



AUTOMATED CROP RECOMMENDATION SYSTEM USING IOT AND MACHINE LEARNING FOR SOIL HEALTH

Ankit Shivkumar Pandey¹, Dr. Sangeeta Vhatkar²

Graduate, Student, Department of Information Technology, Thakur College of Engineering & Technology, Mumbai, Maharashtra, India.¹

Associate Professor, Department of Information Technology, Thakur College of Engineering & Technology, Mumbai, Maharashtra, India.²

Abstract: Using data-driven technologies, modern agriculture promotes sustainability while increasing farming efficiency. Intelligent farming solutions, like the one suggested in this paper, are now possible thanks to developments in machine learning (ML) and the internet of things (IoT). Using an Internet of Things framework, it integrates AI-based crop recommendations with real-time soil monitoring. Together with a temperature and pH sensor, a multi-electrode NPK sensor measures the three main nutrients found in soil: nitrogen (N), phosphorus (P), and potassium (K). An ESP32 microcontroller receives data through an RS485-to-USB interface, processes it, and shows it on a 20x4 LCD I2C screen. Real-time weather data from API integrations is also added. Reliance on conventional lab testing is decreased by cloud-based machine learning models that evaluate environmental variables and soil health to recommend appropriate crops. Because of the system's remote accessibility, farmers—even in remote locations with little technical know-how—to make wise choices. Additional sensors (such as light and moisture sensors) and weather-adaptive predictive models are examples of future improvements that could improve the accuracy of the system and encourage sustainable farming methods.

I. INTRODUCTION

In the fast-evolving world of technology, timely information sharing is crucial to enable agriculturists to reach their potential and make informed choices. Information sharing entails the sharing of relevant and useful information between farmers through informal networks or formal structures. The receptiveness and hospitality of agriculturists to share information heavily influence the extent and quality of the knowledge exchange, which ultimately determines agricultural output and economic stability. In order to enable smooth sharing of information, we use web technologies like HTML, CSS and JAVASCRIPT to create a web-based application. The platform gathers and consolidates agricultural information from various sources, ranging from Kaggle and the government website to include market trends, weather patterns, and past crop prices. The dataset is subsequently used to forecast crop prices so that farmers have access to current market information. The information is stored safely in a MySQL server, and registered farmers are able to view it via the application. Also, to improve ease of access, the system provides for sending updated data in text messages automatically, thus getting rid of the problem of farmers having to go to cities or towns to get updates on markets. We use machine learning algorithms such as DECISIONTREE, LOGISTICREGRESSION, RANDOMFOREST. These algorithms examine past patterns of prices and prevailing market trends to predict crop prices over the next months. The resulting predictions are then subjected to a non-linear filtering process, prioritizing based on accuracy and pertinence. A ranking mechanism is subsequently employed to select the most accurate predictions, helping farmers make informed strategic decisions regarding crop choice and sale. By combining web technologies with machine learning, this system is designed to deliver a strong, data-based solution for agriculturists, enhancing efficiency and diminishing dependence on conventional, time-consuming means of market analysis. Future development could involve adding more predictive models and increasing the dataset to further enhance accuracy and usability, rendering it a useful tool for contemporary agricultural practices.

II. PROBLEM STATEMENT

Crop production largely depends on the quality of soil and nutrient levels. Traditional methods of farming may not pay particular attention to individual soil lacks and environmental factors, resulting in poor crop yields. Most farmers rely on wide-range fertilization and irrigation, which may cause waste of resources and soil erosion. Moreover, limited



monitoring of real-time data and crop-specific information restricts the implementation of sustainable farming practices. This research suggests an ML and IoT-based solution to overcome these constraints, which can be applied to any type of crop farming. The system will use soil classification through ML and a decision tree algorithm to evaluate crops datasets, including soil types, NPK levels, pH, temperature, humidity, and moisture content. The aim is to provide farmers with precise, data-based advice to improve crop yields while ensuring environmental sustainability.

Classification of Soils of India:

India has varied relief features, landforms, climatic realms and vegetation types. These have contributed in the development of various types of soils. On the basis of genesis, color, composition and location, the soils of India have been classified into:

- I. Alluvial soils
- II. Black soils
- III. Red and Yellow soils
- IV. Laterite soils
- V. Arid soils
- VI. Saline soils
- VII. Peaty soils
- VIII. Forest soils.

Seasonal Crops and Soil Requirements:

➤ Kharif Season: Kharif crops are cultivated during the monsoon season in India from June to September and harvested from September to October. These crops are rainfall-dependent and need warm, humid weather to grow. Adequate Soil Types for Kharif Crops:

- Alluvial Soil – Ideal for rice, maize, and sugarcane.
- Black Soil (Regur Soil) – Ideal for cotton, jowar, and soybean.
- Red Soil – Ideal for pulses, millets, and groundnut.
- Laterite Soil – Ideal for tea, coffee, and cashew.
- Sandy Soil – Ideal for crops like millets and groundnut in dry areas.

➤ Rabi Season: Rabi crops are cultivated during the winter season between October and March and harvested in March and April. These crops need cool weather for growth and warm temperature for ripening. They are based on irrigation, not monsoon rains.

Appropriate Soil Types for Rabi Crops:

- Alluvial Soil – Most suitable for wheat, barley, and mustard.
- Black Soil – Most suitable for wheat, chickpeas, and oilseeds.
- Loamy Soil – Suitable for pulses, vegetables, and medicinal plants.
- Clayey Soil – Favors wheat and barley as it has good water-holding capacity.

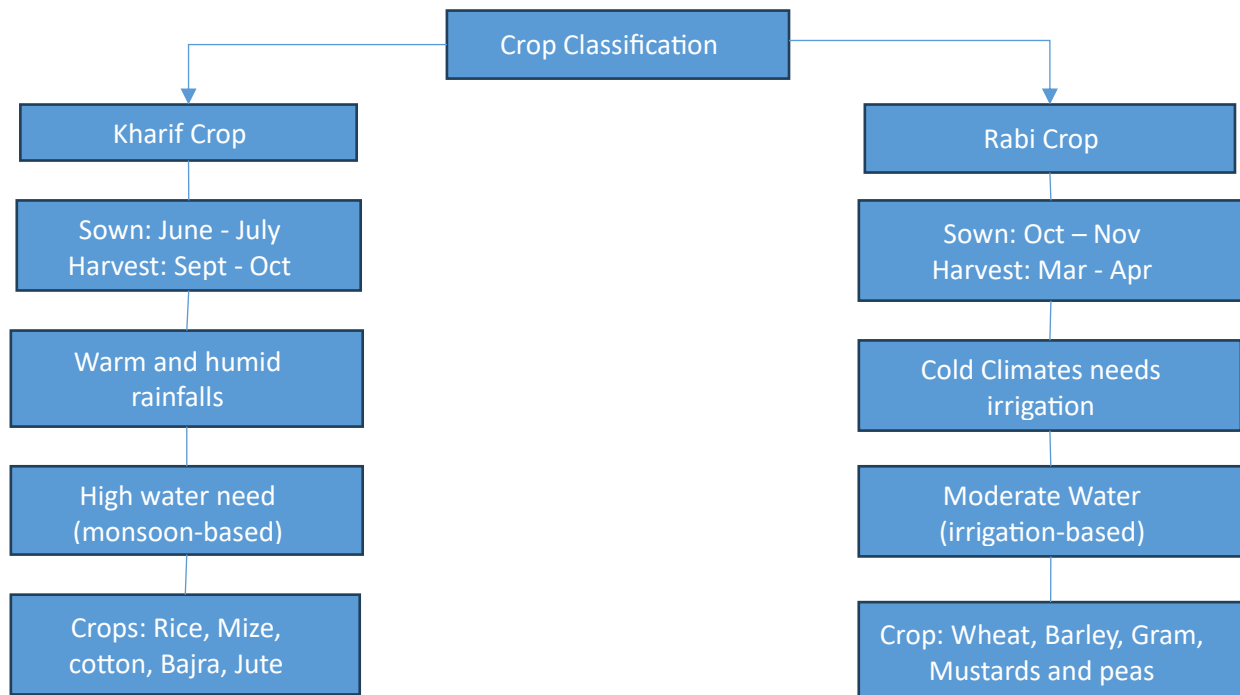


Figure 1: Crop cultivation

III. OBJECTIVE

- **IoT-Based Real-Time Soil Monitoring:** Design an IoT-powered system to constantly evaluate soil health by connecting high-end sensors to monitor crucial nutrients (NPK), moisture, and PH. The gathered data will be fed into a cloud-based server for real-time analysis, which will provide farmers timely and accurate insights. Manual soil testing will no longer be required with this automated method, facilitating precision farming.
- **Machine Learning for Crop Recommendation:** Adopt a machine learning model that scans past agricultural data, environmental factors, and real-time soil data to make individualized crop recommendations. The model will use large databases, such as past yields, climate patterns, and soil composition, to enhance crop selection and increase productivity.
- **Predictive Analytics for Informed Decision-Making:** Use predictive analytics to enable farmers to maximize resource use, minimize input expenses, and achieve maximum yield. By interpreting soil health trends and market trends, the system will facilitate data-driven decision-making, reducing risks related to ambiguous environmental conditions.
- **User-Friendly Visualization Platform:** Create an easy-to-use web and mobile application to present real-time soil health indicators, crop suggestions, and alerts. The platform will have interactive dashboards with graphical displays of soil conditions for easy access to key insights for farmers with limited technical capabilities.
- **Web Application for Data Management:** Create a web-based centralized system for user registration, data collection, and management. Farmers can enter agricultural data, view past records, and get automated updates. The system will offer data security and create a glitch-free interface for farmers to monitor and manage their farming insights in an effective manner.
- **Maximized Crop Selection for Yield Improvement:** Provide farmers with the best crop for every season by considering soil type, climate, and history of yield performance. The system will minimize crop selection uncertainty, maximize yield potential, and ensure sustainable agriculture practices.
- **Crop Yield Prediction with Machine Learning:** Apply machine learning models like Decision Tree, Logistic Regression, Random Forest, to forecast crop yields based on factors like rainfall, soil acidity, temperature, area under cultivation, and season. These will enable farmers to make more effective production and planning decisions for markets.



IV. AIM AND SCOPE OF THE PRESENT INVESTIGATION

Aim: The system can be extensively used in the agricultural industry to enhance the precision of crop yield forecasting. By being integrated into a web or mobile application, it provides a user-friendly interface through which farmers can easily access recommendations. The method can also be generalized to other applications needing classification and ranking methods, e.g., educational performance analysis or industrial purposes.

Scope: Precision Agriculture: The system allows farmers to make informed decisions based on analysing soil conditions, weather patterns, and crop suitability, thus maximizing yield.

IoT Integration: With the integration of IoT sensors, real-time data regarding soil moisture, temperature, and other parameters can be gathered and processed for more precise recommendations.

Scalability: The system can be scaled to accommodate various crops, varied soil types, and varied climatic conditions, hence suitable for application across different geographical areas.

User Accessibility: For ease of integration with web as well as mobile applications, ensuring the convenience of usage for farmers of different technical skill levels.

Sustainability and Cost Savings: Facilitates effective use of resources, preventing wasteful usage of water, fertilizers, and pesticides, resulting in cost savings and saving the environment.

Cross-Domain Applications: The classification and ranking techniques employed within the system are applicable to other domains, for example, student performance analysis, industrial decision making, and diagnosis in healthcare.

V. EXPERIMENTAL OR MATERIALS AND METHODS

RANDOM FOREST ALGORITHM: Random Forest is a popular supervised learning algorithm used for both classification and regression tasks. It employs an ensemble learning strategy, with several decision trees being merged together to enhance accuracy and stability. Rather than having a single decision tree, Random Forest pools predictions from many trees and takes the most frequent or averaged result, minimizing overfitting and improving reliability.

What a Random Forest Works:

- Randomly samples a subset of data points from the training data.
- Builds a separate decision tree for each of several subsets of data.
- Specifies how many decision trees (N) to construct.
- Takes multiple sets of input and produces multiple trees.
- For new data, each tree gives a prediction, and output is decided with majority voting (classification) or averaging (regression).
- Through the addition of more decision trees, the accuracy of the model is enhanced, making it particularly suitable for agricultural yield forecasting.

DECISION TREE: Decision Trees are extensively applied in supervised learning for classification and regression problems. They work by dividing data into branches according to feature values, eventually resulting in a final decision or classification.

- **Root Node:** Represents the entire dataset and initiates the decision-making process.
- **Decision Nodes:** Intermediate nodes that apply decision rules based on feature values.
- **Leaf Nodes:** Represent final outcomes or classifications.

Working Mechanism of Decision Tree:

1. The data is examined to determine the optimal attribute to split on.
2. The root node is established, and the dataset is separated into subsets in accordance with the selected attribute.
3. A decision node is created, and the process is repeated for every subset.
4. The recursive algorithm continues until it is not possible to split anymore, creating the final leaf nodes.
5. Decision Trees are extremely interpretable and effective and, therefore, an incredibly valuable instrument for agricultural forecasting models.

**IoT SYSTEM COMPONENTS:****1. NPK Sensor with three Needles**

- Measures soil nutrient levels (Nitrogen, Phosphorus, and Potassium) accurately.
- Ensures precise soil fertility analysis for better crop recommendations.
- Communicates data via RS485 protocol for high reliability.

2. RS485 USB Dongle

- USB to RS485 conversion for serial communication.
- Half-duplex communication, meaning it can send and receive data but not simultaneously.
- Supports multiple devices (up to 32 or more) on the same bus.
- Built-in transceivers like MAX485 or CH340 for reliable data transmission.
- Plug-and-play support with most operating systems (Windows, Linux, macOS).
- Baud rate support ranging from 300bps to 1Mbps.

3. RS485 TTL Auto-Direction Converter

- TTL to RS485 conversion: Converts standard UART signals (3.3V or 5V) to RS485.
- Auto-direction control: Automatically switches between sending and receiving modes.
- Supports half-duplex communication (one direction at a time).
- Long-distance communication: Up to 1200 meters over twisted-pair cables.
- Supports multiple devices: Can connect up to 32 nodes on a single RS485 bus.
- Voltage compatibility: Typically works with 3.3V and 5V logic levels.

4. ESP32 Microcontroller (38 Pin)

- Processor: Dual-core Xtensa LX6 (240MHz).
- Connectivity: Wi-Fi (802.11 b/g/n) & Bluetooth (Classic + BLE).
- GPIOs: 38 pins with multiple interfaces (I²C, SPI, UART, PWM, ADC, DAC).
- Analog & Digital I/O: 18 ADC (12-bit), 2 DAC (8-bit).
- Touch Sensors: Built-in capacitive touch support.
- Power: Supports deep sleep for low-power applications.
- Flash Memory: Typically, 4MB (varies by model).
- USB Interface: Micro-USB for power and programming.

5. 20×4 LCD I2C Module

- Display: 20 characters × 4 lines (Blue/Green backlight).
- Interface: I2C (Serial SDA, SCL) – requires only 2 pins.
- Chipset: PCF8574 I2C expander for easy communication.
- Operating Voltage: 5V (some support 3.3V).
- Adjustable Contrast: Via onboard potentiometer.

6. 12V DC Supply Connector

- Powers the entire system efficiently, ensuring stable operation in rural areas.
- Protects against voltage fluctuations to maintain sensor accuracy.

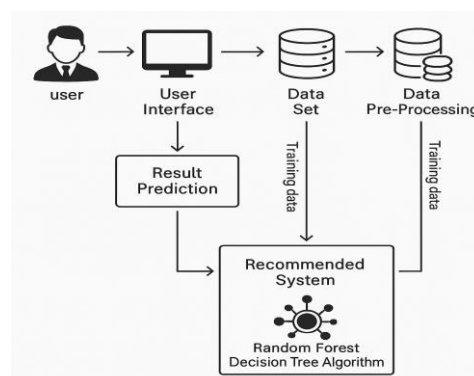
VI. SYSTEM ARCHITECTURE

Figure 2: System Architecture

**SOFTWARE REQUIREMENTS:**

System: Pretium Dual core
Hard disk: 40Mb
Monitor: 15" LED
Ram: 4 GB
Operating system: Windows 10/11
Coding language: python

SOFTWARE ENVIRONMENT:

Python: Python is an interpreted, high-level, object-oriented, interactive, and code readability-focused programming language that aims to be as simple and intuitive as possible. It is one of the best languages for beginners and professionals alike. Python does not use a lot of syntax like most other programming languages do, nor does it use an abundance of punctuation. Instead, it uses English keywords extensively, which makes it easier to read and understand. Second, Python has less syntactic regulation than other computer programming languages and, therefore, makes coding less cumbersome and faster.

MODULES:**1. Admin Login**

The Admin Login module acts as the main access control system, where administrators need to input a registered phone number and a password. When the credentials inputted are matched with the database records stored, the admin gets the access; otherwise, a failed login alert causes re-entry of the proper information. This security feature allows only authorized staff to access and handle the system.

2. Metadata:

The Metadata module standardizes and organizes key dataset data by having different crops assigned with their own unique numerical identifiers. Having data structured in this way makes data processing simpler and enhances machine learning algorithm efficiency. Every crop is given a unique number so as to avoid duplication, hence making things clear and consistent. Metadata has extensive information about over a hundred crops that are grown all over India, and it allows for easy data integration and analysis.

3. Data Pre-processing

The Data Pre-processing module cleans raw crop data by eradicating inconsistencies and incorporating metadata. It entails encoding categorical crop information into numerical values, rendering the data amenable to machine learning models. The process entails loading metadata, combining it with raw data, eliminating irrelevant information, and dividing the dataset into training and testing splits. This enhances prediction accuracy and improves model performance.

4. Crop Prediction Module

The Crop Prediction Module uses machine learning algorithms to analyze environmental and soil parameters, making accurate predictions for yields across different crops. This module enables farmers to make informed choices with the ability to predict crop yield based on current conditions. Providing visibility into likely yields, it facilitates improved crop choice, optimizing farming efficiency and resource utilization.

5. Crop Recommendation Module

The Crop Recommendation Module provides research-based suggestions on the most suitable and lucrative crops based on soil type, weather conditions, and historical yield trends. By providing suggestions on a regional scale, the module helps farmers make optimal crop choices for their fields, achieving maximum productivity and profitability. This system promotes wise farming practices, ensuring optimum and productive agricultural yields.

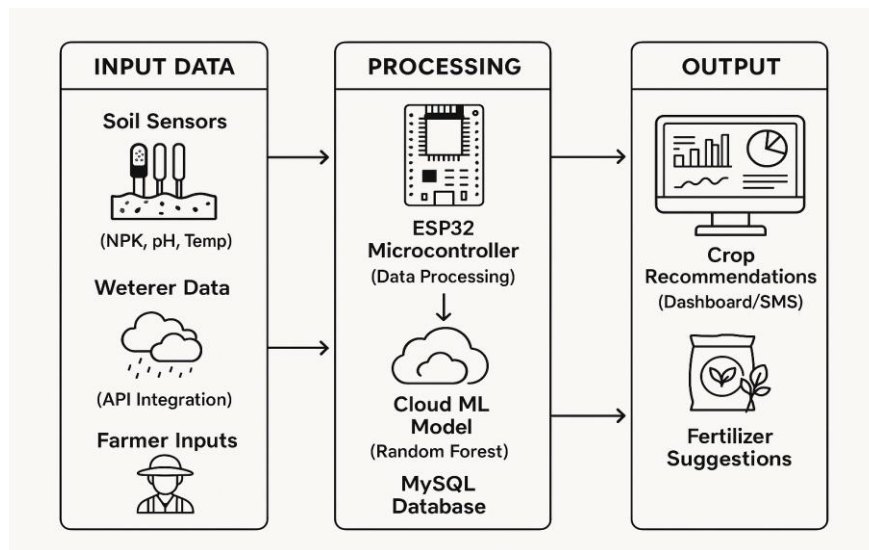


Figure 3: System Design

INPUT DESIGN AND OUTPUT DESIGN:

Input Design:

Involves transforming user-defined input descriptions into a computerized system. This process is essential for reducing data entry errors and ensuring that management receives accurate and reliable information from the system. To achieve this, user-friendly data entry screens are developed to efficiently handle large amounts of data. The main objective is to simplify data input while minimizing mistakes. These screens are designed to support various data manipulations and provide options for viewing records.

During data entry, the system validates the input to ensure accuracy. Users enter data through interactive screens, where relevant messages are displayed as needed to provide guidance and prevent confusion. The ultimate goal of input design is to create a structured and easy-to-navigate layout for seamless data entry.

Output Design:

A high-quality output effectively meets user requirements and presents information in a clear and structured manner. In any system, the results of processing are shared with users and other systems through outputs. **Output design** focuses on how information should be displayed for immediate use and how it should be formatted for printed reports. Since outputs serve as a primary source of information for users, a well-designed output enhances usability and aids in decision-making.

Key Aspects of Output Design:

1. **Systematic Development** – The process of designing outputs should be well-structured and carefully planned. It is essential to develop outputs that are easy to interpret and useful for users. Analysts must determine the specific outputs necessary to meet system requirements.
2. **Effective Information Presentation** – Choosing the most suitable method for displaying information ensures clarity and ease of understanding.
3. **Structured Output Formats** – Documents, reports, and other formats should be designed to present system-generated information in an organized manner.

Objectives of System Outputs:

A well-structured output should fulfil one or more of these functions:

- Provide insights into past activities, current conditions, or future trends.
- Highlight significant events, opportunities, problems, or alerts.
- Prompt necessary actions.
- Confirm successful completion of actions.

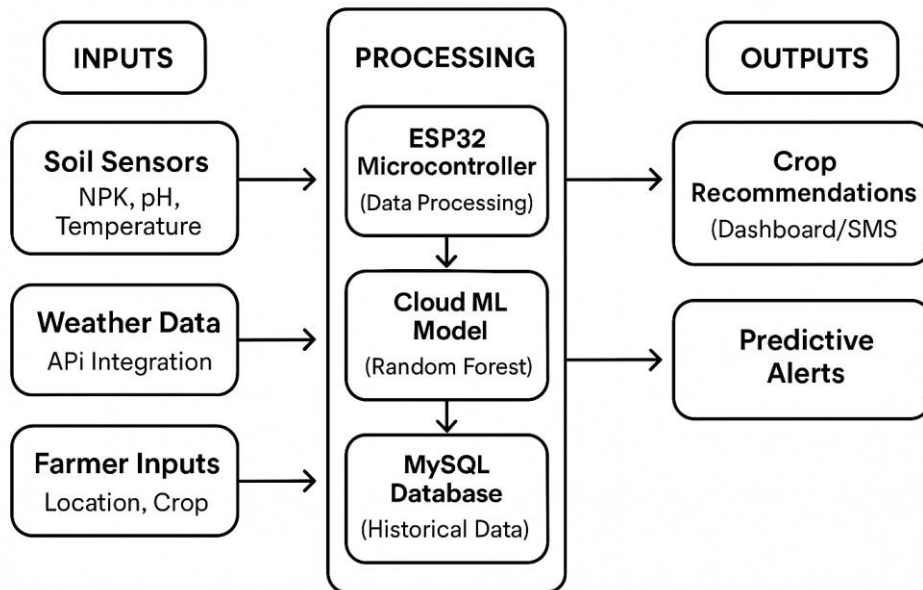


Figure 4:Input and Output Design

VII. RESULTS AND DISCUSSIONS

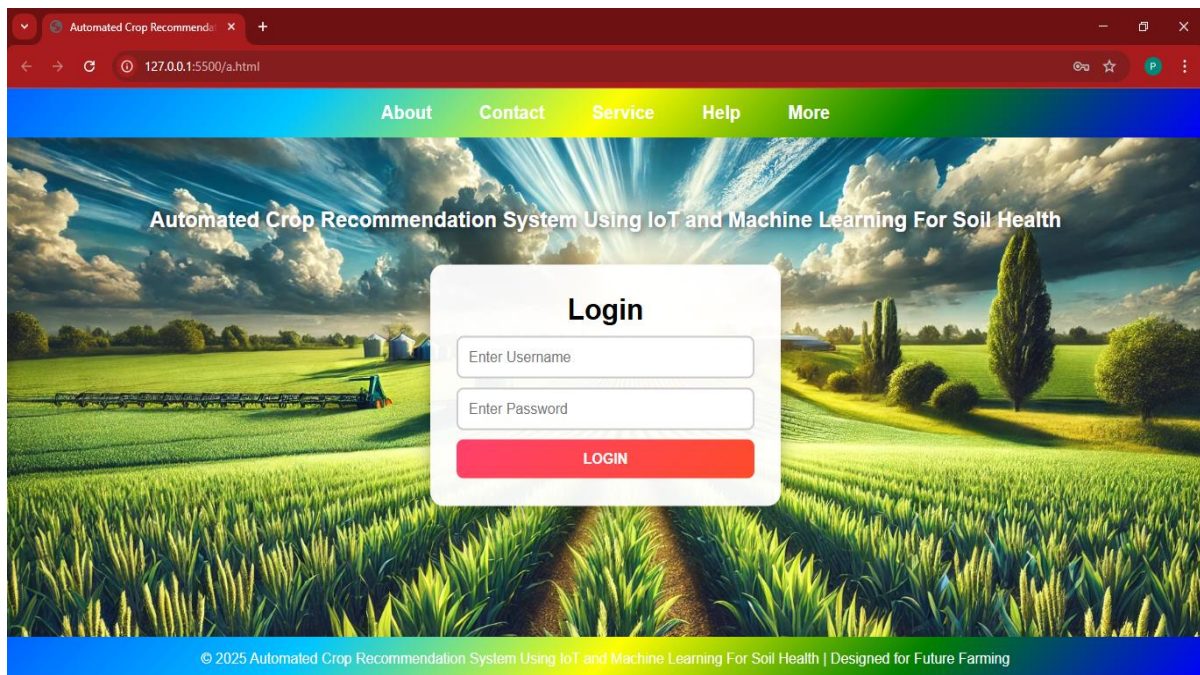
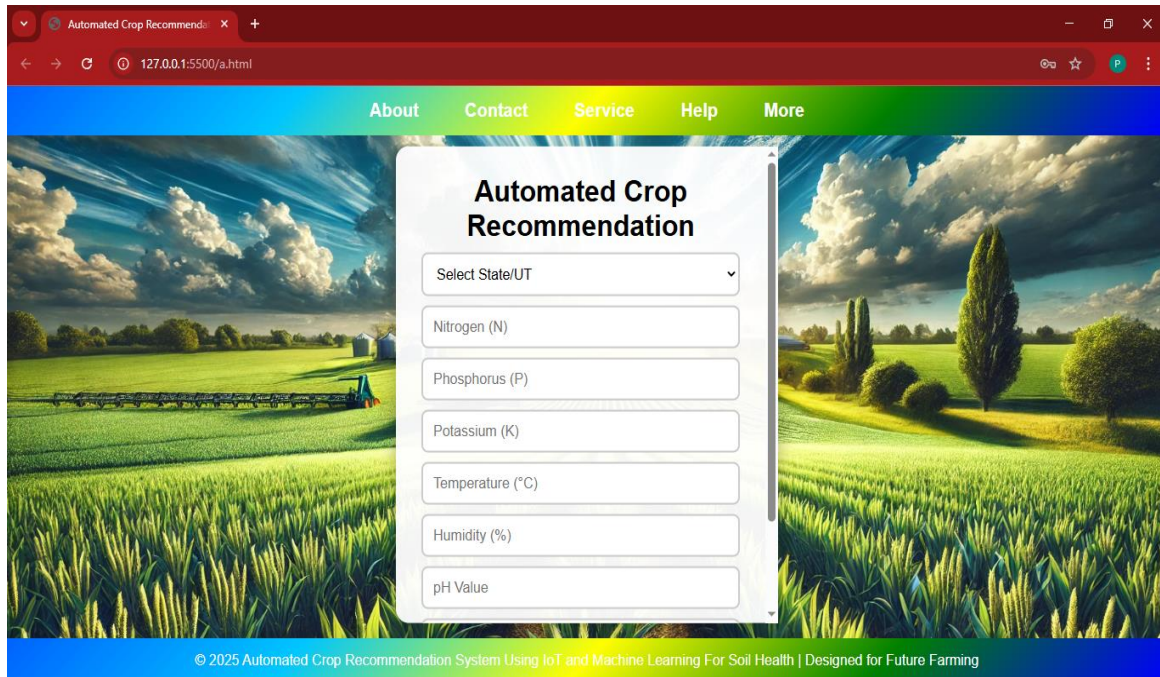


Figure 5:Login page



The screenshot shows a web browser window with the URL 127.0.0.1:5500/a.html. The page has a navigation bar with links: About, Contact, Service, Help, and More. The main content area features a large background image of a green field under a blue sky with clouds. Overlaid on this is a white form titled "Automated Crop Recommendation". The form contains the following fields: "Select State/UT" (a dropdown menu), "Nitrogen (N)" (a text input), "Phosphorus (P)" (a text input), "Potassium (K)" (a text input), "Temperature (°C)" (a text input), "Humidity (%)" (a text input), and "pH Value" (a text input). At the bottom of the page, there is a copyright notice: "© 2025 Automated Crop Recommendation System Using IoT and Machine Learning For Soil Health | Designed for Future Farming".

Figure 6:Home Page

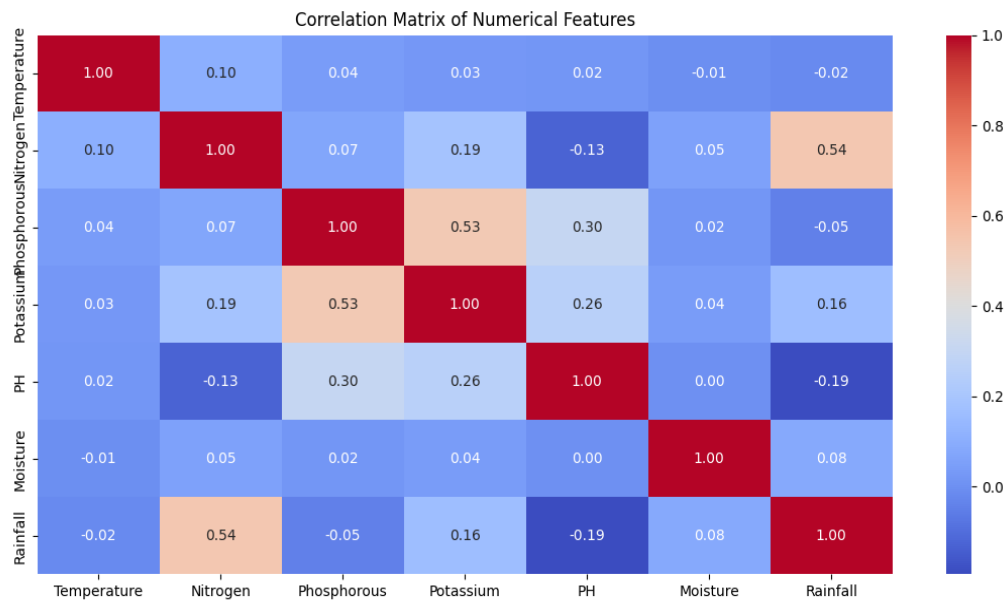


Figure 7:Correlation Matrix of Numerical Features

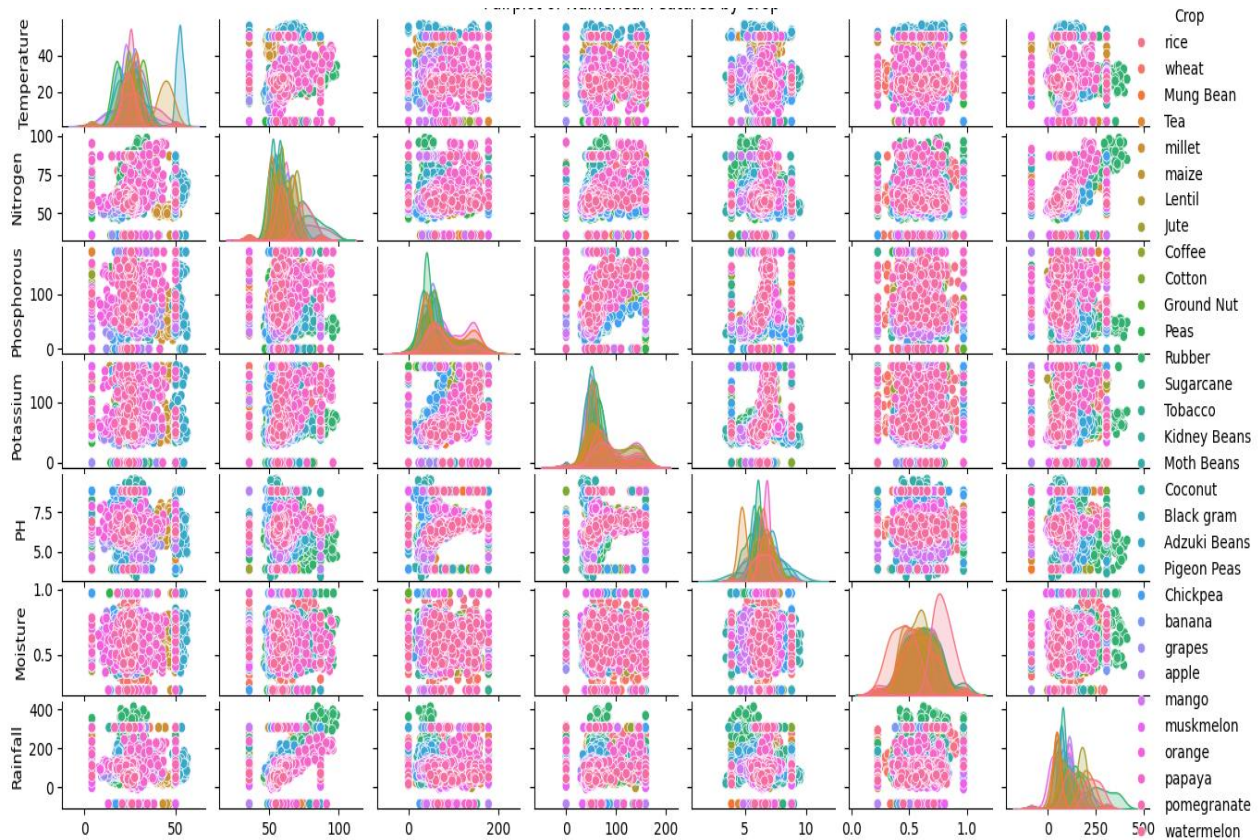


Figure 8: Crop and Soil Feature

DISCUSSIONS:

This work illustrates the success of combining machine learning with IoT for automated crop suggestion. With the use of real-time environmental and soil information, the system is able to supply precise and specific suggestions, enhancing agricultural decision-making as opposed to conventional techniques. The data-based approach meets the trends in smart farming, which maximizes productivity and sustainability.

Nevertheless, some constraints apply. Recommendations' accuracy is sensor data quality dependent, and hardware or connectivity faults could interfere with real-time functionality. Also, the dataset might not represent all soil and climatic variability, reducing model generalizability. These limitations can be overcome with better data collection and robust infrastructure, thereby increasing reliability.

Subsequent work must target further development of the dataset, including novel deep learning methods, and engineering an accessible user interface to further enhance accessibility. The added advancements will further enhance the uptake of smart farming technologies, driving sustainable and efficient agriculture.

VIII. CONCLUSION

This research proves the efficacy of combining machine learning and IoT-based soil analysis for enhanced crop suggestions. Through the use of real-time data from IoT sensors, the system gives precise and customized crop suggestions based on soil parameters like pH, moisture, temperature, and nutrients. Machine learning models also improve decision-making by examining historical and real-time data, resulting in increased productivity and resource utilization. The findings show that with this data-driven strategy, risks in crop choice are minimized, and sustainable agriculture is encouraged. Although the system works effectively, future enhancements can involve increasing datasets, adding further environmental variables, and more precise predictive models for wider use. Generally, the combination of IoT and AI in agriculture is an encouraging move toward precision agriculture, allowing farmers to make informed choices, maximize outputs, and enhance sustainability. With ongoing innovation, this study can play an important role in the future of intelligent agriculture.



REFERENCES

- [1]. Babu, Gurujukota Ramesh, et al. "Integrating IoT for Soil Monitoring and Hybrid Machine Learning in Predicting Tomato Crop Disease in a Typical South India Station." *Sensors*, vol. 24, no. 19, 24 Sept. 2024, pp. 6177–6177, <https://doi.org/10.3390/s24196177>.
- [2]. Farida Siddiqi Prity, et al. "Enhancing Agricultural Productivity: A Machine Learning Approach to Crop Recommendations." *Human-Centric Intelligent Systems*, vol. 4, no. 2, 2 Sept. 2024, <https://doi.org/10.1007/s44230-024-00081-3>.
- [3]. G, Murugesan, and Radha B. "Crop Rotation Based Crop Recommendation System with Soil Deficiency Analysis through Extreme Learning Machine." *International Journal of Engineering Trends and Technology*, vol. 70, no. 4, 25 Apr. 2022, pp. 122–134, <https://doi.org/10.14445/22315381/ijett-v70i4p210>. Accessed 4 May 2022.
- [4]. Gosai, Dhruvi, et al. "Crop Recommendation System Using Machine Learning." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 7, no. 3, 1 June 2021, pp. 558–569, <https://doi.org/10.32628/cseit2173129>.
- [5]. John, Ashwin, et al. "Soil Classification and Crop Recommendation System." *Zenodo (CERN European Organization for Nuclear Research)*, vol. 7, no. 6, 23 June 2022, <https://doi.org/10.5281/zenodo.6692499>. Accessed 11 Apr. 2025.
- [6]. Kulkarni, Mr, et al. "CROP RECOMMENDATION SYSTEM USING MACHINE LEARNING ALGORITHM." *II*, vol. 10, no. 2, 2022, pp. 2320–2882, www.ijert.org/papers/IJCRT2205616.pdf. Accessed 11 Apr. 2025.
- [7]. Manzire, Rushikesh, and Heena Sanghani. "Physico-Chemical Analysis of Soil from Northern Part of Pune District." *II*, vol. 8, no. 7, 2020, p. 5382, www.ijert.org/papers/IJCRT2007599.pdf. Accessed 11 Apr. 2025.
- [8]. Moraye, Kiran, et al. "Crop Yield Prediction Using Random Forest Algorithm for Major Cities in Maharashtra State." *International Journal of Innovative Research in Computer Science & Technology*, vol. 9, no. 2, Mar. 2021, pp. 40–44, <https://doi.org/10.21276/ijircst.2021.9.2.7>. Accessed 28 June 2021.
- [9]. N., Varshitha D., and Savita Choudhary. "An Artificial Intelligence Solution for Crop Recommendation." *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 3, 1 Mar. 2022, p. 1688, <https://doi.org/10.11591/ijeecs.v25.i3.pp1688-1695>.
- [10]. None Yash Gupta, and None Garima Srivastava. "Crop Recommendation System for Growing Best Suitable Crop." *International Journal of Science and Research Archive*, vol. 12, no. 1, 30 June 2024, pp. 2928–2936, [ijrsra.net/content/crop-recommendation-system-growing-best-suitable-crop](https://doi.org/10.30574/ijrsra.2024.12.1.1111), <https://doi.org/10.30574/ijrsra.2024.12.1.1111>.
- [11]. Pavan Kumar.Innamuri, et al. "Soil Classification and Crop Prediction Using Machine." *Zenodo (CERN European Organization for Nuclear Research)*, vol. 7, no. 5, 25 Sept. 2023, <https://doi.org/10.5281/zenodo.8374792>. Accessed 11 Apr. 2025.
- [12]. Prabhu, Shubham, et al. "Soil Analysis and Crop Prediction." *International Journal of Scientific Research in Science and Technology*, vol. 7, no. 4, 29 July 2020, pp. 117–123, www.researchgate.net/publication/347382419_Soil_Analysis_and_Crop_Prediction, <https://doi.org/10.32628/IJSRST207433>.
- [13]. S, SharathKumar B, et al. "Soil Analysis and Crop Recommendation Using Machine Learning." *JETIR*, vol. 11, no. 5, 2025, pp. g212–g218g212–g218, www.jetir.org/view?paper=JETIR2405626. Accessed 11 Apr. 2025.
- [14]. Sangita Changdeo Dandwate. "Analysis of Soil Samples for Its Physicochemical Parameters from Sangamner City." *GSC Biological and Pharmaceutical Sciences*, vol. 12, no. 2, 30 Aug. 2020, pp. 123–128, gsconlinepress.com/journals/gscbps/sites/default/files/GSCBPS-2020-0243.pdf, <https://doi.org/10.30574/gscbps.2020.12.2.0243>.
- [15]. T. Swathi, and S. Sudha. "Crop Classification and Prediction Based on Soil Nutrition Using Machine Learning Methods." *International Journal of Information Technology*, vol. 15, no. 6, 1 July 2023, pp. 2951–2960, <https://doi.org/10.1007/s41870-023-01345-0>. Accessed 11 Apr. 2025.
- [16]. Venkata, Immanni, and Narayana Chowdary. "Crop Recommendation Using Machine Learning." *International Journal of Research Publication and Reviews Journal Homepage: Www.ijrpr.com*, vol. 4, no. 12, 2023, pp. 4007–4009, ijrpr.com/uploads/V4ISSUE12/IJRPR20637.pdf. Accessed 11 Apr. 2025.