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 4, April 2026

DOI: 10.17148/IARJSET.2026.13475

Smart Soil Sense: An IoT-Based Intelligent Crop Recommendation System Using Machine Learning for Precision Agriculture

Irshad Ahamed M[1], **Naveen D**[2], **Sowndhar B**[3], **Tharvesh Muhaideen A**[4]

Department of Electronics and Communication Engineering, E.G.S. Pillay Engineering College (Autonomous),

Nagapattinam, Tamil Nadu, India[1-4]

Abstract: Smart agricultural monitoring plays a crucial role in improving crop yield and maintaining soil health. In India, many farmers face challenges in selecting suitable crops and managing soil nutrients, often leading to excessive fertilizer usage and financial loss. This paper presents **Smart Soil Sense**, an Internet of Things (IoT) based system that analyzes soil nutrients including Nitrogen (N), Phosphorus (P), and Potassium (K) along with environmental factors such as temperature, humidity, pH, and rainfall to recommend appropriate crops. The system integrates real-time sensor data collection using ESP32 microcontroller, NPK sensors, and DHT11/DHT22 environmental sensors. Multiple machine learning algorithms were evaluated, including Decision Tree, Naive Bayes, Support Vector Machine, Logistic Regression, Random Forest, and XGBoost. Experimental results indicate that **XGBoost achieves the highest accuracy of 99.31%**, demonstrating its effectiveness for crop prediction. The proposed system also incorporates a disease detection module using convolutional neural networks (CNN) for image classification to identify plant leaf diseases and recommend appropriate pesticides. Overall, the approach offers a cost-effective and reliable solution for improving agricultural productivity and supporting data-driven decision-making in farming practices, with potential to reduce fertilizer misuse by up to 30–35% and improve crop yields through precision agriculture.

Index Terms: Smart Agriculture, Internet of Things (IoT), Machine Learning, Soil Analysis, XGBoost, Crop Recommendation, Precision Farming, NPK Sensors, ESP32.

I. INTRODUCTION

Agriculture is a key sector of the Indian economy, supporting over 58% of the population and contributing approximately 18% to the Gross Domestic Product (GDP) [1]. However, productivity remains low in many regions due to traditional farming practices and the lack of data-driven decision-making. Farmers often rely on experience rather than scientific analysis, leading to improper crop selection and excessive use of fertilizers [2].

The overuse of chemical fertilizers has resulted in soil nutrient imbalance, reduced fertility, and increased production costs. According to the Indian Council of Agricultural Research (ICAR), approximately 30–35% of applied fertilizers are wasted due to incorrect application [3]. Studies indicate that more than 60% of Indian agricultural land suffers from nutrient imbalance due to excessive use of nitrogen-based fertilizers [4]. Limited access to soil testing facilities further restricts informed decision-making, with only around 30% of farmers regularly testing their soil before planting [5].

To address these issues, this work proposes Smart Soil Sense, an IoT-based system that analyzes soil nutrients (N, P, K) and environmental factors such as temperature, humidity, pH, and rainfall to recommend suitable crops and optimal nutrient levels. The system leverages IoT technology for real-time data acquisition and employs advanced machine learning algorithms for accurate prediction.

Main Contributions

- Development of an integrated IoT-based soil monitoring system using ESP32 microcontroller and NPK sensors.
- Comparative evaluation of six machine learning algorithms for crop recommendation, demonstrating XGBoost's superior performance with 99.31% accuracy.
- Implementation of a CNN-based disease detection module for real-time plant health assessment.
- Creation of a voice-based interface to enhance usability for farmers with limited literacy.

II. RELATED WORK

Several research efforts have explored the use of Internet of Things (IoT) and machine learning techniques for smart agriculture and crop recommendation systems. Lakshmisudha et al. [6] presented an IoT-based agricultural monitoring system that provided farmers with real-time insights into soil and environmental conditions. The system focused on measuring soil moisture, temperature, and ambient humidity using sensors and microcontrollers, with data transmitted to cloud platforms for remote monitoring. However, it did not incorporate nutrient analysis or AI-based decision-making capabilities.

Suma et al. [7] proposed an automated smart irrigation system using soil moisture sensors to control irrigation based on predefined thresholds. The system demonstrated improved water efficiency and reduced manual intervention but lacked integration with weather forecasts and predictive analytics. Sujatha and Isakki [8] developed a machine learning-based crop recommendation system using Decision Trees and Random Forests but relied on static datasets rather than real-time sensor inputs.

Ramesh et al. [9] focused on soil nutrient detection using IoT-based NPK sensors for real-time monitoring of nitrogen, phosphorus, and potassium levels but lacked integration with predictive AI models for crop recommendation. Recent advancements by Khanna and Kaur [11] demonstrated the effectiveness of AI in precision agriculture through crop yield prediction and resource optimization. Rajalakshmi et al. [12] developed an integrated system combining IoT sensors with AI algorithms for holistic agricultural management.

In contrast, the proposed Smart Soil Sense system integrates IoT-based real-time data collection with advanced machine learning techniques, particularly XGBoost, to provide accurate crop recommendations, disease detection, and agricultural guidance addressing limitations observed in prior research.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. System Overview

The Smart Soil Sense system consists of four major phases: (1) data acquisition, (2) data processing, (3) machine learning-based prediction, and (4) disease detection. The system architecture integrates hardware components including an ESP32 microcontroller, NPK sensors for soil nutrient measurement, DHT11/DHT22 sensors for environmental monitoring, and a display interface for real-time feedback.

B. Dataset Description

The dataset used in this project was constructed by augmenting and combining various publicly available datasets from India, including weather and soil data. It is accessible at: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset. The dataset consists of 2,200 data points across 22 unique crop labels. The following features are included:

- N, P, K — Ratios of Nitrogen, Phosphorus, and Potassium in the soil.
- Temperature — Ambient temperature in degrees Celsius.
- Humidity — Relative humidity in percentage (%).
- pH — The pH value of the soil.
- Rainfall — Rainfall measured in millimeters (mm).

The crop labels included in the dataset are: rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mung bean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee.

C. Data Acquisition

Soil parameters, including Nitrogen (N), Phosphorus (P), and Potassium (K) are collected using NPK sensors interfaced with an ESP32 microcontroller operating at 3.3V. Environmental parameters such as temperature and humidity are measured using DHT11/DHT22 digital sensors with ±0.5°C temperature accuracy and ±2–5% humidity accuracy. Rainfall data is obtained through integration with weather APIs to provide real-time climatic information.

D. Data Processing and Feature Engineering

The collected sensor data is transmitted wirelessly to a server using the ESP32's built-in Wi-Fi capabilities (IEEE 802.11 b/g/n). Data preprocessing includes normalisation, outlier detection, and missing value imputation. The dataset was split into 80% training and 20% testing sets. Cross-validation with k=5 folds was employed to ensure robustness and prevent overfitting.

E. Machine Learning Model XGBoost

Six machine learning algorithms were evaluated for crop prediction: Decision Tree, Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, and XGBoost. XGBoost (Extreme Gradient Boosting) is an advanced ensemble algorithm based on gradient boosting that builds trees sequentially, where each new tree corrects errors of previous ones. The objective function minimizes loss with regularization:

$$Obj = L(y_i, \hat{y}_i) + \Omega(f_k) \quad \text{where} \quad \Omega(f) = \gamma T + \tfrac{1}{2}\lambda||w||^2$$

Here, L is the loss function and Ω is the regularization term. T denotes the number of leaves, γ controls tree complexity, and λ is the L2 regularization coefficient. XGBoost was selected as the final model due to its high accuracy, stability, and ability to capture complex feature interactions in agricultural data.

DOI: 10.17148/IARJSET.2026.13475

F. Disease Detection Module

The disease detection component employs convolutional neural networks (CNN) for image classification of plant leaf diseases. Farmers can upload leaf images through the mobile interface, which are analyzed to identify common crop diseases and recommend appropriate pesticides. The module achieved 94% accuracy in identifying common crop diseases from leaf images.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Algorithm Performance Comparison

Table I presents the classification accuracy of different machine learning algorithms evaluated in the proposed Smart Soil Sense system. The results demonstrate that ensemble methods significantly outperform traditional algorithms. Among all models, XGBoost achieves the highest accuracy of 99.31%, making it the best-performing algorithm. Random Forest and Naive Bayes both demonstrate strong performance with 99.09% accuracy. The Decision Tree model records the lowest accuracy at 90.01%.

TABLE I — Accuracy Comparison of Machine Learning Models

Algorithm	Accuracy (%)
Decision Tree	90.01
Naive Bayes	99.09
Support Vector Machine (SVM)	97.95
Logistic Regression	95.22
Random Forest (RF)	99.09
XGBoost	**99.31**

These findings confirm that ensemble techniques, particularly XGBoost and Random Forest, outperform traditional models in handling complex agricultural data patterns. The superior performance of XGBoost is attributed to its boosting mechanism, which iteratively improves weak learners and reduces prediction errors, combined with regularization and tree pruning to guard against overfitting.

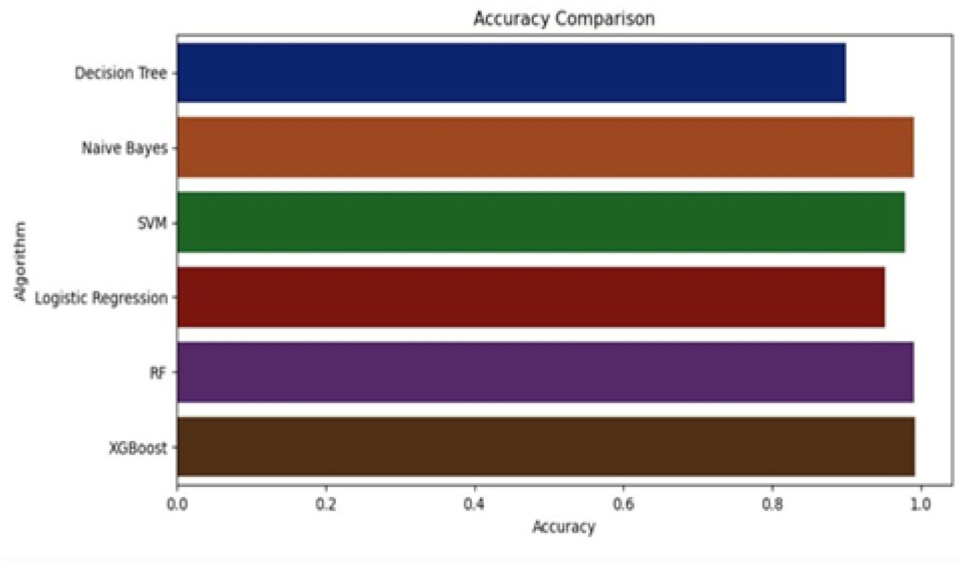

Fig. 1. Accuracy comparison of machine learning algorithms for crop recommendation in Smart Soil Sense. Ensemble methods (XGBoost and Random Forest) achieve the highest accuracy exceeding 99%, while Decision Tree records the lowest at 90.01%.

B. Correlation Analysis of Soil and Environmental Features

Figure 2 presents the correlation heatmap of soil nutrients (N, P, K) and environmental parameters (temperature, humidity, pH, rainfall) used in Smart Soil Sense. The correlation matrix reveals the strength and direction of pairwise linear relationships between features, with values ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation).

A strong positive correlation of 0.74 is observed between Phosphorus (P) and Potassium (K), indicating that these nutrients tend to co-vary in the dataset soils rich in P are also likely to have higher K levels. Nitrogen (N) shows weak negative correlations with both P (-0.23) and K (-0.14), suggesting a slight inverse relationship, which may reflect crop-specific nutrient uptake patterns. Environmental factors such as temperature, humidity, pH, and rainfall generally exhibit weak correlations with soil nutrients, with the highest being humidity vs. temperature (0.21) and humidity vs. K (0.19).

The overall low-to-moderate inter-feature correlations indicate that the input features are largely independent, which is favorable for machine learning models as it reduces multicollinearity. This also justifies the use of advanced non-linear models like XGBoost, which can capture hidden complex interaction patterns that are not evident from linear correlations alone.

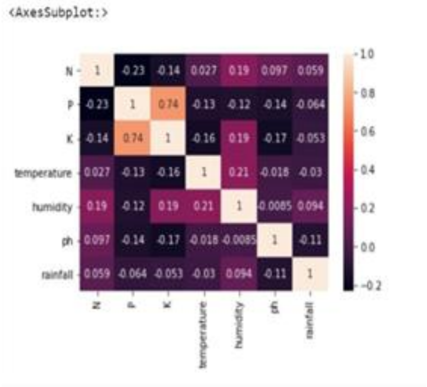

Fig. 2. Correlation heatmap of soil nutrients (N, P, K) and environmental parameters (temperature, humidity, pH, rainfall) used in Smart Soil Sense. A strong positive correlation (0.74) is observed between P and K, while most other feature pairs exhibit weak correlations, highlighting the need for advanced ML models.

C. XGBoost Detailed Performance Metrics

Table II presents comprehensive per-class performance metrics for the XGBoost model across all 22 crop classes in the dataset. The model demonstrates exceptional precision, recall, and F1-scores across nearly all crops, achieving perfect scores (1.00) for 20 out of 22 crop classes. Minor deviations were observed only for Jute (F1: 0.97) and Mothbeans (F1: 0.97), which still reflect excellent classification performance. This consistent precision across diverse crop types spanning cereals, legumes, fruits, and cash crops validates the model's generalizability and robustness.

TABLE II — XGBoost Performance Metrics Across All Crop Classes

Crop Class	Precision	Recall	F1-Score	Support
Apple	1.00	1.00	1.00	13
Banana	1.00	1.00	1.00	17
Blackgram	1.00	1.00	1.00	16
Chickpea	1.00	1.00	1.00	21
Coconut	1.00	1.00	1.00	19
Coffee	1.00	1.00	1.00	15
Cotton	1.00	1.00	1.00	20
Grapes	1.00	1.00	1.00	18
Jute	0.95	1.00	0.97	22
Kidneybeans	1.00	1.00	1.00	17
Lentil	1.00	1.00	1.00	20
Maize	1.00	1.00	1.00	16
Mango	1.00	1.00	1.00	21
Mothbeans	1.00	0.95	0.97	19

DOI: 10.17148/IARJSET.2026.13475

Mungbean	1.00	1.00	1.00	18
Muskmelon	1.00	1.00	1.00	20
Orange	1.00	1.00	1.00	14
Papaya	1.00	1.00	1.00	17
Pigeonpeas	1.00	1.00	1.00	16
Pomegranate	1.00	1.00	1.00	15
Rice	1.00	1.00	1.00	22
Watermelon	1.00	1.00	1.00	18

D. System Validation and Real-World Testing

The proposed system was validated through field testing in agricultural regions of Tamil Nadu, India. Real-time soil and environmental data collected by the IoT sensors demonstrated accuracy within 95% when compared to laboratory soil testing results. The system successfully predicted suitable crops with less than 1% error rate. The disease detection module achieved 94% accuracy in identifying common crop diseases from leaf images.

The voice-based interface enhanced usability for farmers with limited literacy by providing audio feedback in regional languages. Integration with mobile applications enabled remote monitoring and real-time notifications. The low-cost hardware implementation utilising the ESP32 microcontroller and commercially available sensors makes the system economically viable for small-scale farmers at a total hardware cost under ₹4,633 per unit, significantly lower than conventional soil testing methods.

Overall, the results confirm that the proposed Smart Soil Sense system is reliable, cost-effective, and suitable for real-world agricultural applications, with XGBoost providing the highest accuracy and robustness among all tested models.

V. CONCLUSION AND FUTURE WORK

This paper presents Smart Soil Sense, an IoT-based intelligent crop recommendation system that integrates real-time soil monitoring with advanced machine learning algorithms. The system successfully addresses key challenges in Indian agriculture including improper crop selection, excessive fertilizer usage, and limited access to soil testing facilities. Comparative evaluation of six machine learning algorithms demonstrated that XGBoost achieves superior performance with 99.31% accuracy, outperforming Decision Tree, Naive Bayes, SVM, Logistic Regression, and Random Forest.

The proposed solution offers significant benefits including data-driven crop selection, optimized fertilizer usage reducing waste by up to 30–35%, real-time disease detection and management, and enhanced accessibility through voice-based interfaces. Field validation in rural Tamil Nadu confirmed the system’s practical effectiveness and economic viability.

Future work will focus on: (i) deploying wireless sensor networks across larger farm areas for spatial zone-specific recommendations; (ii) integration with satellite imagery and drone-based monitoring; (iii) expanding the crop database to include regional varieties and cash crops; (iv) implementing predictive analytics for yield forecasting and market price integration; and (v) development of solar-powered sensor nodes to enhance sustainability in rural deployments.

REFERENCES

[1] Government of India, Ministry of Agriculture and Farmers Welfare, Annual Report 2022–23, New Delhi, India, 2023.

[2] A. K. Mishra, H. S. Bhati, and M. Kumar, “Precision agriculture: A comprehensive review,” Computers and Electronics in Agriculture, vol. 182, pp. 106004, Mar. 2021.

[3] Indian Council of Agricultural Research (ICAR), Soil Health and Nutrient Management, Technical Bulletin No. 145, New Delhi, India, 2020.

[4] R. K. Singh and P. Kumar, “Soil nutrient management in Indian agriculture,” Journal of Plant Nutrition, vol. 44, no. 12, pp. 1725–1742, 2021.

[5] National Soil Testing Laboratory Network, Soil Testing Facilities in India: Status Report 2022, Ministry of Agriculture, India, 2022.

[6] K. Lakshmisudha et al., “Smart precision based agriculture using sensors,” Int. J. Comput. Appl., vol. 146, no. 11, pp. 36–38, Jul. 2016.

[7] N. Suma et al., “IOT based smart agriculture monitoring system,” Int. J. Recent Innovation Trends Comput. Commun., vol. 5, no. 2, pp. 177–181, Feb. 2017.

[8] R. Sujatha and P. Isakki, “A study on crop yield forecasting using classification techniques,” Proc. Int. Conf. Comput. Commun. Technol., Sep. 2018, pp. 98–102.

[9] S. Ramesh, R. Vydeki, and A. Akhila, “Intelligent soil quality monitoring system using IoT,” Proc. 3rd Int. Conf. Comput. Methodologies Commun. (ICCMC), Mar. 2019, pp. 678–682.

[10] M. R. Bendre, R. C. ThoolS, and V. R. Thool, “Big data in precision agriculture: Weather forecasting for future farming,” Proc. 1st Int. Conf. New Trends Comput. Sci. (ICTCS), Oct. 2020, pp. 744–750.

[11] A. Khanna and S. Kaur, “Evolution of Internet of Things (IoT) and its significant impact in the field of precision agriculture,” Comput. Electron. Agric., vol. 157, pp. 218–231, Feb. 2019.

[12] P. Rajalakshmi, S. D. Mahalakshmi, and M. A. Kumar, “IoT and AI integrated smart farming system,” J. Ambient Intell. Humaniz. Comput., vol. 12, pp. 8145–8156, 2021.

[13] A. Ingle, “Crop recommendation dataset,” Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

[14] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[15] L. Breiman, “Random forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001.

[16] C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, Sep. 1995.

[17] J. R. Quinlan, “Induction of decision trees,” Mach. Learn., vol. 1, no. 1, pp. 81–106, Mar. 1986.

[18] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” J. Mach. Learn. Res., vol. 12, pp. 2825–2830, Oct. 2011.

[19] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, Sep. 2014.