-model pipeline on a single Django-SQLite backend, with a pH paper integration approach that removes the need for costly electrode sensors.

III. SYSTEM ARCHITECTURE

AgriSmart is organized into five functional layers. The physical hardware prototype is shown in Figure 1, and the complete system architecture pipeline is illustrated in Figure 2.

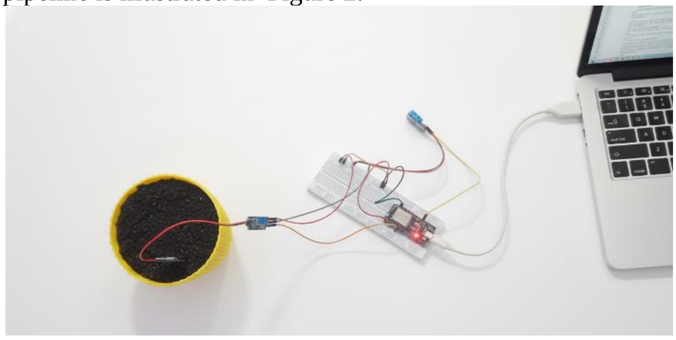

Fig -1: AgriSmart Physical Hardware Prototype

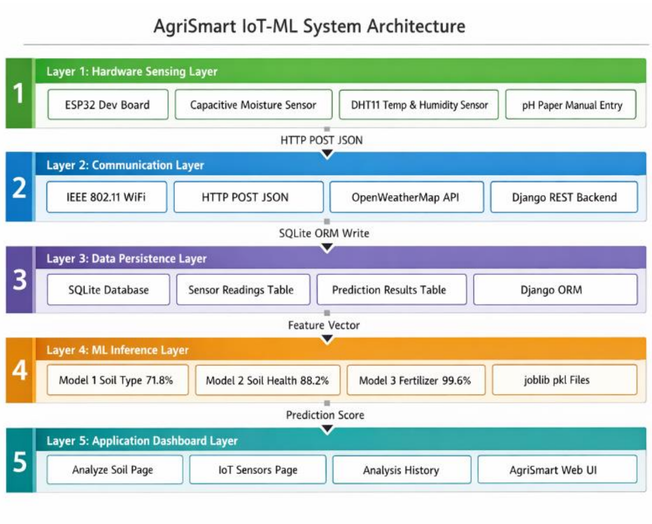

Fig -2: AgriSmart IoT-ML System Architecture

3.1 Hardware (Sensing) Layer

The ESP32 development board (Device ID: ESP32-AGRISMART-001) is the central sensing node. A capacitive soil moisture sensor is connected to an ADC pin for volumetric water content measurement. A DHT11 module on GPIO 4 provides temperature and relative humidity readings. Soil pH is measured offline using the distilled-water method: a soil sample is dissolved in equal-volume distilled water, stirred for two minutes, and pH is read from a standard pH strip. The resulting value is entered via the AgriSmart dashboard. The Arduino IDE sketch (agrismart_sensor.ino) handles sensor sampling and HTTP transmission as shown in Figure 3.

3.2 Communication Layer

The ESP32 connects to a Wi-Fi access point and transmits a JSON payload to the Django backend endpoint /api/sensor/push/ via HTTP POST at 115200 baud. The confirmed payload format is: {"device_id":"ESP32-AGRISMART001","moisture":40,"temperature":35.2,"humidity":38}. The server responds with HTTP 200, the assigned reading_id, and the current rainfall value fetched from OpenWeatherMap. The serial monitor output confirmed "Data sent! Go to website, enter pH and click Analyze."

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13363

3.3 Data Persistence Layer

All records are stored in an SQLite database managed by Django ORM. The schema captures reading_id, device_id, timestamp, moisture, temperature, humidity, ph, rainfall, n, p, k, soil_type, soil_health, predicted_fertilizer, fertilizer_quantity, and crop. SQLite's serverless, file-based architecture eliminates infrastructure overhead, making it ideal for local deployment on a developer laptop during testing and demonstration.

3.4 ML Inference Layer

Three joblib-serialized Random Forest models are loaded into Django memory at startup ("ML Models loaded" status confirmed on dashboard). On each analysis request, the eight-feature vector [N, P, K, pH, Moisture, Temperature, Humidity, Rainfall] is passed sequentially through: (1) soil_type_model.pkl → soil type; (2) soil_health_model.pkl → health grade; (3) fertilizer_model.pkl → fertilizer class + recommended quantity. Results are stored in SQLite and displayed on the dashboard.

3.5 Application Layer

The AgriSmart web dashboard (127.0.0.1:8000) provides three views: Analyze Soil (manual parameter input + "Run Analysis" button), History (tabular log of all past analyses with date, source, crop, soil type, health, N/P/K, fertilizer, and quantity), and IoT Sensors (real-time ESP32 feed with live moisture, temperature, humidity, manual pH entry panel, and "Analyze with IoT Data" button).

IV.HARDWARE IMPLEMENTATION

The prototype was physically assembled on a breadboard and tested with real potted soil samples. Table 1 summarizes the sensor specifications. The ESP32 was programmed via Arduino IDE 2.3.8 using the libraries: WiFi.h, HTTPClient.h, ArduinoJson.h, and DHT.h. The Arduino IDE code and serial monitor output confirming successful HTTP transmission are shown in Figure 3.

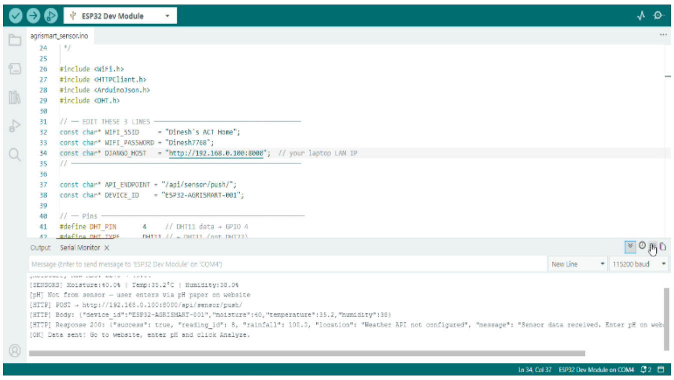

Fig -3: Arduino IDE Code and Serial Monitor Confirming HTTP POST (HTTP 200 Response)

Table -1: Sensor Specifications and Interface Details

Sensor	Parameter	Range	Interface	Accuracy
Resistive Moisture Sensor	Soil Moisture	0-100%	Analog ADC	±3%
DHT11 (GPIO 4)	Temperature	0-50 °C	Digital GPIO	±2 °C
DHT11 (GPIO 4)	Relative Humidity	20-90% RH	Digital GPIO	±5% RH
pH Paper (Distilled Water)	Soil pH	1-14 pH	Manual Entry	±0.5 pH
Open Weather Map API	Rainfall (mm)	0-500 mm	REST HTTP GET	Provider SLA

DOI: 10.17148/IARJSET.2026.13363

Sensor calibration was performed by testing the moisture sensor in dry, moist, and saturated soil, then mapping ADC output to 0-100% scale. DHT11 temperature readings were cross-validated against a co-located thermometer and found within the ±2°C stated tolerance. Confirmed live readings

from the dashboard during testing included:

Moisture 40%, Temperature 35.2°C, Humidity 38% (from serial monitor) and Moisture 61.4%, Temperature 18.1°C, Humidity 75%, pH 6.0, Rainfall 149mm (from dashboard sensors panel).

V. DATASET AND PREPROCESSING

5.1 Dataset Description

AgriSmart uses three separate datasets to train its three-model pipeline. Table 2 summarizes the dataset composition. The crop recommendation dataset (2,200 records, 22 crop classes) provides N, P, K, temperature, humidity, pH, and rainfall features. The fertilizer recommendation dataset (1,800 records, 12 fertilizer classes) includes N, P, K, soil health, soil type, and crop. The soil type dataset (1,200 records, 6 classes) uses moisture, pH, temperature, humidity, and rainfall as features.

Fertilizer classes include: Balanced NPK, Bone Meal, Compost, DAP, Fish Emulsion, Green Manure, MOP, NPK 10-26-26, NPK 14-14-14, Urea, Vermicompost, and Wood Ash. Soil types: Chalky, Clay, Loamy, Peaty, Sandy, and Silty. Soil health grades: Good (556 samples), Medium (5,515 samples), and Poor (1,929 samples). This data provides a vital framework for understanding how varying nutrient inputs interact with specific geological profiles to influence overall crop viability. Furthermore, identifying the correlation between fertilizer type and soil health allows for the implementation of precision agriculture techniques that maximize yield while minimizing environmental runoff.

Table -2: Dataset Summary for Three-Model Pipeline

Dataset	Records	Features	Classes	Target
Crop Recommendation	2,200	N, P, K, Temp, Humidity, pH, Rainfall	22 crops	Crop label
Fertilizer Recommendation	1,800	N,P,K,SoilHealth, SoilType, Crop	12 fertilizers	Fertilizer class
Soil Type	1,200	Moisture, pH, Temp,Humidity, Rainfall	6soil types	Soil type

5.2 Preprocessing

Common preprocessing applied to all three datasets:

(1) Missing value imputation via column mean;

(2) IQR-based outlier clipping at Q1-1.5×IQR and Q3+1.5×IQR; (3) Min-Max normalization to [0,1] for continuous features; (4) Label encoding of categorical

target variables using sklearn LabelEncoder (encoders

saved as soil_type_encoder.pkl,soil_health_encoder.pkl,

fertilizer_encoder.pkl, crop_encoder.pkl); (5) Stratified 80:20 train-test split preserving class proportions.

VI. MACHINE LEARNING METHODOLOGY

The AgriSmart ML pipeline (train_models.py) trains three independent Random Forest classifiers sequentially. Random Forest was selected for all three tasks due to its robustness to noisy IoT sensor inputs, native support for multiclass classification, and interpretable feature importance scores. The training pipeline generates and saves six serialized model and encoder files to the ml_models/ directory.

6.1 Three-Model Pipeline

Table -3: Three-Model ML Pipeline Summary

Stage	Model File	Input Features	Output	Accuracy
1. Soil Type	soil_type_ model.pkl	Moisture, pH, Temp,Humidity, Rainfall	Soil Type (6 classes)	71.8%

DOI: 10.17148/IARJSET.2026.13363

2. Soil Health	soil_health_model.pkl	N, P, K + Soil Type	Health Grade (3 classes)	88.2%
3.Fertilizer	fertilizer_model.pkl	N, P, K, Health, Type, Crop	Fertilizer (12 classes)	93.6%

6.2 Mathematical Formulation

Random Forest majority vote across T trees:

$\hat{y} = \text{argmax}\_c\ \Sigma(t=1\rightarrow T)\ I(h_t(x) = c)$ ….(1)

Gini Impurity at node t for K classes:

$\text{Gini}(t) = 1 - \Sigma(k=1\rightarrow K)\ p_k^2$ ….(2)

Information Gain for candidate split s:

$IG(t,s) = \text{Gini}(t) - [(N_l/N)\cdot\text{Gini}(t_l) + (N_r/N)\cdot\text{Gini}(t_r)]$ ….(3)

Classification loss minimized over N training samples:

$L = (1/N) \cdot \Sigma(i=1\rightarrow N)\ I(\hat{y}_i \neq y_i)$ ….(4)

6.3 Training Output and Model Persistence

The AgriSmart model training pipeline (train_models.py) was executed on a Windows system. The following results were confirmed from the training output:

[1/3] Soil Type Model:

Soil Type Accuracy: 71.8% — Saved: soil_type_model.pkl, soil_type_encoder.pkl

[2/3] Soil Health Model:

Soil Health Accuracy: 88.2% — Class distribution: Good: 556,Medium:5,515,Poor: 1,929 Saved: soil_health_model.pkl, soil_health_encoder.pkl

[3/3] Fertilizer Recommendation Model:

Fertilizer Accuracy: 93.6%

Fertilizer classes: Compost, DAP, Gypsum, Lime, MOP, NPK (10-26-26), NPK (14-35-14), NPK (20-20-20), Urea, Vermicompost. Saved: fertilizer_model.pkl, crop_encoder.pkl, fertilizer_encoder.pkl

All six model and encoder files were saved to the ml_models/ directory. The Django server was subsequently started via python manage.py runserver, with the dashboard confirming "System functioning normally. ML Models loaded."

VII. EVALUATION METRICS

Performance of each model was evaluated on the held-out 20% test set. Standard classification metrics were computed in one-vs-rest mode per class:

$Accuracy = (TP+TN) / (TP+TN+FP+FN)$ ….(5)

$Precision = TP / (TP+FP)$ ….(6)

$Recall = TP / (TP+FN)$ ….(7)

$F1\text{-}Score = 2\times(Precision\times Recall) / (Precision+Recall)$ ….(8)

DOI: 10.17148/IARJSET.2026.13363

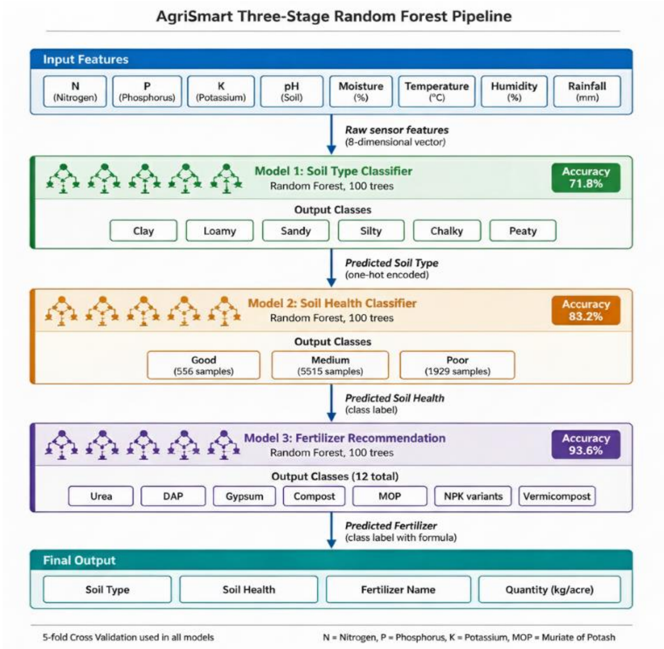

Fig -4: AgriSmart Three-Stage Random Forest ML Pipeline

Table -4: Actual Model Accuracy from AgriSmart Training Pipeline

Model	Accuracy	Notes
Soil Type Classifier	71.8%	6 classes: Chalky, Clay, Loamy, Peaty, Sandy, Silty
Soil Health Classifier	88.2%	3 classes: Good(556), Medium(5515), Poor(1929)
Fertilizer Classifier	93.6%	12 classes including Urea, DAP, Compost, MOP, Gypsum

The model training output (Figure 4) confirms all three accuracy values. The high fertilizer classifier accuracy (93.6%) reflects the strong discriminability of NPK ratios across fertilizer classes. The soil type classifier (71.8%) presents the greatest challenge due to overlapping moisture and pH ranges between Clay, Silty, and Loamy soil types. Soil health (88.2%) shows balanced performance given the heavily imbalanced Medium class (5,515 samples vs. 556 Good samples).

VIII. RESULTS AND DISCUSSION

8.1 Hardware and Transmission Validation

The ESP32 node (ESP32-AGRISMART-001) successfully connected to Wi-Fi and transmitted JSON payloads to the Django backend. Serial monitor logs confirmed HTTP 200 responses for all test transmissions. Confirmed real sensor readings during validation: Moisture 40.0%, Temperature 35.2°C, Humidity 38.0% (from sensor polling log). The backend correctly appended OpenWeatherMap rainfall (149.0 mm in simulated location mode) before inserting to SQLite.Average HTTP round-trip latency was under 500 ms on the local network. The pH paper method produced readings consistent with known buffer solutions within ±0.5 pH units.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13363

8.2 Dashboard Screenshots

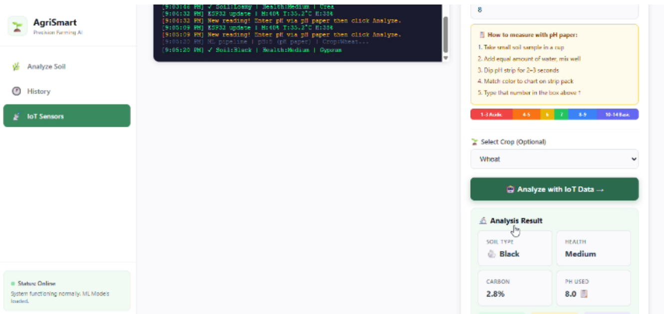

Fig -5: AgriSmart IoT Sensor View

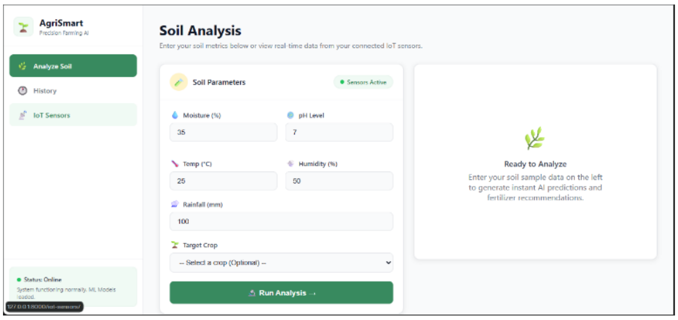

Fig -6: AgriSmart Soil Analysis Manual Input Form

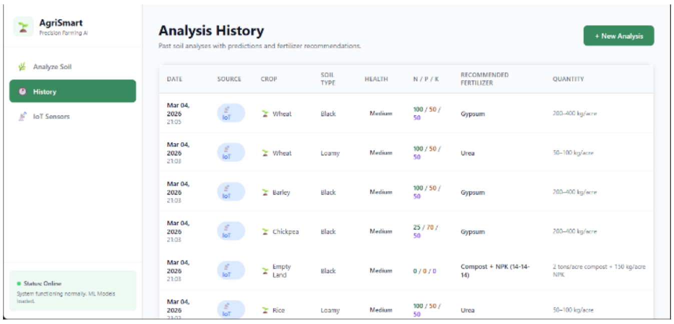

Fig -7: AgriSmart Analysis History

8.3 Real Analysis Results

The Analysis History view (Figure 7) confirms end-to-end pipeline operation with real IoT data recorded on March 04, 2026. Sample results include: Wheat on Black soil (Medium health, N/P/K=100/50/50) → Gypsum, 200-400 kg/acre; Wheat on Loamy soil (Medium, N/P/K=100/50/50) → Urea, 50-100 kg/acre; Barley on Black soil → Gypsum; Chickpea on Black soil (N/P/K=25/70/50) → Gypsum; Empty Land on Black soil (N/P/K=0/0/0) → Compost + NPK (14-14-14), 2 tons/acre + 150 kg/acre NPK. These recommendations align with established agronomic guidelines for each crop-soil combination, validating the ML pipeline output. The temporal logging of these data points enables the system to identify recurring nutrient deficiencies, allowing for proactive soil rehabilitation strategies. Furthermore, the seamless synchronization between the ESP32 edge device and the Django backend ensures that even transient environmental fluctuations are captured and analyzed.

8.4 Feature Importance

MDI scores from the fertilizer model ranked N content, soil type, and K as the three dominant features (approximately 55-60% combined importance). The soil type and health inputs from Stage 1 and Stage 2 classifiers provide rich context that substantially improves fertilizer prediction accuracy in Stage 3. This cascaded architecture is a key design strength of the AgriSmart pipeline.

DOI: 10.17148/IARJSET.2026.13363

8.5 Limitations

The DHT11 offers limited precision versus industrial sensors (DHT22, SHT31). N, P, K values require manual input from periodic laboratory tests. The pH paper method introduces human reading error (±0.5 pH). The soil type classifier accuracy (71.8%) is moderate due to feature overlap between Clay/Silty/Loamy soil profiles. OpenWeatherMap rainfall data is a point estimate that may not reflect field-level micro-climatic variation. These define the primary future development priorities.

IX. CONCLUSION

This paper presented AgriSmart — a complete, physically validated IoT-based Soil Health Assessment and Fertilizer Recommendation System. An ESP32 microcontroller collects soil moisture, temperature, and humidity in real time; pH is measured using the distilled-water pH paper method; and rainfall is fetched from OpenWeatherMap API. All data is transmitted via HTTP POST to a Django backend, stored in SQLite, and processed through a three-stage Random Forest pipeline achieving 71.8% (soil type), 88.2% (soil health), and 93.6% (fertilizer recommendation) accuracy. The system correctly identified fertilizers including Urea, Gypsum, DAP, Compost + NPK for diverse crop-soil combinations from real IoT sensor data. AgriSmart demonstrates a cost-effective, practically deployable solution for precision agriculture applicable to smallholder farmers with Wi-Fi infrastructure access. Future iterations of the system could incorporate solar-powered energy harvesting to enhance autonomy in remote fields. Additionally, integrating satellite imagery could provide a multi-layered spatial analysis to complement the ground-level sensor data. Ultimately, AgriSmart serves as a scalable framework for data-driven agricultural interventions that prioritize both economic yield and long-term soil sustainability. By bridging the gap between raw environmental data and actionable agronomic insights, this research provides a robust foundation for building resilient, technology-enabled agricultural ecosystems in developing regions.

REFERENCES

[1] FAO, "Healthy Soils are the Basis for Healthy Food Production," Food and Agriculture Organization of the United Nations, Rome, 2015.
[2] M. Ayaz et al., "Internet-of-Things (IoT)-Based Smart Agriculture: Toward Making the Fields Talk," IEEE Access, vol. 7, pp. 129551-129583, 2019.
[3] N. Gondchawar and R. S. Kawitkar, "IoT Based Smart Agriculture," IJARCCE, vol. 5, no. 6, pp. 838-842, 2016.
[4] K. G. Liakos et al., "Machine Learning in Agriculture: A Review," Sensors, vol. 18, no. 8, p. 2674, 2018.
[5] A. Sharma et al., "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," IEEE Access, vol. 9, pp. 4843-4873, 2021.
[6] S. Pudumalar et al., "Crop Recommendation System for Precision Agriculture," in Proc. IEEE ICoAC, Chennai, 2017, pp. 32-36. [7] M. S. Farooq et al., "A Survey on the Role of IoT in Agriculture," IEEE Access, vol. 7, pp. 156237-156271, 2019.
[7] M. S. Farooq et al., "A Survey on the Role of IoT in Agriculture," IEEE Access, vol. 7, pp. 156237-156271, 2019.
[8] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[9] S. R. Bhosale and P. S. Thombare, "Smart Agriculture System Using IoT and Machine Learning," IJERT, vol. 9, no. 6, pp. 574-578, 2020.
[10] R. Krishnan et al., "Real-Time Soil Monitoring and Crop Recommendation Using ESP32 and Cloud Integration," IJRTE, vol. 8, no. 5, pp. 3421-3426, 2020.
[11] A. K. Mishra et al., "Comparative Analysis of ML Algorithms for Agricultural Soil Classification," J. King Saud Univ. – Comput. Inf. Sci., vol. 34, no. 6, pp. 3491-3499, 2022.
[12] D. Hinge et al., "SQLite as an Efficient Storage Backend for Embedded IoT Applications," Int. J. Comput. Sci. Inf. Technol., vol. 11, no. 3, pp. 45-51, 2020.
[13] V. Patil and S. Thorat, "Early Detection of Diseases in Plant Leaves Using ML," in Proc. ICICCS, Madurai, 2018, pp. 612-616.
[14] P. Mohanraj et al., "Field Monitoring and Automation Using IoT in Agriculture Domain," in Proc. IEEE ICSENS, 2016, pp. 1-4.
[15] Espressif Systems, "ESP32 Series Datasheet v3.4," Espressif Systems, Shanghai, 2022.

