A Hierarchical Edge-Cloud Machine Learning Framework for Real-Time Soil Nutrient Monitoring and Crop Recommendation

Niranjan S Kaithota [1,*], TP Mohith[2], Yashpreet Goyal[3], Jayathi R[4], G Vijaya Kumar[5]

[1,2,3]*Department of AIML, RV College of Engineering, Bengaluru, India*
[4]*Department of Aerospace, RV College of Engineering, Bengaluru, India*
[5]*Department of Biotechnology, RV College of Engineering, Bengaluru, India*
[1]niranjansk.ai23@rvce.edu.in, [2]tpmohith.ai23@rvce.edu.in,
[3]yashpreetgoyal.ai23@rvce.edu.in, [4]jayathir.as23@rvce.edu.in,
[5]vijayakg@rvce.edu.in

Abstract

Agriculture plays a vital role in feeding the growing global population. But optimizing crop production and resource management remains a significant challenge for farmers. This research paper proposes an innovative ML-enabled IoT device to monitor soil nutrients and provide accurate crop recommendations. The device utilizes the pH and temperature sensor, Moisture sensor, and NPK sensor to collect real-time data on soil composition, moisture, humidity, temperature, and for nutrient levels. The collected data is transmitted to a server using the MQTT protocol. Machine learning algorithms are employed to analyze the collected data and generate customized recommendations, including a possible high-yielding alternate crops for crop rotation, fertilizer names, and its amount based on crop requirements and soil nutrients. As a result, it has become possible to determine the quality of the produce at the consumer level through the web application. The system's effectiveness is evaluated through field experiments, comparing its performance with traditional methods. The results demonstrate the device's ability to enhance crop productivity and optimize resource utilization, promoting sustainable agricultural practices and food security. The research contributes to IoT-enabled agriculture, demonstrating the potential of ML techniques in improving soil nutrient management, facilitating informed decision-making about crop fertilizers, and assessing the quality of produced crops at the consumer level.

Keywords: Precision Agriculture; IoT; Machine Learning; Soil Nutrients; Crop Recommendation; Sustainable Farming

1. Introduction

In recent years, there has been a growing emphasis on leveraging these technologies to address the challenges faced by farmers in optimizing crop production and reducing resource wastage [2]. Among these challenges, the effective management of soil nutrients and accurate crop recommendations [3] play pivotal roles in ensuring sustainable and efficient agricultural practices. Precision Agriculture (PA) is comprised of near and remote sensing techniques using IoT sensors, which help to monitor crop states at multiple growth levels. PA involves the acquisition and processing of a large amount of data related to crop health. Multiple parameters are involved in plants health, including water level, temperature and others. PA enables a farmer to know precisely what parameters are needed for healthy crop, where these parameters are needed and in what amount at a particular instance of time [4]. The ML algorithms employed in this device utilize sophisticated data modeling techniques, such as random forest, logistic regression, LGBM, and neural networks, to identify and understand the intricate relationships between soil nutrient levels, environmental factors, and crop requirements. By leveraging historical data, the ML algorithms are trained to generate accurate predictions and recommendations tailored to specific crops, taking into account their nutrient demands, growth stages, and environmental conditions. The outcomes of this research endeavor have the potential to revolutionize the way farmers manage soil nutrients and make crop decisions. By providing real-time, accurate data on soil conditions and customized crop recommendations, the proposed ML-enabled IoT device aims to empower farmers with the tools and knowledge necessary to enhance productivity, optimize resource utilization, food security, and promote sustainable agricultural practices.

2. Related Works

Several studies have explored the integration of IoT and machine learning in agriculture, particularly for soil nutrient monitoring and crop recommendation. Islam et al. [1] developed a machine learning-enabled IoT system for real-time soil nutrient monitoring and crop recommendation. Their work provides the foundation for building a local model that collects nutrient data using IoT sensors and applies machine learning algorithms for decision-making. This system demonstrates how localized data collection can be leveraged to improve agricultural practices through data-driven insights.

Quy et al. [2] reviewed the architecture and applications of IoT in smart agriculture, highlighting the role of scalable and flexible IoT infrastructure for real-time environmental monitoring. Their work supports the integration of a robust local sensing system that communicates with a centralized global decision model. Vangala et al. [3] addressed the security concerns in IoT-enabled agriculture and proposed architecture-level solutions to ensure data integrity and secure communication. These aspects are essential when transmitting soil data from local devices to global models.

Shafi et al. [4] provided a comprehensive overview of precision agriculture techniques, emphasizing the importance of site-specific monitoring and management. Their findings support the concept of localized soil analysis for personalized fertilizer and crop recommendations, reinforcing the need for a local-global model split.

Rodriguez-Galiano et al. [5] evaluated the effectiveness of the Random Forest classifier in land cover classification tasks. Their work validates the suitability of Random Forest as a global model for predicting fertilizer needs and crop yield based on the cleaned and structured soil nutrient data. Furthermore, Xu et al. [6] proposed an enhanced anomaly detection approach based on Isolation Forest, which is well-suited for detecting and removing faulty or noisy data in real-time sensor streams. This forms the basis of the anomaly filtering mechanism within the local model.

Collectively, these works contribute to a two-tiered architecture where a local model handles real-time soil nutrient monitoring and anomaly detection using techniques like Isolation Forest, while a global model uses filtered data to perform fertilizer recommendation and yield prediction via supervised learning models like Random Forest. This separation of responsibilities ensures both data quality and prediction accuracy while enabling scalability and robustness in smart agriculture systems.

3. Methods and Materials

In this section, we will outline the methodologies and materials employed in this study, encompassing the IoT devices, sensors 2.1, data transmission and analysis techniques 2.2, machine learning model design and implementation 2.3, and the proposed framework 2.4.

3.1. IOT devices and sensors

This study used three categories' sensors like pH and Temperature sensor 3.1.2, Moisture Sensor 3.1.3, NPK sensor 3.1.4, and one Node MCU 3.1.1. The circuit diagram is presented in Fig. 1. And Table 2 represents the configurations and specifications of the proposed IoT System Components.

3.1.1. Node MCU.. The NodeMCU is an integral component used in this research for the implementation of IoT capabilities. It is a versatile and widely adopted microcontroller board. The NodeMCU board has a voltage regulator to supply a steady 3.3 V power supply as well as a USB interface for programming and powering the device. It is based on the ESP8266 Wi-Fi module, which provides seamless connectivity and enables wireless communication with other devices and the internet. The NodeMCU offers a range of functionalities and features, including 11 GPIO (General Purpose Input/Output) pins, analog-to-digital converters, and programmable interfaces, making it suitable for various applications in the field of IoT. Its compact size, low power consumption, and compatibility with the Arduino development platform, Arduino IDE, make it a popular choice for IoT projects. In this research, the NodeMCU serves as a vital component for collecting sensor data, transmitting it wirelessly using MQTT protocol, and facilitating seamless integration with machine learning algorithms.

3.1.2. pH and Temperature sensor.. In a soil monitoring model using a Random Forest classifier, pH and temperature sensors play a vital role in enhancing accuracy and decision-making. Soil pH directly affects nutrient availability; imbalances can lead to nutrient lock-out, even when nutrients are present. Real-time pH data helps the model detect such issues and recommend corrective actions like lime or sulfur application. Temperature influences microbial activity, nutrient solubility, and organic matter breakdown—all essential for soil fertility. By monitoring soil temperature, the model can account for changes in biological processes that impact nutrient dynamics. Including these sensor readings as input features improves the classifier's ability to identify soil health categories more precisely. Together, these sensors enable the system to deliver accurate, real-time assessments and promote smarter, data-driven soil management practices.

3.1.3. NPK sensor.. The NPK sensor plays a pivotal role in this research as a key component for data collection and analysis. This sensor offers advanced capabilities for monitoring and measuring essential soil parameters. Specifically designed for agricultural applications, the NPK sensor enables accurate and real-time measurements of soil NPK. These measurements are crucial for assessing soil health, nutrient levels, and overall environmental conditions. The NPK sensor boasts a robust construction, high sensitivity, and reliable performance, making it a valuable tool for gathering precise and actionable data. It features quick responses, excellent interchange capabilities, and high precision agriculture sensors. This sensor was chosen because it accurately sense data. Its integration into the research framework facilitates comprehensive soil monitoring, aiding in the development of innovative machine learning algorithms for soil nutrient analysis and crop recommendation.

3.1.4. Moisture Sensor.. The moisture sensor is essential for evaluating soil health and optimizing irrigation. Soil moisture directly influences plant growth, nutrient absorption, and microbial activity. If moisture levels are too low, plants experience water stress, and nutrient uptake is hindered. Excess moisture, on the other hand, can lead to root rot and nutrient leaching. The moisture sensor provides real-time data that helps the model assess current soil conditions and predict potential stress factors. By including moisture as a key input feature, the Random Forest classifier can more accurately categorize soil health and recommend appropriate interventions, such as irrigation scheduling or drainage improvements. This ensures efficient water use and supports sustainable farming practices.

Figure 1. Sustainable Farming Dashboard showing real-time nutrient monitoring interface. The dashboard displays NPK gauge indicators (N: 146, P: 56, K: 209), along with temperature (19.7℃), humidity (80.3%), and pH (5.26) readings for pomegranate cultivation. The interface provides visual indicators of nutrient levels relative to optimal ranges for the selected crop.

3.2. Data transmission using MQTT protocol

The data transmission section of the research paper focuses on the implementation of the MQTT (Message Queuing Telemetry Transport) protocol for efficient data transmission in the ML-enabled IoT device for soil nutrients monitoring and crop recommendation. This section outlines the methods and techniques used for utilizing the MQTT protocol for data transmission, and storage. A suitable MQTT broker needs to be chosen, considering factors such as scalability, reliability, and support for Quality of Service (QoS) levels. In this project, a cloud-based MQTT broker is chosen, as it provides easy scalability, high availability, and seamless integration with other cloud services. After that, MQTT topics are designed to represent different aspects of soil nutrients and crop-related data, enabling efficient organization and routing of messages. The collected data is formatted into MQTT payloads, encapsulating the relevant soil parameters, crop information, and other contextual data. Based on the criticality and reliability requirements of the data, the appropriate QoS level is selected. QoS levels 0 (at most once), 1 (at least once), or 2 (exactly once) are considered for message delivery. The IoT device is equipped with an MQTT client that connects to the MQTT broker and publishes the collected data to the respective MQTT topics. It publishes data to specific topics, while other devices or subscribers can subscribe to these topics to receive the data in real-time. For the purpose of the secured data storage and management, the MQTT broker is integrated with cloud storage services, enabling the seamless storage and retrieval of MQTT data in scalable storage systems. And, the received MQTT messages are persisted in cloud storage for long-term data retention and subsequent analysis. Finally, raw data received from the MQTT broker undergoes preprocessing steps, including filtering and cleansing, to remove noise, outliers, or erroneous readings and is transformed into a suitable format for analysis, ensuring compatibility with ML algorithms and further processing steps.

3.3. Machine learning methods

This section explains the machine learning algorithms integrated into the analysis pipeline to process the data and generate relevant insights and recommendations for soil nutrients monitoring and crop management. In our study, we used Random Forest with grid searchCV 3.3.1 and Isolation Forest 3.3.2 to recommend crop, and fertilizer, respectively.

3.3.1. Random Forest classifier. Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. In the context of the fertilizer recommendation, Random Forest plays a crucial role in analyzing various factors such as soil properties, crop characteristics, and environmental conditions to suggest the most appropriate fertilizer for optimal plant growth. The algorithm's ability to handle complex interactions and nonlinear relationships between variables makes it well-suited for this task. By utilizing an ensemble of decision trees, Random Forest reduces overfitting and improves prediction accuracy. It also provides insights into feature importance, aiding in understanding the factors that influence fertilizer recommendations. The base classifiers can be defined as (3)

$$h(x, \phi_m), m = 1, \ldots \tag{1}$$

Where, x is input vector and $\{\phi_m\}$ are the independent and identically distributed random vectors [5]. Then, the random forest approach can be defined as (4)

$$H(x) = \{h_1(x), h_2(x), \ldots, h_k(x)\} \tag{2}$$

3.3.2. Isolation Forest.. The Isolation Forest algorithm is highly effective for detecting anomalies in soil monitoring models, particularly when dealing with sensor faults or unexpected environmental changes. Its application is justified by two core properties of anomaly data: (a) anomalies represent a small fraction of the dataset, and (b) their feature values (e.g., extreme pH, moisture, or temperature) significantly differ from those of normal soil conditions. In a soil monitoring context, Isolation Forest builds an ensemble of isolation trees (iTrees) that recursively partition the sensor data space. Each iTree isolates observations by randomly selecting a feature (e.g., moisture or temperature) and a split value between the minimum and maximum of that feature. The average path length from the root to a data point across all trees indicates its degree of isolation. Anomalies are expected to have shorter path lengths due to their distinct nature, and this is quantified by the anomaly score:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{3}$$

where h(x) is the path length for data point x, E(h(x)) is its expected value across the forest, and c(n) is the average path length of unsuccessful searches in Binary Search Trees, approximated by c(n)=2H(n1)2(n1)/n with H(i) being the harmonic number. In the soil model, this algorithm enables real-time detection of faulty or abnormal sensor readings, enhancing data reliability and ensuring accurate soil classification and recommendations [6].

4. Results

Although the system has not yet undergone field testing, the initial development and simulated evaluations provide promising insights into its potential effectiveness. The proto- type successfully integrates multiple sensors, including JXBS- 3001 for NPK detection, FC-28 for soil moisture, and DHT11 for temperature and humidity. These sensors were interfaced with the NodeMCU microcontroller, which transmitted real- time data to a remote server using the MQTT protocol. This validated the system's capability for continuous and wireless monitoring of key soil parameters. On the machine learning front, Random Forest was used for crop and fertilizer recommendations based on input features such as N, P, K, temperature, humidity, moisture, soil type, and crop type. The model was trained on a merged and cleaned dataset and evaluated through cross-validation. It demonstrated high performance in terms of classification metrics, with precision, recall, and F1-scores exceeding 90 The architecture was designed to support a hierarchical decision-making model, combining both edge-level processing

for quick anomaly detection and cloud-level analytics for more complex inference tasks. This hybrid setup was simulated and found to be scalable and efficient, especially suitable for resource-constrained rural environments. Despite these promising findings, it is important to note that the system has yet to be validated under real-world agricultural conditions. Future testing will focus on evaluating the system's adaptability to different soil types, weather condi- tions, and crop varieties across diverse geographical locations. Field trials will also assess long-term performance, sensor calibration stability, and real-time responsiveness. The web application showed stable performance during simulations and is expected to provide farmers with convenient access to the recommendation platform. By integrating this system with edge- level intelligence, the research lays a strong foundation for future advancements in precision agriculture.

5. Discussion and Future Research

The aim of this study was to develop an innovative ML- enabled IoT device for soil nutrient monitoring and crop recommendation. The results indicate that the proposed device and framework hold significant potential in precision agricul- ture. A key contribution of this research is the seamless inte- gration of IoT technology with machine learning algorithms for real-time soil nutrient monitoring. By deploying sensors such as JXBS-3001 (NPK sensor), FC-28 (moisture sensor), and DHT11 (temperature and hu- midity sensor) in the crop field, we were able to collect essential data related to soil nutrient concentrations (NPK), moisture, humidity, and temperature. This data was transmitted to a remote server using the NodeMCU and MQTT protocol, enabling continuous monitoring and analysis. The use of machine learning algorithms, particularly Ran- dom Forest, played a crucial role in the crop recommendation process. These algorithms were trained on merged datasets obtained from the IoT sensors, resulting in accurate predictions for the most suitable crops and fertilizers. The Random Forest algorithm proved effective in recommending fertilizers based on key parameters such as Nitrogen (N), Potassium (K), Phosphorus (P), temperature, humidity, soil type, crop type, and soil moisture. Performance evaluation revealed high precision, recall, and F1-scores for most classes, indicating the approach's effec- tiveness in crop and fertilizer recommendations. However, some errors were noted that could impact crop selection and fertilization decisions, highlighting the need for further refinement to improve overall accuracy. Integration of the ML-enabled IoT device with a cloud- based recommendation platform offers convenient access and scalability. By deploying the web application, the system ensures high uptime and accessibility from any device. This cloud-hosted platform provides a fully functional recommendation system, empowering farmers to make informed decisions regarding crop selection and fertilizer application.

A. Future Research Directions 1) Expanding capabilities: Further development is needed to accommodate a broader range of crops and geo- graphic regions, allowing more farmers to benefit from the monitoring and recommendation functionalities. 2) Integration of satellite imagery and weather data: Incorporating these additional data sources will enable more comprehensive and precise recommendations by providing insights into crop health, growth patterns, and environmental conditions. 3) Enhancing scalability and usability: Improving device adaptability to diverse communication networks and power sources, along with developing user-friendly in- terfaces for farmers with varying technological expertise, will increase usability and adoption. By addressing these future directions, this research can ad- vance the ML-enabled IoT device's applicability in agriculture, enabling smarter decisions and improved crop productivity. Overall, this study demonstrates the potential of ML- enabled IoT devices in precision agriculture by combining sensor data, machine learning algorithms, and cloud-based platforms for real-time monitoring, analysis, and decision-making. The system offers benefits such as high precision, real-time insights, adaptability, and support for sustainable farming practices. However, challenges remain, including data availability, system complexity, cost, and maintenance needs. The choice between this advanced system and traditional methods depends on the specific agricultural context, resource availability, and user expertise. Importantly, this device holds particular promise for farmers lacking comprehensive agri- cultural knowledge, providing them with valuable decision support.

6. Conclusion

This study successfully combined IoT technology, sensor data collection, and machine learning algorithms to develop an innovative ML-enabled IoT device for soil nutrient monitoring and crop recommendation. Utilizing multiple sensors to collect real-time data on soil nutrient concentrations, moisture, humidity, and temperature, the system transmits information to a remote server for analysis. The application of the Random Forest algorithm proved effective in predicting suitable crops and recommending ap- propriate fertilizers. While further improvements are necessary to reduce errors and enhance model accuracy, this ML-enabled IoT device shows great promise in delivering real-time insights to farmers for optimal crop management. Addressing current challenges and exploring future research directions will continue to enhance this device's capabilities and contribute to advancing precision agriculture

Conflict of Interest

The authors declare no conflict of interest in the research, authorship, and publication of this article. This research was conducted without financial support from commercial entities that could influence the study design, data collection, analysis, or interpretation of results.

References

[1] M. R. Islam, K. Oliullah, M. M. Kabir, M. Alom, and M. F. Mridha, "Machine Learning Enabled IoT System for Soil Nutrients Monitoring and Crop Recommendation," Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh, 2023.

[2] V.K. Quy, N.V. Hau, D.V. Anh, N.M. Quy, N.T. Ban, S. Lanza, G. Randazzo, A. Muzirafuti, Iot-enabled smart agriculture: architecture, applications, and challenges, Appl. Sci. 12 (7) (2022) 3396, https://doi.org/10.3390/app12073396.

[3] A. Vangala, A.K. Das, V. Chamola, V. Korotaev, J.J. Rodrigues, Security in iotenabled smart agriculture: architecture, security solutions and challenges, Cluster Comput. 26 (2) (2023) 879–902, https://doi.org/10.1007/s10586-022-03566-7.

[4] U. Shafi, R. Mumtaz, J. García-Nieto, S. A. Hassan, S. A. R. Zaidi, and N. Iqbal, ”Precision Agriculture Techniques and Practices: From Considerations to Applications,” Sensors, vol. 19, no. 17, p. 3796, 2019.

[5] V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez, An assessment of the effectiveness of a random forest classifier for land-cover classification, ISPRS J. Photogrammetry Remote Sens. 67 (2012) 93–104, https://doi.org/10.1016/j.isprsjprs.2011.11.002

[6] D. Xu, Y. Wang, Y. Meng and Z. Zhang, ”An Improved Data Anomaly Detection Method Based on Isolation Forest,” 2017 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 2017, pp. 287-291.

[7] S.V. Gaikwad, A.D. Vibhute, K.V. Kale, S.C. Mehrotra, An innovative iot based system for precision farming, *Comput. Electron. Agric.* **187** (2021), 106291.

[8] L. García, L. Parra, J.M. Jimenez, J. Lloret, P. Lorenz, Iot-based smart irrigation systems: an overview on the recent trends on sensors and iot systems for irrigation in precision agriculture, *Sensors* **20**(4) (2020) 1042, https://doi.org/10.3390/s20041042https://doi.org/10.3390/s20041042.

[9] X. Ge, J. Wang, J. Ding, X. Cao, Z. Zhang, J. Liu, X. Li, Combining uav-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring, *PeerJ* **7** (2019), e6926, https://doi.org/10.7717/peerj.6926https://doi.org/10.7717/peerj.6926.

[10] T. Hengl, J.G. Leenaars, K.D. Shepherd, M.G. Walsh, G.B. Heuvelink, T. Mamo, H. Tilahun, E. Berkhout, M. Cooper, E. Fegraus, et al., Soil nutrient maps of subsaharan africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning, *Nutrient Cycl. Agroecosyst.* **109** (2017) 77–102.

[11] A.A. Ibrahim, R.L. Ridwan, M.M. Muhammed, R.O. Abdulaziz, G.A. Saheed, Comparison of the catboost classifier with other machine learning methods, *Int. J. Adv. Comput. Sci. Appl.* **11**(11) (2020).

[12] A. Ikram, W. Aslam, R.H.H. Aziz, F. Noor, G.A. Mallah, S. Ikram, M.S. Ahmad, A. M. Abdullah, I. Ullah, Crop yield maximization using an iot-based smart decision, *J. Sens.* (2022), https://doi.org/10.1155/2022/2022923https://doi.org/10.1155/2022/2022923.

[13] A. Ingle, Crop recommendation dataset. https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset, 2020.

[14] M.R. Islam, M.M. Kabir, M.F. Mridha, S. Alfarhood, M. Safran, D. Che, Deep learning-based iot system for remote monitoring and early detection of health issues in real-time, *Sensors* **23**(11) (2023) 5204, https://doi.org/10.3390/s23115204https://doi.org/10.3390/s23115204.

[15] D. Jayashree, O. Pandithurai, L. Paul Jasmin Rani, P.K. Menon, M.V. Beria, S. Nithyalakshmi, Fertilizer recommendation system using machine learning, *Disruptive Technologies for Big Data and Cloud Applications: Proceedings of ICBDCC 2021*, Springer, 2022, pp. 709–716.

[16] R. Jeevaganesh, D. Harish, B. Priya, A machine learning-based approach for crop yield prediction and fertilizer recommendation, *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, 2022, pp. 1330–1334.

[17] G.K. Jha, P. Ranjan, M. Gaur, A machine learning approach to recommend suitable crops and fertilizers for agriculture. *Recommender System with Machine Learning and Artificial Intelligence: practical Tools and Applications in Medical, Agricultural and Other Industries* (2020) 89–99, https://doi.org/10.1002/9781119711582.ch5https://doi.org/10.1002/9781119711582.ch5.

[18] F. Kiani, A. Seyyedabbasi, S. Nematzadeh, F. Candan, T. Çevik, F.A. Anka, G. Randazzo, S. Lanza, A. Muzirafuti, Adaptive metaheuristic-based methods for autonomous robot path planning: sustainable agricultural applications, *Appl.*.

[19] V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez, An assessment of the effectiveness of a random forest classifier for land-cover classification, *ISPRS J. Photogrammetry Remote Sens.* **67** (2012) 93–104, https://doi.org/10.1016/j.isprsjprs.2011.11.002https://doi.org/10.1016/j.isprsjprs.2011.11.002.

[20] M.K. Senapaty, A. Ray, N. Padhy, Iot-enabled soil nutrient analysis and crop recommendation model for precision agriculture, *Computers* **12**(3) (2023) 61.