Power System Protection and Control
ISSN:1674-3415

A HYBRID INTERNET OF THINGS AND MACHINE LEARNING APPROACH FOR A USER-FRIENDLY PRECISION AGRICULTURE FRAMEWORK.

Punya Prabha V[1], M.D. Nandeesh[2], Tejaswini S[3], Jaseema Feroze Mohideen[4]

[1]Department of Electronics and Communication Engineering, Ramaiah Institute of Technology, Bengaluru, India. Email: punya.v@msrit.edu
[2]Department of Electronics and Instrumentation Engineering, Ramaiah Institute of Technology, Bengaluru, India. Email: mdnandeesh@msrit.edu
3Department of Medical Electronics and Engineering, Ramaiah Institute of Technology, Bengaluru, India. Email: tejaswini.s@msrit.edu
[4]Department of Electronics and Communication Engineering, Dayananda Sagar College of Engineering, Bengaluru, India. Email: jaseemaferoze725@gmail.com

Received: 09/05/2026 **Revised:** 12/06/2026 **Accepted:** 13/07/2026

***ABSTRACT*:**

Objectives: To develop an IoT and Machine Learning (ML) based decision-making framework that enables farmers to make informed decisions through data-driven approaches on the suitable type of crop to be planted, corresponding to respective soil conditions. Additionally, a Nutrient management approach to maintain soil health and optimize healthy yields.

Methods: Two primary models have been implemented: a Crop Recommendation model that suggests crops suitable for the respective soil type by analysing soil health using real-time crucial parameters such as nitrogen (N), phosphorus (P), potassium (K), pH, rainfall, temperature, and humidity obtained via a NPK sensor, and a Nutrient Sufficiency Classification model that analyses the sensed nutrient content and classifies nutrient sufficiency (low, sufficient, high) for N, P, and K. The output of the Nutrient Sufficiency Classification model is linked to a rule-based feedback system that provides suggestions or solutions for the detected shortcomings to improve soil health.

Findings: The Crop Recommendation model demonstrated an accuracy of 90.35% and the Nutrient Sufficiency Classification model exhibited an accuracy of 85.12%. Soil parameters analysed by the model, which include: nitrogen (N), phosphorus (P), potassium (K), pH, Rainfall, Temperature, and humidity, were either manually given as input through the web-based interface or a combination of real-time sensing and third-party APIs, including Google and weather APIs, that auto-populated the input fields in the user-friendly HTML-based web interface. The entire system was validated for its performance. The predictive results of the ML model were displayed on the web interface.

Novelty: This study implements a practical and deployable framework with a real-time sensing IoT architecture to sense nutrient content in the soil, an ML framework for data analysis with an integrated rule-based feedback engine to auto-populate the web interface that gives crop recommendation outputs and actionable feedback in case of nutrient deficiencies.

***Keywords**: IoT, Machine Learning, Crop Recommendation, Nutrient Sufficiency Classification, third-party APIs.*

INTRODUCTION

Agriculture is the cornerstone of the Indian Economy, validated by statistics that show that almost half of the nation's population is employed in the agricultural sector [1]. Despite its crucial role, it contributes to only 18% of India's Gross Domestic Product (GDP) [2] [3] as of 2024-2025. These numbers reflect a critical imbalance in the sector that needs to be addressed. This profitability gap is primarily due to the rising cost of cultivation and reduced return on investments caused by decreasing yields, which can be traced back to lack of knowledge among the farmers to facilitate the crops with the right resources and effective agronomic principles. Farmers lack the tools and facilities to recognize and treat soil degradation, the absence of essential nutrients and minerals, and the early onset of crop diseases. Poor fertilizer practices and traditional " one size fits all" methods [4].

IoT-based implementations have significantly aided farmers, a similar study where an IoT-based hydroponic monitoring framework was proposed in [5] pH, EC, and temperature sensors were interfaced with an ESP32 controller, and the data was transmitted over Wi-Fi for visualization on a web dashboard, showed a significant applicability of layered IoT architectures in agriculture. Precision farming [6] has recently bloomed into a major area of research due to its various advantages to farmers. Its primary goal is to effectively manage the resources, aiding in cost efficiency and improving the yield through informed decision-making.

1.1 Literature Review

Analysing 97 studies from 2017 to 2024, the review by Shawon et al [7] outlines significant parameters like NPK and pH of the soil, Rainfall, vegetation, etc., along with popular models such as Random Forest, SVM, Linear Regression, CNN, and LSTM. It notes a growing shift toward hybrid models and outlines key challenges, including data scarcity, inconsistent feature use, and limited model interpretability. Liu et al [8] used MODIS and Landsat images, terrain data, and soil samples to study cropping patterns and soil properties over a period of 20 years in Punjab and Pakistan. Multiple ML models were used to predict soil properties, primarily the N, P, K, PH, and electrical conductivity. Among which Random Forest Regression was found to be the best. The study shows that remote sensing and AI are effective tools for regional-scale agricultural analysis; however, the direct applicability of the findings to farms on a real-time basis is limited. This points to a research gap which our project aims to fill by enabling sensor-based, farmer-friendly technologies providing localized crop and fertilizer recommendations. Another study by Dey et al. [9] used five machine learning models for recommending agriculture and horticulture crops corresponding to the NPK Levels, soil PH, and climatic factors. The results of their study, suggested that XGBoost provided the highest accuracy (up to 99.3%). The study also reveals the potential of ML in crop analogy and the importance of specific targeted models for sites instead of general models over single or mixed crop datasets. This aligns with the goal of our IoT-based Crop Recommendation system and promotes the creation of AI-based decision support tools. Inspired by likewise applications of precision farming that integrate IOT and Machine Learning in the field of agriculture, a noteworthy study by Islam et al [10] presented an IoT-based soil nutrient monitoring and Crop Recommendations system by using real-time ML using sensors for measuring NPK, moisture, humidity, and temperature sensors. The model deployed CatBoost for crop prediction and RF with GridSearchCV for fertilizer recommendation. This work further demonstrates how real-time IoT systems can enable precision agriculture for smallholder farmers with intuitive user-centered platforms. Another recent advancement in smart Crop Recommendation includes Alam et al [11], who propose a novel soft-voting ensemble model by blending six machine learning models to predict the optimal crop, and it also takes agro-climatic attributes like NPK, temperature, humidity, pH, and rainfall. A key novelty lies in the adoption of entropy-based uncertainty quantification, where the confidence of the prediction in each case is a function of the softmax probability distributions on the classifiers. To improve interpretability further, the model also uses SHAP (SHapley Additive Explanations). Such transparency improves reliability and enhances the practical utility of the model. In order to deal with the problem of crop selection over a specific location considering different soil and climatic conditions, Cheema and Pires [12] proposed a real-time IoT Crop Recommendation system that collects data from real-time NPK, pH, temperature, and humidity sensors to recommend appropriate crops. The system utilizes 5 ML models. Decision Trees combined with AdaBoost gave the highest accuracy (98%). NodeMCU sends the data to ThingSpeak, and that is processed in Google Cloud. Crop suggestions are presented by a mobile application, and based on Pareto analysis, N, P, K, pH, and precipitation are the most contributing attributes. This research provided an operative, sensor-based, and energy-saving approach that is suitable for precision agriculture.

To aid farmers and resolve such issues, IoT-enabled frameworks have gained immense traction. Various research has been implemented [12, 13, 14] where IoT architectures with feature-specific sensors such as temperature, pH, NPK, humidity and rainfall integrated with cloud platforms and machine learning models to generate crop recommendations. However, these systems are strictly predictive and solely report acquired data to the farmers [12-14]. Interpretation of the data, whether the acquired nutrient levels are low, sufficient, or excessive. It also does not explain the agronomic implications of the measurements, nor does it recommend any corrective measures or strategies to resolve discrepancies if detected. The majority of the prior research relies on low-cost sensors (DHT11, FC-28, etc.) [13, 14, 15], which are known for their high error margins, poor repeatability and calibration drift. These research does not report accurate soil depth or the sampling protocol used, such as at what depth the sensor was inserted, the number of readings per plot, etc. Most existing frameworks also ignore critical contextual variables such as rainfall, temperature and local climatic suitability, which are essential for realistic crop recommendations. Additionally, many IoT–ML studies rely on small or Kaggle-derived datasets with limited regional representation, reducing model generalizability in real-world settings. Frameworks that support multiple

 DOI: 10.46121/pspc.54.3.30

ML algorithms, cloud deployment, and SMS alerting were implemented to improve system accessibility [16]. While this represents a significant step towards practical deployable platforms, it still has several limitations. The dataset used lacked transparent provenance, making it difficult to assess geographic bias or sample redundancy. Furthermore, remote-sensing driven soil mapping models [8] offer large-scale estimations but cannot provide plot-level, real-time nutrient intelligence required for daily decision-making by small-holder farmers. Another major weakness is the fact that the majority of the literature lacks a real-time, deployable and farmer-friendly IoT pipeline and solely focuses on classical ML classifiers being trained on pre-existing datasets [9]. Beyond these, centralized cloud-based learning pipelines are raising concerns on data privacy, communication overheads and scalability of the approach especially in cases where farm-level data are distributed across geographically diverse and resource-impoverished environments.

Federated learning has been recently investigated as a decentralized counterpart to centralized agricultural ML. A study [17] taken place in a smart farm scenario shows federated crop classification with climate and soil-based features, by training models locally at farm nodes while sharing only the model parameters. Obtaining performances close to 90% on classification, it ensured privacy and reduced reliance on centralised data aggregation. While effective at proving that decentralized learning is feasible, the approach does not include real-time nutrient sensing or an agronomic interpretation. Implementations [18] directed towards scalability of federated learning models have addressed crucial challenges in the smart agriculture sector such as non-IID farm data, intermittent connectivity, and resource-constrained edge devices. These approaches report classification accuracies exceeding 90%, while reducing communication rounds by approximately 30% and computational overhead by nearly 40% compared to centralized training.

Albeit such systems scale well and offer some amount of privacy, they focus mainly on system level optimization and distributed training efficiency casting aside the actual crop nutrient interpretation in real time or sending out localised crop and fertilizer advisory necessary for field level decision support.

Edge–federated learning [19] has also been studied for precision agriculture, by fusing the edge with decentralized model training. These methods allowed local training and inference directly on farm-level devices achieving performance within 2–5% of centralized models while significantly reducing data transmission and latency. While these have potential with real-time decision making to assist farming, existing systems concentrate primarily on architectural feasibility and distributed computation enabled by technologies or platforms, lacking attention in agronomic explainability, sensor calibration protocols and cropping/fertilization recommendations that are actionable for farmers. These ML-focused implementations rely on supervised classifiers such as SVM and Decision Tree models, demonstrating reasonable predictive capabilities but lacking real-time nutrient sensing and contextual climatic integration [20,21]. Extensive implementations of Deep learning models using CNNs, LSTM, Autoencoders, and feature-based classifiers remain entirely analytical and lack a field-deployable architecture as they operate on static datasets while lacking standardised sensor calibration and soil sampling protocols [20].

1.2 Research Gap

Despite all this extensive research in this area to use technological advancements, including the rise of AI to make agriculture more effective and profitable for farmers, existing implementations are predominantly predictive, relying heavily on low-cost sensors, limited static datasets, and non-transparent data provenance. The burden of data interpretation is solely left on the farmers, which entirely negates the purpose of simplifying agronomic decision-making for farmers. The lack of corrective or prescriptive agronomic recommendations is also observed in prior implementations. No existing framework combines the use of high-precision real-time sensing of crucial soil parameters with API based real-time climate data and a nutrient sufficiency model integrated with a feedback engine to guide the farmers in case of deficiencies into a single deployable and user-friendly framework for farmers.

Therefore, a clear research gap exists in the development of an integrated, real-time, high-precision IoT architecture that not only senses and predicts but also interprets nutrient sufficiency and provides actionable agronomic recommendations. To help farmers make informed decisions and improve their crop yield, this study proposes an integrated framework of a real-time IoT architecture combined with a machine learning (ML) implementation with the purpose of addressing various limitations exhibited by prior implementations. First, the proposed IoT architecture integrates a high-precision, real-time sensor to fetch crucial soil nutrient parameters. These parameters were fed into a dual pipeline with two models: A Crop Recommendation model and a second Nutrient Sufficiency Classification model that analyses the sensed nutrient content and classifies nutrient

DOI: 10.46121/pspc.54.3.30

sufficiency (low, sufficient, high) for N, P, and K. This makes data interpretation very effective and simple for the farmers. The output of the Nutrient Sufficiency Classification model is linked to a rule-based feedback system that provides suggestions or solutions for the detected shortcomings to improve soil health. Finally, this model is made lightweight to be integrated via the cloud with a user-friendly website and an API framework to get real-time climatic information. To cover critical gaps observed in this field, this high-precision, lightweight, real-time and deployable IoT pipeline has been developed.

1.3 Our Contribution

An NPK sensor is used to gather nutrient data such as Nitrogen (N), Phosphorus (P), and Potassium (K) content in soil. The sensed data is then transmitted to an Arduino microcontroller via the RS-485 protocol to route the input data to be fed into an ML model capable of making the required informed decisions. The results are displayed in the user-friendly web interface. A key contribution of this work is the dual-model ML core, where a Crop Recommendation model is coupled with a Nutrient Sufficiency Classification model and a rule-based feedback engine, enabling direct agronomic interpretation and actionable recommendations rather than standalone predictions .The output of the Nutrient Sufficiency Classification model (low, high, sufficient) for N, P, and K content in the soil, is linked to a rule-based suggestion engine that can map the predicted deficiencies to implementable recommendations to improve soil conditions. A high-precision NPK sensor is used to fetch N, P, K, and pH values from the soil. Various soil samples were tested to determine nutrient sufficiency and recommend a crop for the corresponding soil type. The dataset used to train these models was manually curated using credible sources for agronomic data and merged with pre-existing datasets available on Kaggle. To obtain the specific weather data for the respective location of the farmer, the location is either manually entered or retrieved using the Google Geocoding API, integrated at the application layer, to enhance efficiency and precision while reducing human error. The external weather API then fetches the temperature and rainfall data. These APIs autofill the required parameters, such as temperature and rainfall, in the web interface that feeds the data to the ML model. To illustrate the quality and relationships within the collected dataset, correlation heatmaps are generated and analysed for both the Nutrient Sufficiency Classification and Crop Recommendation Model. To illustrate the relationships within the collected dataset, correlation heatmaps are generated and analysed for both the Nutrient Sufficiency Classification and Crop Recommendation Model.

METHODOLOGY

A multi-layered, comprehensive, and informed decision-making framework, as summarized in Figure 1, has been curated using real-time sensors, local climate data, and leveraging machine learning for efficient predictive analysis. In the perception layer, real-time data acquisition through real-time sensing is executed using an NPK sensor to retrieve N, P, K, and pH values from the soil samples obtained from various agricultural fields. The soil sensor was inserted at a consistent depth of 10–15 cm (topsoil layer), following standard agronomic sampling practice. For each plot, three readings were taken and averaged to reduce noise and improve reliability. The sensor transmitter outputs the data as per the RS-485 protocol, which is then converted from a differential communication interface to TTL-level serial signals through the MAX485 module and received by the Arduino Nano microcontroller that uses the Modbus RTU protocol. After connecting the sensor, the Arduino sends the data in JSON-formatted packets to the farmer's local device through UART, which triggers a client application that parses the JSON package received in the application layer on a local device. On that note, the application layer on the local device, after parsing the JSON package, builds the corresponding HTTP requests by invoking the API of geolocation, weather, and third-party sensor data. The live data received from the sensors, network, and third-party APIs are automatically displayed on the web interface coded in HTML, which auto-populates the updated live data for users and is given as input to the ML models. This study implements two primary ML models in the processing layer: Crop Recommendation and another Nutrient Sufficiency Classification model. The Nutrient Sufficiency Classification model output is directly linked with a rule-based feedback mechanism to address any observed shortcomings with respect to soil health. In line with prior studies that demonstrated the effectiveness of RF in handling agricultural datasets, the proposed framework employs RF-based models for both implementations. Both models are trained using a comprehensive dataset that is a combination of a pre-existing dataset publicly available on Kaggle, merged with other credible agronomic sources.

DOI: 10.46121/pspc.54.3.30

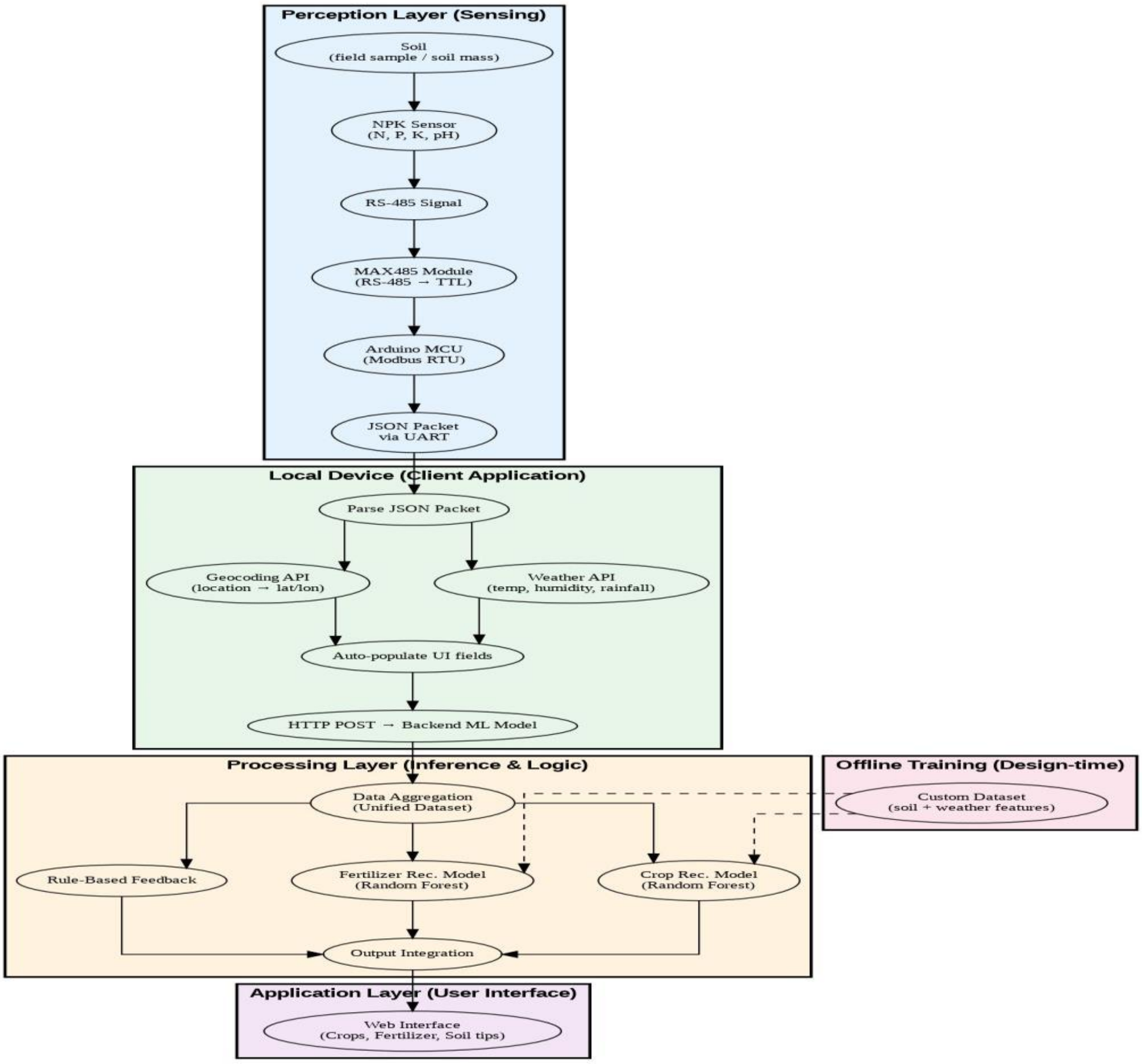


Fig. 1 End-to-end IoT and ML pipeline for smart farming-enabled decision framework

2.1 Perception layer

This is the foundational layer of the IoT architecture. The NPK sensor for in situ acquisition of the required data, including N (mg/kg), P (mg/kg), K (mg/kg), and pH (0-14) values from soil samples. All soil and environmental sensors were, calibrated as recommended by manufacturers before recording the data. NPK and pH sensors in soil were calibrated with standard reference solutions, and temperature and humidity sensors were cross-validated with laboratory-grade instruments. To reduce noise and drift generated in the sensors, each sensor value was measured multiple times per sampling point and averaged before further handling. The soil sensor was inserted at a consistent depth of 10–15 cm (topsoil layer), following standard agronomic sampling practice. For each plot, three readings were taken and averaged to reduce noise and improve reliability. The schematic diagram of the implemented IoT architecture is shown in Figure 2. The MAX485 TTL–RS-485 transceiver model enables communication from the sensor to the Arduino microcontroller. The data transmitted by the sensor in the form of Modbus RTU frames is transported to the physical layer via an RS-485 interface, which is suitable for long-distance transmission. To facilitate communication with the Arduino Microcontroller, a serial communication protocol, Modbus RTU, is implemented via the MAX485 transceiver to convert the incoming differential signal into 0–5 V TTL signals. In this system, the Arduino behaves as the master for the Modbus RTU protocol implemented for data exchange with the NPK sensor that acts as a slave using the function code 0x03. The received raw 16-bit values obtained from the sensor are stabilized using a moving average filter by the Arduino. These readings are converted into meaningful agronomic units and

DOI: 10.46121/pspc.54.3.30

encapsulated into a lightweight JSON packet to be sent over the UART interface to the client device.

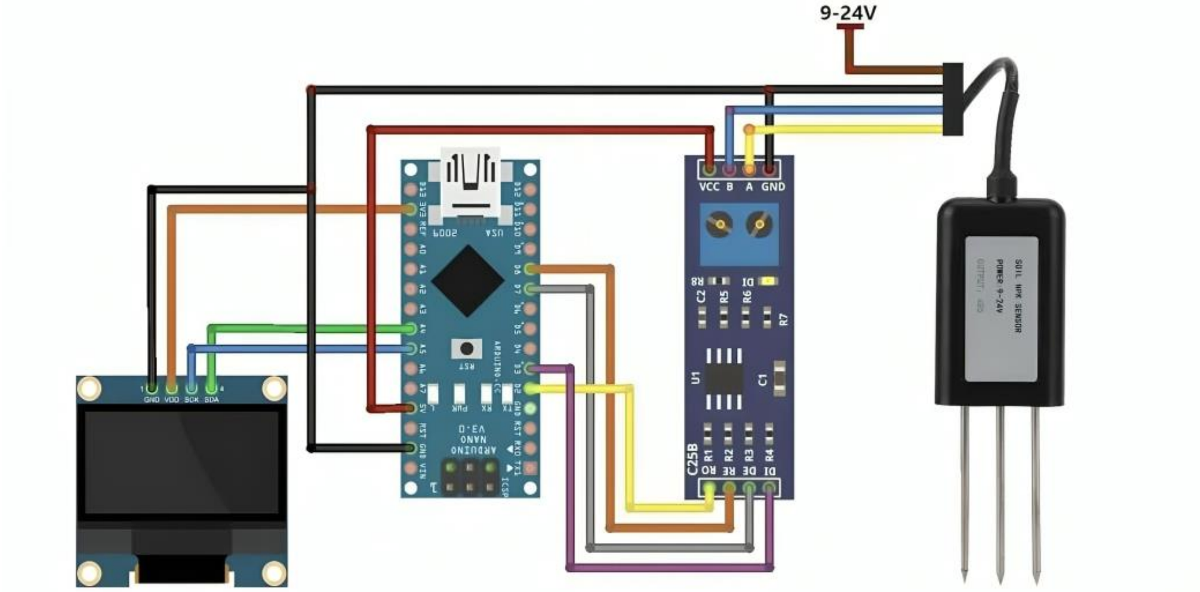


Fig. 2 Schematic Representation of the perception layer: NPK sensor, MAX485 interface, Arduino Nano, and OLED display. B. Network Layer

2.2 Network layer

The network layer links the data acquisition layer and the client device in the application layer with primary backend interfaces crucial for the transmission of data and processing via the backend ML interfaces. The acquired values of N, P, K, and pH from the NPK sensor are encapsulated into a lightweight and easy-to-parse JSON packet and transmitted via the UART interface to the client device. JSON formats are platform-independent. The receiving of these packets at the client end executes external third-party API's such as OpenWeatherMap API to complete the JSON packets with environmental data such as circulation temperature, rainfall, and humidity. Google Geocoding API is also simultaneously called to retrieve the latitude (lat) and longitude (lon) coordinates of the farmer's location. The final data vector is encapsulated into a structured JSON payload and transmitted using HTTPS POST requests for the cloud-based backend ML models. Retry mechanisms and timeout schedules are used to keep the system robust in rural or unreliable networks, so that partial or invalid data does not disrupt the process. ML model outputs (crop recommendation, nutrient classification) sent back in JSON format. The client parses these results and forwards them to the application layer (Web UI) for visualization. Climatic parameters obtained through third-party weather and location-based APIs are subject to spatial resolution constraints and intermittent availability. While such APIs may not fully capture micro-climatic variations at the plot level, they provide sufficiently reliable contextual information for crop-level decision support in the proposed framework.

2.3 Processing Layer

2.3.1 Dataset Curation and Pre-processing

Various published literature and research articles show that RF is used as the primary model to train agricultural data. Agriculture data sets often have a nonlinear structure with restricted sample sizes. They are prone to noise and missing values. RF handles these challenges efficiently and reduces the risk of overfitting by training on bootstrapped samples. Hence, this study employs an RF model that is seamlessly integrated with the IoT architecture. Two separate RF classifiers were trained, one for Crop Recommendation and one for the Nutrient Sufficiency Classification model. The dataset, as shown in Figure 3, was curated by merging structured agronomic data from multiple credible sources, including the FAO Soil Fertility and Plant Nutrition Data (UN Food and Agriculture Organization), ICAR–CRIDA (Central Research Institute for Dryland Agriculture) for crop–soil suitability guidelines, the ICAR-KVK Ballari crop-weather calendars, IMD, various peer-reviewed agricultural journals [12], and government soil testing reports were also analysed. The data collected include’s multiple growing seasons and depict soil and climatic conditions that are generally found in the semi-arid and tropical regions of India. Data sources were chosen for agronomic validity and applicability in smallholder cropping

DOI: 10.46121/pspc.54.3.30

systems with a focus on crops commonly grown under rain-fed and irrigated conditions The obtained data was combined with the dataset available on Kaggle to increase diversity and acquire a larger training database. The required input features for both models, N (mg/kg), P (mg/kg), K (mg/kg), pH, temperature (°C), humidity (%), and rainfall (mm), were obtained, and their units were standardized. Missing values were treated by simple mean or median imputation. Apart from imputation, raw data was validated by exclusion of physically implausible observations (eg negative nutrient concentrations or out of range pH values). Repeated and inconsistent records were removed, and distributions of features were scrutinized to ensure they were in accordance with agronomically reasonable ranges before model fitting. Proper imputation of missing labels was substituted based on the rules of agronomic suitability or omitted if unsure. To ensure compatibility with all platforms, the final dataset was UTF-8 encoded. In total, the dataset consisted of 2200 entries. The data consisted of 22 unique crop classes, containing mild class imbalance, which was rectified through randomly under-sampling while training. Class imbalance between the minority categories of crops was also observed. This issue was addressed by applying the Synthetic Minority Over-sampling Technique (SMOTE) with k = 5 nearest neighbours. However, the resampling process was only carried out in training folds during cross-validation to prevent leakage of information with validation or test sets. When working with the Nutrient Sufficiency Classification model, there was no need for resampling due to the data distribution's uniform nature. The dataset was split in an 80:20 ratio for training and testing subsets using a stratified split to preserve class distribution across 22 crop categories.

Crop_recommendation

N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.87974371	82.00274423	6.502985292000000	202.9355362	rice
85	58	41	21.77046169	80.31964408	7.038096361	226.6555374	rice
60	55	44	23.00445915	82.3207629	7.840207144	263.9642476	rice
74	35	40	26.49109635	80.15836264	6.980400905	242.8640342	rice
78	42	42	20.13017482	81.60487287	7.628472891	262.7173405	rice
69	37	42	23.05804872	83.37011772	7.073453503	251.0549998	rice
69	55	38	22.70883798	82.63941394	5.70080568	271.3248604	rice
94	53	40	20.27774362	82.89408619	5.718627178000000	241.9741949	rice
89	54	38	24.51588066	83.53521630000000	6.685346424	230.4462359	rice
68	58	38	23.22397386	83.03322691	6.336253525	221.2091958	rice
91	53	40	26.52723513	81.41753846	5.386167788	264.6148697	rice
90	46	42	23.97898217	81.45061596	7.50283396	250.0832336	rice
78	58	44	26.80079604	80.88684822	5.108681786	284.4364567	rice
93	56	36	24.01497622	82.05687182	6.98435366	185.2773389	rice
94	50	37	25.66585205	80.66385045	6.94801983	209.5869708	rice
60	48	39	24.28209415	80.30025587	7.0422990690000000	231.0863347	rice
85	38	41	21.58711777	82.7883708	6.2490506560000000	276.6552459000000	rice
91	35	39	23.79391957	80.41817957	6.970859754	206.2611855	rice
77	38	36	21.8652524	80.1923008	5.953933276	224.5550169000000	rice
88	35	40	23.57943626	83.58760316	5.85393208	291.2986618000000	rice
89	45	36	21.32504158	80.47476396	6.442475375	185.4974732	rice
76	40	43	25.15745531	83.11713476	5.070175667	231.3843163	rice
67	59	41	21.94766735	80.97384195	6.012632591	213.3560921	rice
83	41	43	21.0525355	82.67839517	6.254028451	233.1075816	rice
98	47	37	23.48381344	81.33265073	7.375482851	224.0581164	rice
66	53	41	25.0756354	80.52389148	7.778915154	257.0038865	rice
97	59	43	26.35927159	84.04403589	6.2865001760000000	271.3586137000000	rice
97	50	41	24.52922681	80.54498576	7.070959995	260.2634026	rice

Fig. 3. Self-curated dataset of the Crop Recommendation model

Although the dataset represents enough variation in soil nutrient levels, crop species and climate to test the proposed framework, all agroecological zones or cultivars of crops may not be represented. Therefore, the dataset can be scaled down to local crop recommendation rather than universal agricultural practices and conditions.

DOI: 10.46121/pspc.54.3.30

Power System Protection and Control
ISSN:1674-3415

2.3.2 Crop Recommendation Model

The RF model for Crop Recommendation was trained on N (mg/kg), P (mg/kg), K (mg/kg), pH, temperature (°C), humidity (%), and rainfall (mm). After this data was cleaned and standardized, they were directly label encoded into integers of the corresponding target crop, as shown in Figure 4. The climatic variables were either manually entered or obtained via the Google Geocoding API. Hyperparameter tuning primarily focused on the training subset included n_estimators, max_depth, min_samples_split, and max_features. RandomizedSearchCV was used to efficiently explore the model's domain space. This was followed by GridSearchCV for terminal fine-tuning. 5-fold cross-validation was also implemented to improve generalization and hyperparameter selection. The hyperparameter tuning and cross-validation were performed exclusively on the training set, while the final model performance metrics were computed, including 300 decision trees with the maximum depth of each tree set at 22. The minimum number of samples required to split a node was 4, and the number of features considered for splitting followed the 'sqrt' strategy used along with a consistent seed of 42. The model's predicted crop label was then transmitted to the client application in the simulation environment for visualization and farmer decision-making.

	Unnamed: 0	Crop	N	P	K
0	0	rice	80	40	40
1	3	maize	80	40	20
2	5	chickpea	40	60	80
3	12	kidneybeans	20	60	20
4	13	pigeonpeas	20	60	20

Fig.4 Label encoding of the crops into corresponding integer values

2.3.3 Nutrient Sufficiency Classification model integrated with the rule-based suggestion engine

The Nutrient Sufficiency Classification model focuses on analysing soil health by assessing the concentrations of N, P, and K and soil pH. The input features are standardized and cleaned, after which appropriate threshold-based classification was implemented for the Nutrient Sufficiency Classification model, where the raw nutrient concentration values of the soil are categorized into classes such as Low, Sufficient, and High based on standard agronomic thresholds. This threshold-based classification forms the basis for subsequent decision-making in the rule-based recommendation engine. The RF model employed for nutrient classification was trained similarly to the model used for Crop Recommendation. The resulting classification was displayed on an HTML-based Web interface that automatically triggered the rules-based suggestion engine to display the appropriate feedback. Each deficiency is mapped to a predefined suggestion string.

Example rules: IF N_status = Low → Recommend Urea or Ammonium Sulphate + dosage. IF P_status = Low → Recommend Single Superphosphate (SSP) + dosage.

To ensure reproducibility, all experiments were executed with a fixed random seed (42), using Python 3.10, scikit-learn 1.3.0, and NumPy 1.25.0 for both the models.

The fixed nutrient sufficiency thresholds in this study were based on accepted agronomic recommendations and national soil-fertility standards. These limits give consistent nutrient status interpretations over a variety of farming systems. Although nutrient requirements may differ by crop variety, soil type, and regional practices, the

DOI: 10.46121/pspc.54.3.30

use of standardized cut-offs is justified to make sure our decisions are interpretable and comparable across sites. Adaptive and location specific thresholding is highlighted as an important extension for future development.

2.3.4 Application Layer

The application layer is a simple HTML-based Web User Interface (Web UI). The parameters are either manually entered or received from real-time sensing. Third-party APIs are also used to fill in other parameter fields in the UI. The SIL architecture is integrated with this layer and receives the data vector in JSON format. This layer is also used to visualize the outputs from the ML models and the rule-based suggestion engine, as the JSON payload vector is received from the models.

RESULTS AND DISCUSSION

This section displays the outputs of the ML models, specifically the Crop Recommendation and Nutrient Sufficiency Classification models, along with the integrated rule-based suggestion engine, as shown on the web-based user-friendly interface. The performance of each model has been analysed, focusing on the feature correlation heatmaps to highlight the interdependencies between variables, and dataset trend visualizations are provided to illustrate the variability of soil nutrient concentrations across the collected samples.

3.1 Model Evaluation

The performance of both models was evaluated on the held-out 20% test set. The Crop Recommendation model demonstrated an accuracy of 90.35% and the Nutrient Sufficiency Classification model exhibited an accuracy of 85.12%. The models were evaluated using standard metrics as shown in Table 1.

Table 1. Evaluated performance metrics for both models

Models	Accuracy (%)	Precision	Recall	F1-score
Crop Recommendation	90.35	91.1	89.8	90.4
Nutrient Sufficiency Classification	85.12	86.0	84.2	85.0

The error trends for both the Crop Recommendation and Nutrient Sufficiency Classification models are exhibited. The variability observed is primarily associated with the use of the diverse heterogeneous dataset that has been used to train the model. The steady decline in the error rate shows efficient training. The promising performance metrics combined with these trends show the model's ability to sufficiently generalize. The heatmaps shown in Figure 5 and Figure 6 show the correlation between variables that are observed in each of the heterogeneous datasets used to train both models. The Crop Recommendation model heatmap shows a strong positive correlation between P and K (0.74), while N is negatively correlated with most features. The climatic factors also have low correlation levels, denoting that these are independent of the nutrient concentration and often have minimal effect on the nutrient levels of the soil. A similar trend is seen between the N, P, and K concentrations in the soil for the Nutrient Sufficiency Classification model.

DOI: 10.46121/pspc.54.3.30

Power System Protection and Control
ISSN:1674-3415

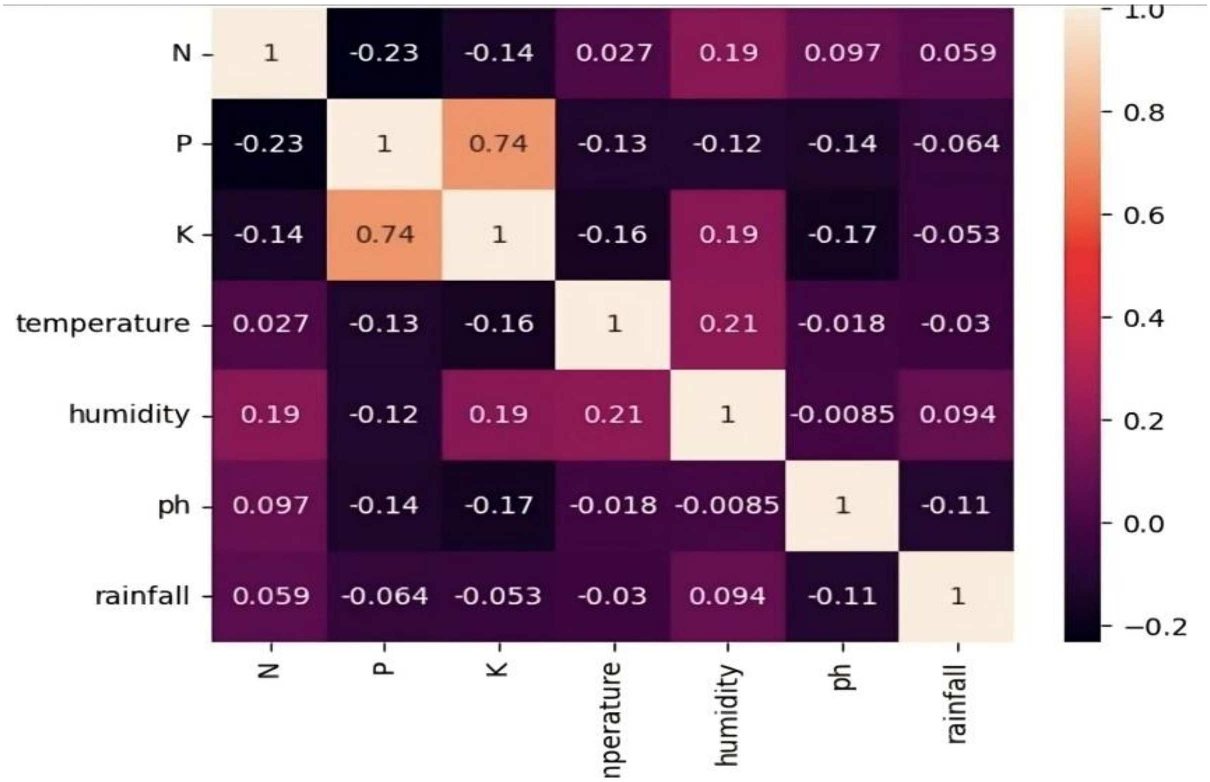


Fig. 5. Heatmap showing the correlation between variables in the Crop Recommendation model

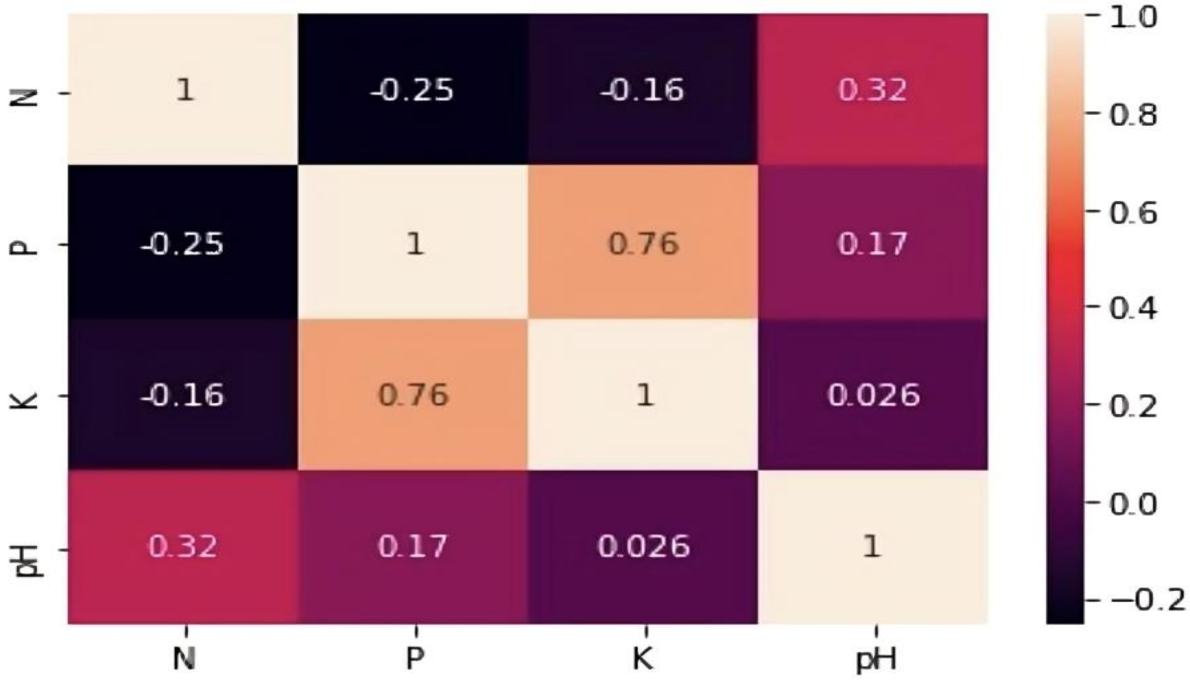


Fig. 6. Heatmap showing the correlation between variables in the Nutrient Sufficiency Classification model

3.2 Web UI Output

The HTML-based web interface is used as the initial window for manual data entry, or the required parameters are auto-filled using real-time sensing. Here, the N, P, and K values are received from the NPK sensor, and the location and climatic information are obtained via third-party APIs such as Google Geocoding API and

DOI: 10.46121/pspc.54.3.30

OpenWeatherMap API. The web UI of the Crop Recommendation model serves as the application layer endpoint where the inputs that are received in the interface shown in Figure 7 are transmitted to the RF model, which displays the output of the suitable crop to be planted in accordance with the soil type, as shown in Figure 8.

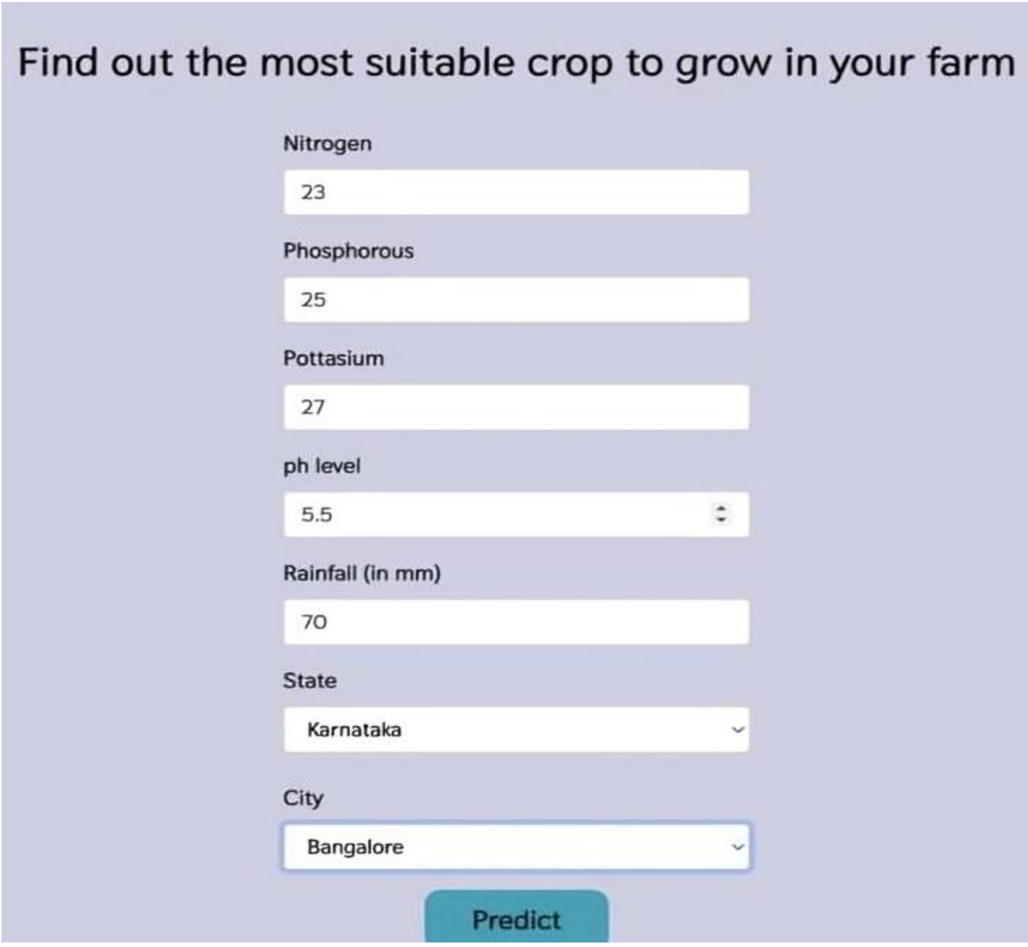


Fig. 7. Input Interface for manual or auto-populating soil and environmental parameters for the Crop Recommendation model

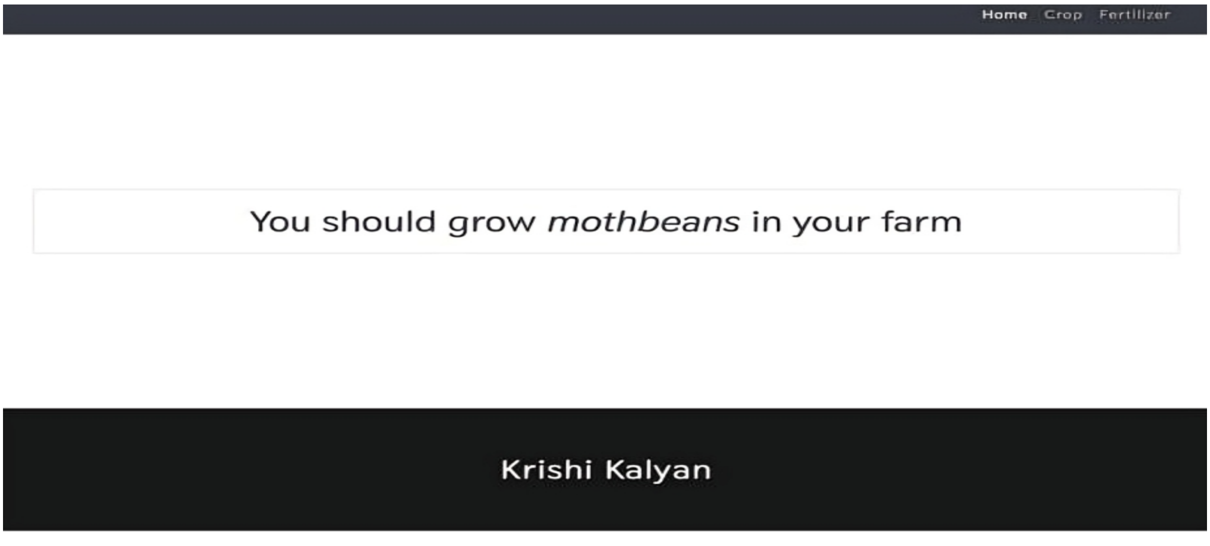
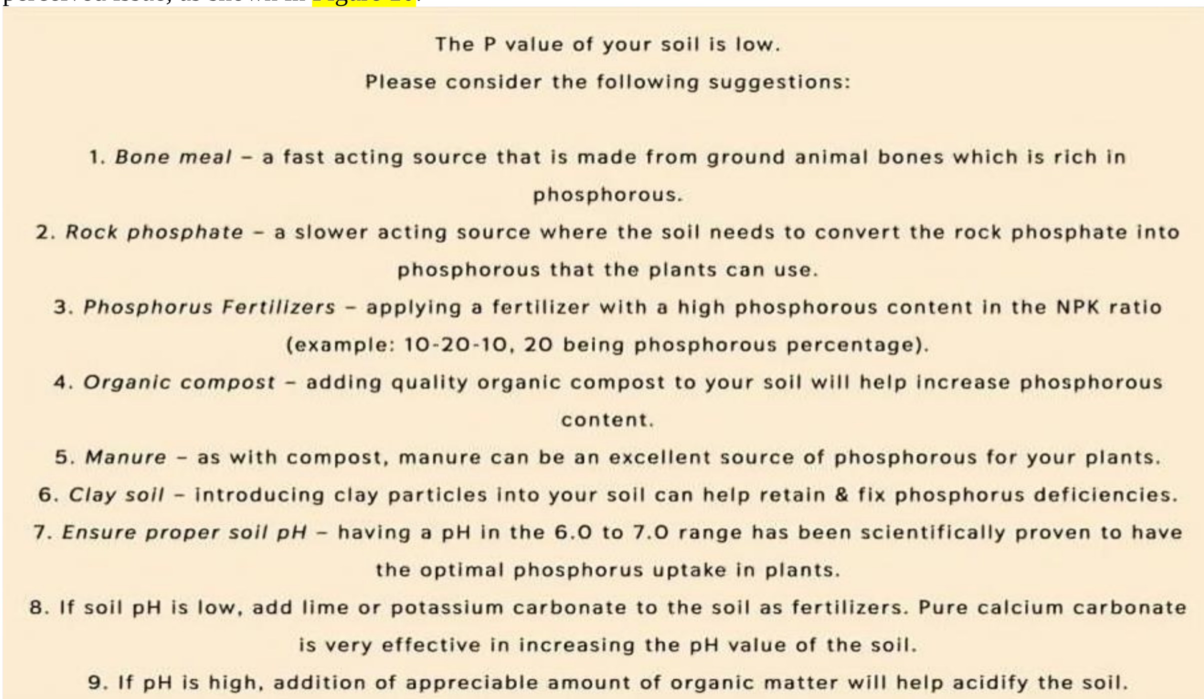


Fig. 8. Output Interface showing suitable crop recommended for cultivation for the corresponding soil type.

Fig. 9. Input Interface of the nutrient classification model integrated with an NPK sensor for real-time acquisition of values

Similarly, the Nutrient Sufficiency Classification model also has an interface, as shown in Figure 9, where N, P, and K concentrations received from the sensor are analysed. The output is classified into low, sufficient, and high accordingly. Following this, an integrated rule-based suggestion engine also gives feedback or a solution to the perceived issue, as shown in Figure 10.


The P value of your soil is low.
Please consider the following suggestions:

1. *Bone meal* – a fast acting source that is made from ground animal bones which is rich in phosphorous.
2. *Rock phosphate* – a slower acting source where the soil needs to convert the rock phosphate into phosphorous that the plants can use.
3. *Phosphorus Fertilizers* – applying a fertilizer with a high phosphorous content in the NPK ratio (example: 10-20-10, 20 being phosphorous percentage).
4. *Organic compost* – adding quality organic compost to your soil will help increase phosphorous content.
5. *Manure* – as with compost, manure can be an excellent source of phosphorous for your plants.
6. *Clay soil* – introducing clay particles into your soil can help retain & fix phosphorus deficiencies.
7. *Ensure proper soil pH* – having a pH in the 6.0 to 7.0 range has been scientifically proven to have the optimal phosphorus uptake in plants.
8. If soil pH is low, add lime or potassium carbonate to the soil as fertilizers. Pure calcium carbonate is very effective in increasing the pH value of the soil.
9. If pH is high, addition of appreciable amount of organic matter will help acidify the soil.

Fig. 10. Rule-based mechanism output showing recommendations for low phosphorus levels

DOI: 10.46121/pspc.54.3.30

To assess the statistical stability of the proposed models, performance was evaluated across multiple training–testing splits. The Crop Recommendation model achieved a mean accuracy of 90.35% ± 1.8%, while the Nutrient Sufficiency Classification model achieved 85.12% ± 2.3%, indicating consistent performance with limited variance across runs.

The investigation of misclassified cases indicated that the errors were mainly due to boundary problems or boundary cases where soil nutrient values were close to predefined threshold limits (e.g., marginal nitrogen sufficiency). Further inaccuracies were noticed at very fast variations of the climate, especially if rainfall or humidity values greatly changed over short periods. These examples illustrate the natural uncertainty in real-world sensing scenarios and underscore the importance of further consideration of adaptive thresholding and temporal smoothing in future system revisions.

Preliminary feature sensitivity analysis showed that excluding primary soil nutrients (N, P, K) reduced crop recommendation accuracy by approximately 6–8%, whereas excluding individual climatic features resulted in a smaller reduction of 2–3%, indicating the dominant role of soil chemistry in the proposed framework

3.1 Comparative Analysis and Validation of Results

The proposed framework has been implemented, prioritising practical deployability and a user-friendly approach to help farmers make informed agronomic decisions to improve yield and maximise profits. It can be observed that the proposed IoT–ML framework surpasses several limitations of previous implementations and research in the field. The accuracy of 90.35% for crop recommendation and 85.12% for the Nutrient Sufficiency Classification model is consistent with the upper range of Random Forest performance reported in various literature, including the systematic review by Shawon et al. [7], where it can be seen that RF typically achieves an accuracy of 80–95% across diverse agricultural datasets. Additionally, the findings of Singh et al. [16] show a clear dominance of tree-based ensemble methods, an RF-driven hybrid, achieving the highest accuracy (94%) compared to the performances of various baseline models like KNN and other Deep learning models such as CNN and LSTM. Shawon et al. (2025), which systematically reviewed 162 peer-reviewed studies, and the Scientific Reports survey by Singh & Sharma [16], which benchmarked over 30 ML and Deep learning-based crop recommendation frameworks, have concluded similar significant limitations and highlighted major gaps. It shows that most implementations suffer from limited feature selection and overreliance on climactic parameters, minimal implementation of deployable IoT frameworks, lacking real-time sensing and interpretability. Most of these black-box models have inconsistent evaluation metrics and lack a feedback mechanism to suggest any corrective measures or strategies to resolve discrepancies if detected. The proposed framework in this study generates very transparent, real-time nutrient sufficiency insights using multi-parameter soil sensing with an interpretable rule-based advisory engine, thereby demonstrating stronger practical validity and applicability for farmer-centric decision-making. IoT-based models such as Islam et al. [10] and Cheema & Pires [12] similarly report accuracies between 90–96%, supporting the suitability of RF for heterogeneous soil–climate data; however, these studies relied on low-cost sensors that introduce drift and noise, whereas the present work uses high-precision RS-485 nutrient sensing, which likely contributes to the stable performance observed. Cloud-integrated systems such as Singh and Sharma's TCRM [16] reach ~94% accuracy but lack real-time soil sensing and nutrient sufficiency interpretation, suggesting that the slightly higher accuracy achieved in this study derives from combining real-time soil data with climatic API inputs. While dataset-based analytical studies, including XGBoost models reported [9], show higher accuracies (~97–99%), these rely on clean, static datasets without sensor variability and therefore do not contradict the more realistic field-level accuracy obtained here. Remote-sensing studies such as Liu et al. [8] report strong correlations between climate, soil properties, and cropping patterns at regional scales, which complement our findings by confirming the independence of climatic variables and the nutrient–nutrient correlations observed in our heatmaps. Overall, the comparative evaluation demonstrates that the present system is both statistically consistent with existing literature and uniquely enhanced through real-time sensing, dual-model classification, and actionable feedback mechanisms, collectively supporting the validity and practical relevance of the developed framework.

The current work mainly concentrates on system design and controlled experimental validation; large-scale field tests were out of the scope. Nevertheless, the framework was designed with deployability in mind and field-level pilot testing is identified as a critical area for future research.

Power System Protection and Control
ISSN:1674-3415

CONCLUSION

This research proposes a hybrid IoT-ML framework that bridges major gaps existing in agricultural decision-support systems built to aid farmers in making informed agricultural decisions. The perception layer transmitted real-time N, P, K, and pH values from the soil to the ML models. This study implemented two RF models: a Crop Recommendation model that suggests a suitable crop to be cultivated corresponding to the soil conditions, and the Nutrient Sufficiency Classification model that detects deficiencies and also provides tailored, data-based advice on how to increase crop yield and balance the nutrients in case of deficiencies. The Crop Recommendation model demonstrated an accuracy of 90.35% and the Nutrient Sufficiency Classification model exhibited an accuracy of 85.12%. Soil parameters analysed by the model, which include: nitrogen (N), phosphorus (P), potassium (K), pH, Rainfall, Temperature, and humidity, were either manually given as input through the web-based interface or a combination of real-time sensing and third-party APIs, including Google and weather APIs, was used. The entire system was validated for its performance. The predictive results of the ML model were displayed on the web interface. Heatmaps were used to analyse the quality of the dataset and study correlations between input features. The framework is the first to unify high-precision sensing, real-time climate data, ML-based crop prediction, nutrient sufficiency interpretation, and automated corrective feedback into a single deployable pipeline suitable for field use on a user-friendly interface. This research attempted to surpass various limitations of previous research by implementing a dual pipeline framework for not only data acquisition, but also data interpretation. Sampling depth has clearly followed standard protocols as mentioned. The Nutrient Sufficiency Classification model, implemented alongside the Crop Recommendation model, provides real-time interpretability and supports efficient crop planning while addressing fertiliser overuse among farmers. An end-to-end deployment system using high precision sensors with real-time feedback was implemented to overcome shortcomings of previous research that used low-cost equipment's affecting the accuracy of obtained results. Future development may also integrate mobile app deployment, offline capabilities, or remote-sensing data to build a more comprehensive farm management ecosystem. Overall, the proposed framework contributes a novel, practical, and scalable approach to precision agriculture and has strong potential to support sustainable farming practices, improved resource management, and better yield outcomes for smallholder farmers. From a real-world deployment perspective, several limitations and challenges continue to persist which can be looked into in the context of future work. The price of high-precision soil nutrient sensors, although improving data accuracy, could limit widespread adoption among smallholder farmers; the continued venture plans to investigate hybrid architectures that harmonize selective high-accuracy sensing with periodic calibration using economical sensors to lower overall system costs. Localization continues to be a major obstacle because nutrient thresholds, crop suitability, and fertilizer recommendations vary widely among regions that can be resolved by using region-specific agronomic databases and adaptive thresholding strategies based on local soil characteristics (soil type) and weather. With respect to industry integration, seamless rollout would depend on interoperability with current agricultural extension services, fertilizer advisement platforms and the government-backed digital farming initiatives. Future extensions of this work will focus on field-level pilot deployments, mobile and SMS-based interfaces for low-connectivity environments, and integration with standardized agricultural data platforms to enable scalable and industry-ready adoption

REFERENCES

1. Bais P, Bahadur P S. Agriculture in Indian Economy and Contribution of Science and Technology. Asian Journal of Applied Science and Technology (AJAST). 2023 Jan–Mar; 7(1): 167–176. https://doi.org/10.38177/ajast.2023.7116
2. Bhoir R. Agriculture and employment: Impact and trends in India. International Research Journal of Marketing & Economics. 2025 Feb; 12(2): 50–57.
3. Trading Economics. India – GDP from Agriculture [Internet]. Available from: https://tradingeconomics.com/india/gdp-from-agriculture [cited 2025 Oct 30].
4. Sinha A, Basu D, Priyadarshi P, Ghosh A, Sohane R K. Farm typology for targeting extension interventions among smallholders in tribal villages in Jharkhand State of India. Frontiers in Environmental Science. 2022; 10: 823338. https://doi.org/10.3389/fenvs.2022.823338

5. Prabha P, Sarala S M, Suttur Sharmila C S. Robust smart irrigation system using hydroponic farming based on data science and IoT. In: Proceedings of the IEEE Bangalore Humanitarian Technology Conference (B-HTC); 2020. p. 1–4. https://doi.org/10.1109/B-HTC50970.2020.9297842
6. Saha S, Kucher O D, Utkina A O, Rebouh N Y. Precision agriculture for improving crop yield predictions: A literature review. Frontiers in Agronomy. 2025; 7: 1566201. https://doi.org/10.3389/fagro.2025.1566201
7. Shawon S M, Ema F B, Mahi A K, Niha F L, Zubair H T. Crop yield prediction using machine learning: An extensive and systematic literature review. Smart Agricultural Technology. 2025; 10: 100718. https://doi.org/10.1016/j.atech.2024.100718
8. Liu J, Yang K, Tariq A, Lu L, Soufan W, El Sabagh A. Interaction of climate, topography and soil properties with cropland and cropping pattern using remote sensing data and machine learning methods. The Egyptian Journal of Remote Sensing and Space Sciences. 2023; 26(3): 415–426. https://doi.org/10.1016/j.ejrs.2023.05.005
9. Dey B, Ferdous J, Ahmed R. Machine learning-based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables. Heliyon. 2024 Mar; 10(3): e25112. https://doi.org/10.1016/j.heliyon.2024.e25112
10. Islam M R, Oliullah K, Kabir M M, Alom M, Mridha M F. Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. Journal of Agriculture and Food Research. 2023; 14: 100880. https://doi.org/10.1016/j.jafr.2023.100880
11. Alam M S B, Esichaikul V, Lameesa A, Ahmed S F, Gandomi A H. An approach for crop recommendation with uncertainty quantification based on machine learning for sustainable agricultural decision-making. Results in Engineering. 2025; 26: 105505. https://doi.org/10.1016/j.rineng.2025.105505
12. Munawar Cheema S, Pires IM. A IoT based soil nutrient analysis and recommendation system for crops using machine learning. Smart Agric Technol. 2025;11:100924. doi:10.1016/j.atech.2025.100924.
13. Joshi S, Shelke MV. Soil moisture detection and crop recommendation with crop disease detection. Int J Technol Eng Arts Math Sci. 2020;3(2):26–30. doi:10.11591/eei.v9i3.xxxx.
14. Senapaty MK, Ray A, Padhy N. IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture. Computers. 2023;12(3):61. doi:10.3390/computers12030061.
15. Aarthi S, Manimegalai S, Sakthivel R. AI-based Smart Crop Recommendation System for Sustainable Agricultural Production: A Data-driven Approach to Minimize Resource Use and Maximize Yield. Madras Agric J. 2025;112:135–9. doi:10.29321/MAJ.10.701221.
16. Singh G, Sharma S. Enhancing precision agriculture through cloud based transformative crop recommendation model. Sci Rep. 2025;15:9138. doi:10.1038/s41598-025-93417-3.
17. Idoje G, Dagiuklas T, Iqbal M. Federated learning: Crop classification in a smart farm decentralised network. Smart Agric Technol. 2023;5:100277. doi:10.1016/j.atech.2023.100277.

18. Janga R, Dave R. Enhancing smart farming through federated learning: A secure, scalable, and efficient approach for AI-driven agriculture. Artif Intell Appl. 2025. doi:10.47852/bonviewAIA52025089.
19. Jagadeesan J, Nagarajan R. IoT-enabled federated learning model for crop yield prediction using machine learning and optimization algorithms in agriculture sector. ICTACT J Soft Comput. 2025;16(2):3833. doi:10.21917/ijsc.2025.0532.
20. Chunduri S, Raj H, Narendra VG. Intelligent Systems for Crop Recommendation using Machine Learning. WSEAS Trans Comput. 2025;24.2. doi:10.37394/23205.2025.24.2.
21. Latha P, Kumaresan P. Deep learning for soil nutrient prediction and strategic crop recommendations: An analytic perspective. Nat Environ Pollut Technol. 2025;24(1):B4205. doi:10.46488/NEPT.2025.v24i01.B4205.