World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Review Article)

AI-driven soil analysis and crop recommendation system

Abhin S Shetty, Deeksha Kamath, Joyvi Rodrigues, Sonal Dsouza * and Maryjo M George

Department of Artificial Intelligence and Machine Learning, Mangalore Institute of Technology and Engineering, Moodabidri, India.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(02), 2634–2643

Publication history: Received on 09 April 2025; revised on 16 May 2025; accepted on 19 May 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.2.0739

Abstract

This research introduces an innovative IoT-enabled Soil Analysis and Crop Recommendation System, aimed at transforming agricultural decision-making through the integration of advanced sensor technologies, cloud platforms, and machine learning techniques. Devices such as DHT11 sensor for temperature and humidity, alongside NPK nutrient and pH sensors, gather critical soil and environmental data. The ESP8266 microcontroller, in conjunction with the Blynk IoT platform, facilitates real-time data transmission and analysis, giving farmers useful information on climate and soil health. At the base of this system is a Random Forest Classifier that decides which crops to recommend based on NPK levels, pH, humidity, temperature, and rainfall for a particular set of environmental conditions. A multi-factor recommendation algorithm further refines these predictions by including soil nutrient profiles, pH measurements, temperature variation, and localized climate data for even more accurate crop recommendations. Experimental validation in several sites of agriculture was shown with 98% accuracy on crop selection. IoT and AI technologies will thus become the new future for farming practices. This system helps the farmer to use the resources much more efficiently and decrease the input cost with improved yield

Keywords: IoT Agriculture; Crop Recommendation; Random Forest; Sensor Integration; Precision Farming.

1. Introduction

Optimization of crop selection and cultivation strategies is a challenge for agriculture. Conventional techniques may result in less-than-ideal yields since they rely primarily on empirical knowledge and minimal soil testing.The proposed IoT-enabled soil analysis and crop recommendation system provides real-time, data-driven insights to overcome these issues. These systems use DHT11 as the temperature and humidity sensors. A pH sensor checks if the acidity or alkalinity of soil is appropriate. Finally, a sensor measures NPK level in terms of nitrate, phosphate, and potassium. All of these senses together give the entire meaning of the chemical and physical condition of soil, a requirement to make smart decisions.

An ESP8266 microcontroller is linked to the sensors that collects data and forwards it to Blynk's IoT Cloud-based Platform. This platform includes real-time viewing, monitoring, and computation of sensor information. Farmers and other stakeholder can see this data for real-time, which will enable him or her to make intervention timely and accurately in determining the health of the soils and the crop selection in his or her farm. The combination of automation and cloud technology makes this system connect traditional practices with modern precision agriculture. In addition to integrating the hardware, the system involves a user-friendly interface designed using Flask, fetching data from Blynk's IoT platform directly for visualization. Integration with sensor data and historical datasets is done using the Random Forest Classifier, which is a robust machine-learning algorithm. This classifier evaluates soil conditions and predicts the suitability of specific crops, making recommendations that optimize productivity. By leveraging historical insights and

* Corresponding author: Sonal Dsouza

real-time sensor inputs, the system transforms traditional agricultural practices into a data-driven approach, improving yields, resource utilization, and sustainability.

2. Literature Survey

Table 1 Literature Review on AI-Driven Soil Analysis and Crop Recommendation Systems — Methodologies, Applications, and Scope for Improvement

S. NO	Paper Title (Year of Publication)	Methodology	Scope for Improvement
1	IoT Based Smart Greenhouse, 2022	Sensor Integration and Data Transmission	Improve sensor accuracy, scalability, and real-time decision algorithms.
2	Smart Farming Using IoT, 2021	Data Monitoring and Automation	Develop offline-compatible IoT solutions and cost-effective setups for small-scale farms.
3	Crop Selection and IoT-Based Monitoring, 2020	Sensor Deployment and Analysis	Integrate advanced machine learning and enhance system scalability for diverse setups.
4	Data-Driven Smart Farming, 2021	Data Analytics and IoT	Use advanced machine learning for soil and crop analysis and expand datasets for adaptability.
5	Impact of IoT in Smart Agriculture, 2022	IoT-Based Smart Agriculture	Add AI-driven analytics, enhance interoperability, and address connectivity and security issues.
6	Smart Agriculture Sensors in IoT	Sensor-Driven Data Collection	Focus on data verification and robust peer-review processes for credibility.
7	Innovations in Soil Management, 2022	Interdisciplinary Soil Management	Integrate AI and IoT for resource efficiency and address socio-economic barriers.
8	AI and High-Throughput Phenotyping for Crop Improvement, 2024	AI-Driven Phenotyping	Address data standardization and innovate low-cost phenotyping platforms.
9	Crop Yield Prediction Review, 2024	Systematic Review of AI Applications	Integrate diverse data sources and improve deep learning model interpretability.
10	Soil Erosion Prevention Using Fog-Based Smart Agriculture, 2024	Fog-Based IoT Smart Agriculture	Enhance real-time weather integration and extend algorithm applications to diverse crops

Motwani et al. [11] discussed the soil analysis and crop recommendation through ML by highlighting the character of the soil that enhances yield in precision agriculture. Arooj et al. [12] used the concept of evaluating predictive data mining algorithms to the classification of soil in optimization crop recommendation systems and to derive knowledge on algorithm performance. Rajak et al. [13] have developed an ML-based crop recommendation system considering weather and soil-related factors, which produced encouraging results. In an extension of their previous research, Rajak et al. [14] incorporated supplementary environmental data to enhance the effectiveness of the system, highlighting the significance of data integration in agriculture.

Barshe et al. [15–17] proposed an AI-based framework for crop recommendation that offered impressive improvements over traditional methods. Their research included advanced techniques based on AI to analyze soil and crops, ensuring scalability as well as strong accuracy for various agricultural applications. Pudumalar et al. [18] proposed a precision agriculture system based on the integration of soil and environmental data, while Kulkarni et al. [19] have used ensemble techniques that increase the accuracy of the model's recommendation and reduce its variance and bias.

Gosai et al. [20] created a model for machine learning takes soil and climatic information, which was justified through empirical datasets, and thus enabled data-driven decision-making in agriculture. Reddy et al. [21] developed a crop recommendation system tailored to the Ramtek region, where localized data play a more significant role in achieving greater accuracy. Chauhan and Chaudhary [22] conducted a comparative study of machine learning methods using Support Vector Machines as a basis(SVM) and Random Forest in predicting agricultural yields. Doshi et al. [23] proposed AgroConsultant as an advanced system that integrates information on soil, meteorological, and geographic information for its applicability in practical farming contexts.

Sharma et al. in [24] developed the AI-Farm: using data-driven insights to enhance and surpass traditional crop recommendation systems and techniques. Ray et al. in [25], proposed a crop recommendation system that is useful to provide accurate recommendations for their real-time applications. More or less, Rawat et al. in [26], introduced several ML models and analyzed the accuracy and computational expense trade-off for crop recommenders. Balakrishnan et al. combined algorithms in [27], to enhance prediction accuracy and align research innovation with farm-ready applications.

Choudhury et al. [28] focused on the optimization of machine learning models to maximize the predictability of prediction systems and demonstrated how the optimization could significantly improve system performance.
Dhabal et al. [29] proposed a scalable, cloud-centric crop recommendation framework that focuses on accessibility and efficiency for agricultural producers. Finally, Parameswari et al. [30] explored different machine-learning approaches tailored for specific datasets and helped in the efficient deployment of crop recommendation systems. This body of research goes on to underscore the potential of ML and AI in altering agricultural practices through better decision-making, high yield, and sustainability of farming solutions.

3. System architecture

The System Architecture diagram below shows the process of collecting and analyzing agricultural data using IoT technology. Sensors are placed in the soil to collect data on temperature, pH, and NPK (nitrogen, phosphorus, potassium). The data is sent to an IoT gateway, which transfers it to the cloud. The results are therefore displayed on a UI to help farmers make informed decisions about their crops.

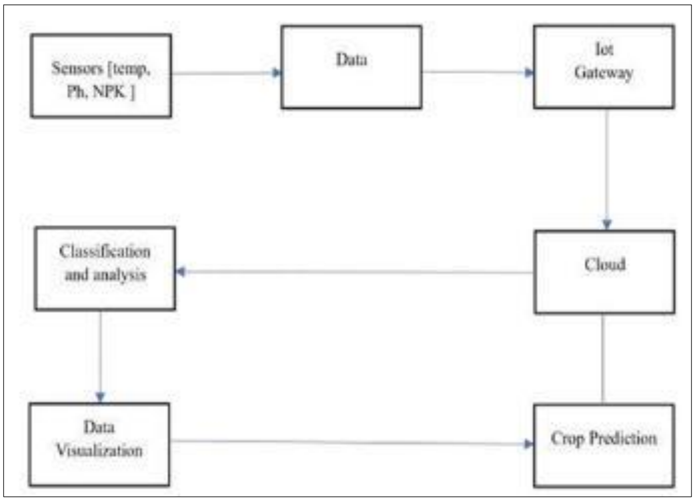

Figure 1 System Architecture diagram

3.1. Hardware Components

The proposed system integrates multiple sensors to capture comprehensive soil and environmental data:

3.1.1. DHT11 Temperature and Humidity Sensor

The DHT11 sensor is used to measure temperature and humidity, providing crucial data for understanding climate conditions. With an accuracy of ±5% for humidity and ±2°C for temperature, it offers reliable insights.

3.1.2. NPK Sensor

The NPK sensor is designed to detect the levels of essential soil nutrients—Nitrogen (N), Phosphorus (P), and Potassium (K). This data is critical for assessing soil fertility, enabling precise fertilization strategies that maximise crop output and reduce unnecessary chemical usage.

3.1.3. pH Sensor

The pH sensor measures the degree of acidity or alkalinity of the soil, providing key insights into the chemical properties. This information helps determine the suitability of soil for various crops, ensuring optimal growth conditions.

3.2. Microcontroller Integration

ESP8266 NodeMCU: The ESP8266 NodeMCU serves as the central data processing and transmission unit of the system. It offers built-in Wi-Fi connectivity, making it ideal for IoT applications. The microcontroller is coded using the Arduino IDE which ensures easy development and customization. Its low-power consumption design enhances the system's energy efficiency while maintaining seamless data collection and transmission to the cloud platform.

4. Experimental setup and implementation

4.1. Dataset Description

The dataset consists of 2200 entries and includes 7 features related to soil and environmental conditions: Nitrogen, Phosphorus, Potassium, Temperature, pH value, Humidity, and Rainfall. The target variable, labeled as Crop, represents the class to which each entry belongs, corresponding to the recommended crop type . The dataset encompasses 22 different crops with 100 data points for each of the crops.

4.2. Data Collection and Processing

The system uses a strong, multi-stage data collection and processing pipeline to ensure high accuracy and reliability in crop recommendations. The key stages include:

- ***Sensor Calibration:*** Sensors are used for data acquisition, such as those measuring soil nutrients, pH, temperature, and humidity, are calibrated periodically to ensure precision. This step helps in minimizing measurement errors and ensures consistency in data quality.
- ***Data Normalization:*** Data preprocessing is done in order to normalize units and scales. Nutrient levels, are scaled uniformly using MinMaxScaler.
- ***Real-time Cloud Transmission:*** Using IoT protocols, sensor data is transmitted in real-time to a cloud platform, therefore, easier integration into the prediction system and remote monitoring capabilities on platforms like Blynk.

4.3. Crop Recommendation Algorithm

START

4.3.1. Step 1: Import Libraries

IMPORT numpy, pandas, sklearn (MinMaxScaler, train_test_split, models, metrics)

4.3.2. Step 2: Load and Explore Dataset

LOAD dataset

HANDLE null values and duplicates

PRINT data summary and target distribution

4.3.3. Step 3: Preprocess Data

MAP crop names to numerical values

X ← feature columns, y ← target column

4.3.4. Step 4: Split and Normalize Data

SPLIT X and y into X_train, X_test, y_train, y_test (80:20)

SCALE X_train and X_test using MinMaxScaler

4.3.5. Step 5: Define and Train Models

MODELS ← {"Logistic Regression", "SVM", " Decision Tree" , " Random Forest" }

BEST_MODEL, BEST_ACCURACY ← NULL, 0

FOR EACH model IN MODELS:

TRAIN model on X_train, y_train

PREDICT y_test

ACCURACY ← calculate accuracy

IF ACCURACY > BEST_ACCURACY:

BEST_MODEL, BEST_ACCURACY ← model, ACCURACY

END FOR

4.3.6. Step 6: Output Results

PRINT BEST_MODEL and BEST_ACCURACY

OPTIONALLY save BEST_MODEL

STOP

4.4. Arduino IDE

The Arduino IDE code sets up an ESP32 microcontroller to read sensor data and send it to the Blynk platform for monitoring purpose.

4.4.1. Initialization and Setup

The program sets up the Blynk IoT platform and Modbus. It has credentials in its contents: Blynk token/auth, Wi-Fi name, password/pass for Internet connectivity of ESP32. TX_PIN and RX_PIN are allocated for UART communication, as well as RE and DE pins for the RS485 module. In setup(), UART is initialized at 9600 baud and the RS485 module is set to receive mode. ESP32 connects to Blynk, and a timer periodically calls the sendSensor function.

4.4.2. Modbus Communication and Data Handling

Modbus communication is done using the sendModbusQuery and readModbusResponse functions. The sendModbusQuery sends a predefined Modbus query for NPK (Nitrogen, Phosphorus, Potassium) sensor data, switching the RS485 module to transmit mode. The readModbusResponse reads the Modbus device's response, extracts NPK values, and sends them to Blynk virtual pins (V0, V1 , V2 , V3 , V4 , V5 , V6) for remote monitoring. By integrating Modbus communication with the Blynk IoT platform, the system enables remote and real-time monitoring of data.

4.5. Equations

4.5.1. Random Forest

Random Forest is an ensemble learning method. It combines multiple decision trees to improve prediction accuracy.

Mathematically, for a classification task, the final prediction y of the Random Forest model is:

$$y=\text{mode}(T1(x),T2(x),\ldots,Tk(x))$$

4.5.2. Gini Impurity

Gini impurity is used in decision trees and random forests to measure the "impurity" of a node.. The formula for Gini impurity at a node is:

$$\text{Gini}(D)=1-\sum_{i=1}^{C} pi^2$$

Where:

- D is the dataset at the node.
- C is the number of classes.
- Pi is the percentage of samples of class i in the dataset D.

5. Result

The study tested some of the popular machine learning models, which included Random Forest, Decision Tree, Support Vector Machine (SVM), and Logistic Regression. In the models, performance on the ability to make results predictions on the data was checked by training and validating with three different types of datasets for each of them to check the authenticity of the outcomes.

The results showed that, for every dataset, the Random Forest model performed better than the other models.

Table 2 Performance Comparison of Machine Learning Models for AI-Driven Soil Analysis and Crop Recommendation Based on Accuracy Across Train, Test, and Validation Datasets

Method	**Dataset**	**Accuracy**
Logistic Regression	Train	0.94
	Test	0.91
	Valid	0.91
Support Vector Machine	Train	0.97
	Test	0.90
	Valid	0.95
Decision Tree	Train	0.89
	Test	0.81
	Valid	0.86
Random Forest	Train	0.98
	Test	0.98
	Valid	0.98

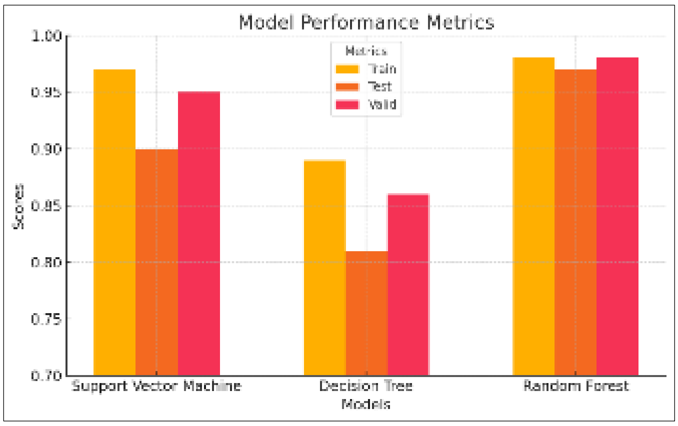

Figure 2 Bar Chart of Model Performance

This System was tested using real-time data that was collected from sensors deployed in agricultural fields. These sensors ensured precise and location-specific input data by measuring important environmental and soil characteristics, such as temperature, humidity, pH, rainfall, nitrogen, phosphorus, and potassium (NPK levels). As seen in the accompanying image, the gathered data was subsequently input into the system through an easy-to-use user interface. The algorithm in the system suggested the best crop after analysing these inputs.

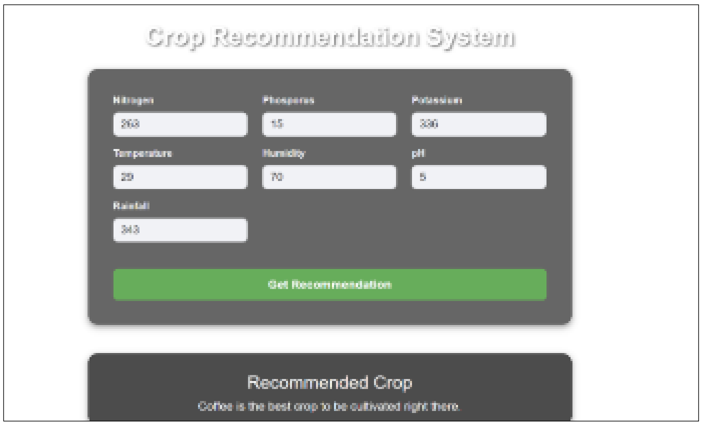

Figure 3 Crop Prediction Result

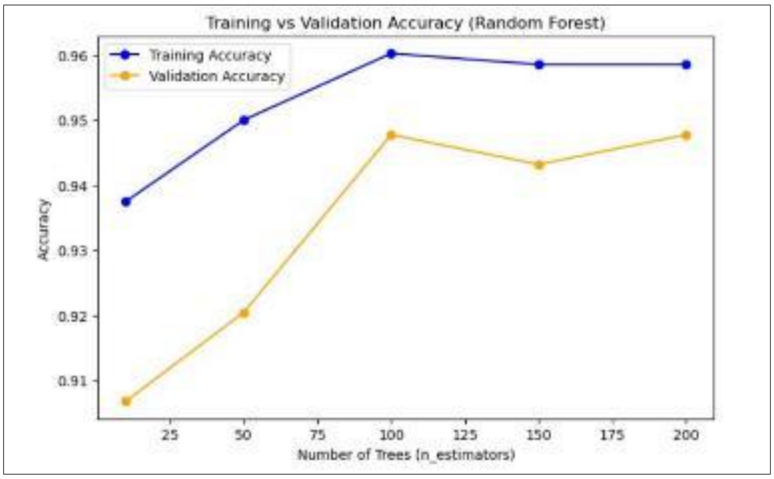

Figure 4 Accuracy Graph

The graph below shows F1 scores for crop classes using Logistic Regression, Decision Tree, Random Forest, and SVM on a crop recommendation dataset. The x-axis represents crop classes, and the y-axis shows F1 scores (0 to 1), indicating model performance. Higher scores near 1 reflect strong performance, while lower scores highlight areas needing improvement, guiding efforts to enhance accuracy for specific crops.

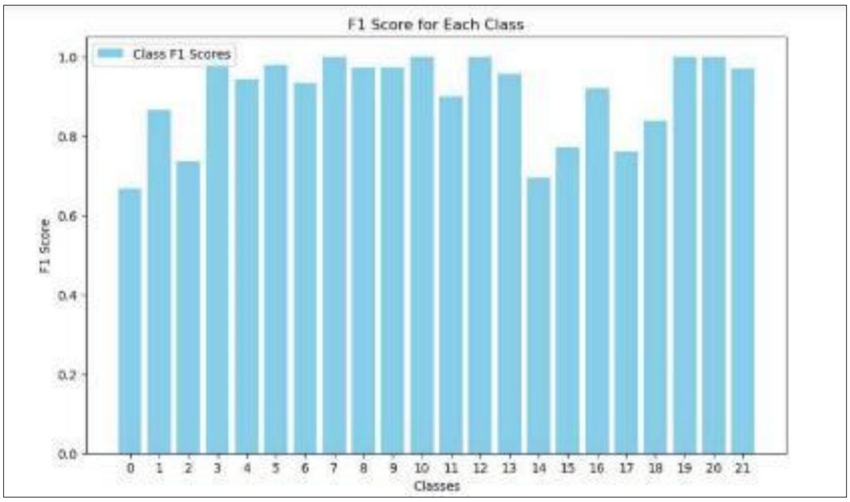

Figure 5 F1 Score

The confusion matrix evaluates classification models by comparing predicted and actual labels. Diagonal elements show correct predictions, while off-diagonal highlight errors like false positives and false negatives. It helps assess accuracy, identify weaknesses, and guide model improvements.

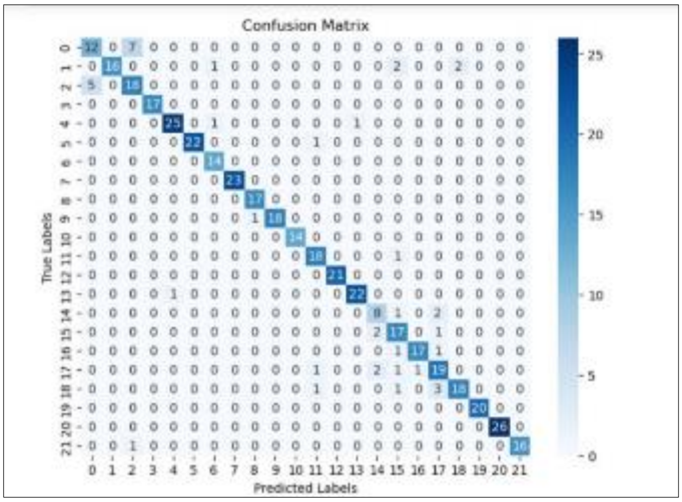

Figure 6 Confusion Matrix

6. Conclusion

The soil analysis and crop recommendation system is capable of demonstrating significant potential to help farmers take decisions based on data regarding sustainable agriculture. The system determines soil properties and gives advice on crop selection based on the acquisition of data from sensor technologies and further analyzes it using a machine learning model. The proper data flow and user engagement increase have been facilitated through the incorporation of hardware parts in software tools like Blynk IoT and Flask. This project embodies the possibility of integrating technology with agriculture, to increase productivity and resourcefulness.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] O. Sakpal, S. Shetye, S. Patil, S. Surve, and A. Jadiya, "IoT Based Smart Greenhouse," 2021.

[2] C. H. Nishanthi, D. Naveen, C. Sai Ram, K. Divya, and R. A. Kumar, "Smart Farming Using IoT," 2021.

[3] Y. Bhojwani, R. Singh, R. Reddy, and B. Perumal, "Crop Selection and IoT-Based Monitoring System for Precision Agriculture," 2020.

[4] S. K. Sangeetha, N. C., M. D., K. R., and K. K., "Data-Driven Smart Farming Using IoT and Data Analytics," 2021.

[5] O. V. Priya and R. Sudha, "Impact of Internet of Things (IoT) in Smart Agriculture," 2022.

[6] S. Ratnaparkhi, S. Khan, C. Arya, S. Khapre, P. Singh, M. Diwakar, and A. Shankar, "Smart Agriculture Sensors in IoT," 2020.

[7] R. K. Srivastava, S. Purohit, E. Alam, and M. K. Islam, "Advancements in Soil Management: Optimizing Crop Production Through Interdisciplinary Approaches," Dec. 2024.

[8] M. Sheikh, F. Iqra, H. Ambreen, K. A. Pravin, M. Ikra, and Y. S. Chung, "Integrating Artificial Intelligence and High-Throughput Phenotyping for Crop Improvement," Jun. 2024.

[9] M. A. Jabed and M. A. Murad, "Crop Yield Prediction in Agriculture: A Comprehensive Review of Machine Learning and Deep Learning Approaches, with Insights for Future Research and Sustainability," 2024.

[10] S. Mohanty, S. K. Pani, N. Tripathy, R. Rout, M. Acharya, and P. K. Raut, "Prevention of Soil Erosion, Prediction of Soil NPK and Moisture for Protecting Structural Deformities in Mining Area Using Fog-Assisted Smart Agriculture System," 2024.

[11] A. Motwani, P. Patil, V. Nagaria, S. Verma, and S. Ghane, "Soil analysis and crop recommendation using machine learning," in *2022 International Conference for Advancement in Technology (ICONAT)*, Jan. 2022, pp. 1–7.

[12] A. Arooj, M. Riaz, and M. N. Akram, "Evaluation of predictive data mining algorithms in soil data classification for optimized crop recommendation," in *2018 International Conference on Advancements in Computational Sciences (ICACS)*, Feb. 2018, pp. 1–6.

[13] R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A. Devare, "Crop recommendation system to maximize crop yield using machine learning technique," *International Research Journal of Engineering and Technology*, vol. 4, no. 12, pp. 950–953, 2017.

[14] R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A. Devare, "Crop recommendation system to maximize crop yield using machine learning technique," *International Research Journal of Engineering and Technology*, vol. 4, no. 12, pp. 950–953, 2017.

[15] S. B. Barshe, A. S. Kamble, P. Ramanathan, M. V. Deshmukh, N. R. Patil, and S. D. Jadhav, "Comprehensive Analysis of Artificial Intelligence based Crop Recommendation and Soil Analysis," in *2024 Second International Conference on Data Science and Information System (ICDSIS)*, May 2024, pp. 1–5.

[16] S. B. Barshe, A. S. Kamble, P. Ramanathan, M. V. Deshmukh, N. R. Patil, and S. D. Jadhav, "Comprehensive Analysis of Artificial Intelligence based Crop Recommendation and Soil Analysis," in *2024 Second International Conference on Data Science and Information System (ICDSIS)*, May 2024, pp. 1–5.

[17] S. B. Barshe, A. S. Kamble, P. Ramanathan, M. V. Deshmukh, N. R. Patil, and S. D. Jadhav, "Comprehensive Analysis of Artificial Intelligence based Crop Recommendation and Soil Analysis," in *2024 Second International Conference on Data Science and Information System (ICDSIS)*, May 2024, pp. 1–5.

[18] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in *2016 Eighth International Conference on Advanced Computing (ICoAC)*, Jan. 2017, pp. 32–36.

[19] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, "Improving crop productivity through a crop recommendation system using ensembling technique," in *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, Dec. 2018, pp. 114–119.

[20] D. Gosai, C. Raval, R. Nayak, H. Jayswal, and A. Patel, "Crop recommendation system using machine learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 7, no. 3, pp. 558–569, 2021.

[21] D. A. Reddy, B. Dadore, and A. Watekar, "Crop recommendation system to maximize crop yield in Ramtek region using machine learning," *International Journal of Scientific Research in Science and Technology*, vol. 6, no. 1, pp. 485–489, 2019.

[22] G. Chauhan and A. Chaudhary, "Crop recommendation system using machine learning algorithms," in *2021 10th International Conference on System Modeling and Advancement in Research Trends (SMART)*, Dec. 2021, pp. 109–112.

[23] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "AgroConsultant: intelligent crop recommendation system using machine learning algorithms," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Aug. 2018, pp. 1–6.

[24] A. Sharma, M. Bhargava, and A. V. Khanna, "AI-Farm: A crop recommendation system," in *2021 International Conference on Advances in Computing and Communications (ICACC)*, Oct. 2021, pp. 1–7.

[25] R. K. Ray, S. K. Das, and S. Chakravarty, "Smart crop recommender system—a machine learning approach," in *2022 12th International Conference on Cloud Computing, Data Science and Engineering (Confluence)*, Jan. 2022, pp. 494–499.

[26] P. Rawat, M. Bajaj, S. Vats, and V. Sharma, "An Analysis of Crop Recommendation Systems Employing Diverse Machine Learning Methodologies," in *2023 International Conference on Device Intelligence, Computing and Communication Technologies (DICCT)*, Mar. 2023, pp. 619–624.

[27] D. Balakrishnan, A. P. Kumar, K. S. K. Reddy, R. R. Kumar, K. Aadith, and S. Madhan, "Agricultural crop recommendation system," in *2023 3rd International Conference on Intelligent Technologies (CONIT)*, Jun. 2023, pp. 1–5.

[28] S. S. Choudhury, P. B. Pandharbale, S. N. Mohanty, and A. K. Jagadev, "An acquisition-based optimized crop recommendation system with machine learning algorithm," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 11, no. 1, 2024.

[29] G. Dhabal, J. Lachure, and R. Doriya, "Crop recommendation system with cloud computing," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, Sep. 2021, pp. 1404–1411.

[30] P. Parameswari, N. Rajathi, and K. J. Harshanaa, "Machine learning approaches for crop recommendation," in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, Oct. 2021, pp. 1–5.