IoT-Enabled K-Nearest Neighbors (KNN)-Based Soil Nutrient Recommendation System for Rice (*Oryza Sativa L.*) Cultivation

Rannie M. Sumacot[1]* and Jessie R. Paragas[2]
[1]Public Administration Department, Faculty of Governance and Development Studies,
Southern Leyte State University
Southern Leyte, Philippines 6608
[2]Information Technology Department, College of Engineering, Eastern Visayas State University
Leyte, Philippines 6500
Email: [1]rsumacot@southernleytestateu.edu.ph, [2]jessie.paragas@evsu.edu.ph

***Abstract*—Soil nutrient assessment is essential for optimizing crop yield. Still, existing machine learning-based soil nutrient recommendation systems face several challenges, including limited real-time adaptability, inconsistent integration with IoT frameworks, and a lack of scalability for smallholder use. Many of these systems rely heavily on static, pre-collected datasets and lack the capability to respond dynamically to field conditions. These limitations reduce the practical effectiveness of such models in achieving precision agriculture goals, particularly in resource-constrained environments. These limitations also hinder efficient soil fertility management, leading to ineffective fertilizer application, nutrient imbalances, and reduced crop productivity. To address these issues, the researchers develop an Internet of Things (IoT)-enabled K-Nearest Neighbors (KNN)-based soil nutrient recommendation system specifically for rice (Oryza Sativa L.) cultivation. The system integrates an RS485 Integrated Soil Nitrogen, Phosphorus, and Potassium (NPK) Sensor with an Arduino-based IoT framework to continuously monitor essential nutrients: nitrogen (N), phosphorus (P), and potassium (K). The collected data are processed using Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Linear Regression, and KNN. After rigorous model training on Google Colab, KNN outperforms other models with an accuracy of 98%, making it the optimal choice for predictive soil fertility assessment. This system provides accurate and automated fertilizer recommendations to improve soil management efficiency and sustainability by combining real-time IoT monitoring with machine learning. The research contributes to precision agriculture by offering a scalable data-driven approach that enhances crop yield, reduces fertilizer waste, and minimizes environmental impact.**

***Index Terms*—Soil Nutrient Recommendation, Internet of Things (IoT), Microcontroller, Soil Nitrogen, Phosphorus, and Potassium (NPK), Machine Learning, K-Nearest Neighbor (KNN)**

Received: Nov. 14, 2024; received in revised form: April 30, 2025; accepted: May 03, 2025; available online: Sep. 04, 2025.
*Corresponding Author

I. Introduction

THE transformation of agriculture into a technology-driven sector creates new opportunities for improving productivity, efficiency, and sustainability in crop cultivation [1–3]. Rice (Oryza Sativa L.) remains a staple food for much of the global population, making advancements in its cultivation vital for addressing food security [4]. In the Philippines, rice production has experienced significant fluctuations [5]. According to the Philippine Rice Research Institute [6], production peaked at 8,651 metric tons in 2017 before dropping to 4,068 metric tons in 2019, with yield per hectare ranging between 2.74 and 2.95 metric tons. These variations underscore the dynamic and often challenging nature of rice cultivation in response to shifting soil and environmental conditions. Thus, increasing farm-level production efficiency has become critical, necessitating technological interventions to address these agricultural challenges [7].

According to the previous studies [8, 9], effective rice farming depends on accurate soil nutrient management, particularly for nitrogen (N), phosphorus (P), and potassium (K). It is collectively known as NPK. Traditional laboratory methods for soil testing are labor-intensive and time-consuming and often lack real-time adaptability to field conditions [10–12]. Recent advancements in IoT offer a promising solution through real-time soil nutrient monitoring, enabling farmers to access immediate insights into nutrient levels and adjust fertilization practices accordingly [13, 14]. Previous research has demonstrated the potential of IoT-enabled soil nutrient classification and crop recommendation models to support precision agriculture. It reduces unnecessary fertilizer use and optimizes productivity [15].

Several studies have attempted to enhance soil nutrient assessment using machine learning techniques. The first research has utilized machine learning algorithms to evaluate soil nutrients in forest ecosystems [16]. Meanwhile, the second research has focused on optimizing extreme learning machine parameters to improve soil classification accuracy [17]. More advanced approaches, such as deep learning, have been explored by the third and fourth research [18, 19]. They have utilized deep learning models for classifying soil nutrients and pH levels in agricultural settings. However, despite these advancements, deep learning models often require high computational resources, large labeled datasets, and sophisticated infrastructure, making them less practical for real-time and cost-effective soil fertility monitoring in small-scale farming environments.

The research addresses these gaps by developing an IoT-enabled and real-time soil nutrient recommendation system using the K-Nearest Neighbors (KNN) algorithm. The algorithm is computationally efficient and well-suited for deployment in resource-constrained environments. By integrating an RS485 Integrated Soil NPK Sensor with an Arduino-based IoT framework, this system provides continuous and real-time monitoring of soil nutrients and automated fertilizer recommendations, which is cost-effective for small-scale farmers.

The research also aims to further innovate in this area by integrating IoT with machine learning algorithms to enhance the accuracy of soil nutrient testing and analysis. The research focuses on the convergence of Internet of Things (IoT) devices and machine learning algorithms to enhance rice farming, particularly in monitoring essential soil nutrients that drive crop yield. Utilizing Arduino-based IoT devices for real-time soil monitoring, the researchers explore the application of five machine learning algorithms, such as Naive Bayes, Support Vector Machine (SVM), KNN, Linear Regression, and Logistic Regression, to assess soil nutrient data and make predictive recommendations. Each algorithm brings unique advantages: Naive Bayes offers probabilistic classification [20], and SVM is optimized for complex classification tasks [21]. Meanwhile, KNN classifies things based on spatial proximity [22], while Linear and Logistic Regression provide predictive analytics based on variable relationships [23, 24]. By leveraging these capabilities within a supervised learning framework, the researchers aim to enhance soil testing efficiency and empower farmers with data-driven insights.

Ultimately, this research advances a machine learning-based soil nutrient recommendation system designed to optimize fertilizer usage, increase crop yields, and contribute to sustainable farming practices in rice cultivation. Aligned with the United Nations Sustainable Development Goals (SDGs) for Zero Hunger (SDG 2) and Industry, Innovation, and Infrastructure (SDG 9), the research underscores the potential of IoT and machine learning integration in fostering food security and resilient agricultural infrastructure [25, 26]. Through this multidimensional approach, the research addresses urgent needs in agriculture by providing a scalable and technology-driven solution for sustainable crop management.

II. Research Method

A. Research Design

The research employs a quantitative research methodology due to its ability to provide precise and objective measurements in evaluating the integration of Arduino-based Internet of Things (IoT) devices and machine learning algorithms in rice farming [27]. The quantitative approach facilitates the numerical assessment of key parameters, such as nutrient levels and crop yield, which enables statistical analysis and enhances the generalizability of findings across diverse agricultural contexts. Specifically, the research aims to compare the performance of various machine learning algorithms, to identify the algorithm that exhibits the highest predictive accuracy. This methodology ensures a systematic evaluation of algorithmic effectiveness and offers valuable insights into their practical applicability within the agricultural sector.

Figure 1 illustrates the sequential stages of the research methodology. It starts with data collection, where IoT sensors gather real-time soil NPK data. The collected data undergoes preprocessing to clean and prepare it for machine learning analysis. Next, multiple machine learning models (KNN, SVM, Naive Bayes, and others) are trained and evaluated. The selected model is then integrated into the IoT-based system, enabling real-time soil nutrient monitoring. As IoT sensors continuously track soil conditions, the system utilizes the trained model to generate fertilizer recommendations, ensuring optimal nutrient management. This structured approach enhances precision agriculture by combining real-time IoT data with machine learning-driven decision-making.

B. System Architecture

The system, as shown in Fig. 2, begins with a Soil NPK Sensor that collects real-time data on essential soil nutrients (nitrogen (N), phosphorus (P),

Cite this article as: R. M. Sumacot and J. R. Paragas, "IoT-Enabled K-Nearest Neighbors (KNN)-Based Soil Nutrient Recommendation System for Rice (*Oryza Sativa L.*) Cultivation", CommIT Journal 19(2), 151–165, 2025.

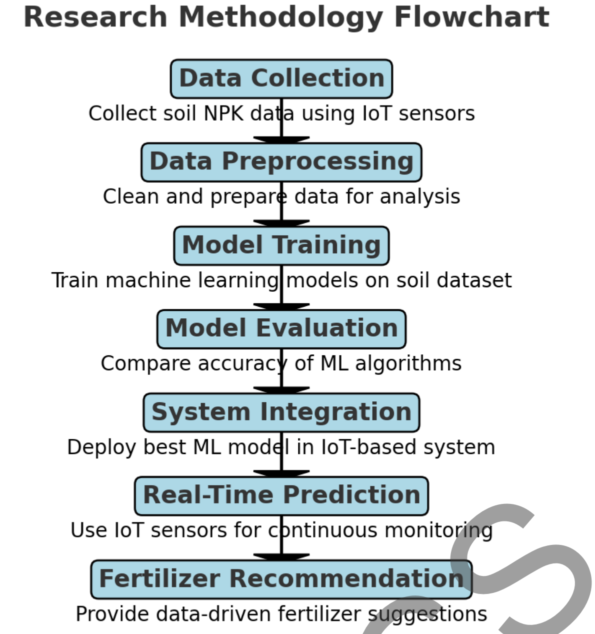

Fig. 1. Research methodology flowchart. Note: Nitrogen (N), Phosphorus (P), and Potassium (K), Internet of Things (IoT), and Machine Learning (ML).

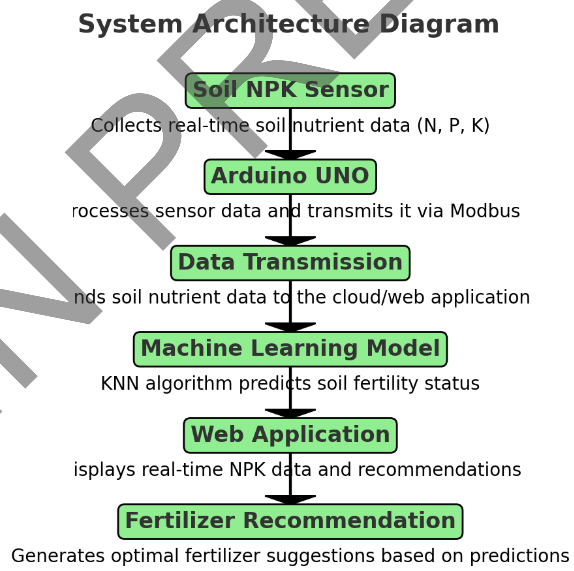

Fig. 2. System architecture diagram. Note: Nitrogen (N), Phosphorus (P), Potassium (K), and K-Nearest Neighbors (KNN).

and potassium (K)). The data are sent to the Arduino UNO, which acts as the microcontroller by processing sensor readings and transmitting them via the Modbus communication protocol. The processed data are then transmitted to a cloud-based web application, where the machine learning model (KNN algorithm) analyzes the soil data to predict soil fertility levels. Next, the web application serves as the user interface, displaying real-time NPK values and machine learning-based fertilizer recommendations. The final output is a fertilizer recommendation, guiding farmers on the optimal nutrient requirements for improving soil fertility.

C. Software

Software plays a pivotal role in the system's implementation, enabling essential functions for programming, integration, and execution. Together, the software tools establish a cohesive infrastructure for coding, debugging, and executing the different system components. Each tool plays a specific role, ensuring that the system functions effectively and fulfills its objectives.

The Arduino Software Integrated Development Environment (IDE) 2.0 is utilized as a core platform for coding the Arduino board to which the soil nutrient sensor is connected. This open-source software provides a user-friendly environment for writing, compiling, and uploading code, as well as integrating necessary libraries and tools to ensure seamless communication between the board and connected sensors [28].

Next, Microsoft Visual Studio 2019 serves as the development environment for the Windows application that displays real-time NPK values from the soil sensor. This comprehensive IDE offers a suite of tools, including a code editor and debugger, to streamline the coding, testing, and deployment of the application [29]. Visual Studio 2019's robust capabilities make it a suitable choice for developing applications that require consistent and accurate data display.

In addition, Visual Studio Code (VS Code) is used for programming the Python code that integrates with the system's web application. This application incorporates the machine learning model, displaying predictive results based on soil data. VS Code's versatility, extensive extensions, and support for multiple programming languages make it an ideal tool for building, testing, and deploying the Python code central to this web application [30].

Moreover, Google Colaboratory (Colab) is essential in training the dataset to create the machine learning model for the soil nutrient recommendation system. Developed by Google Research, Colab offers a browser-based platform for Python coding, with free access to computing resources such as Graphics Processing Units (GPUs). Hence, it is particularly advantageous for handling large datasets and developing machine learning models [31].

D. Hardware

The hardware components serve as the primary input devices for the system, playing a crucial role in data measurement and processing. Collectively, the hardware components are the foundation of the system, facilitating the collection of crucial data, such as soil nutrient levels, and ensuring robust communication across various modules. Their seamless integration enables the efficient operation of the system, supporting the achievement of the system's objectives.

One of the key devices used is the RS485 Integrated Soil NPK Sensor, which is responsible for measuring soil nutrient levels, specifically nitrogen (N), phosphorus (P), and potassium (K). This sensor utilizes the RS485 communication protocol, allowing for long-distance data transmission and reliable communication in industrial and agricultural applications. A significant advantage of this sensor is its high-speed response time, as it can detect and transmit soil nutrient readings in less than one second [32]. This real-time data acquisition capability enhances the efficiency of the system by ensuring that nutrient levels are updated almost instantaneously, making it highly effective for precision agriculture applications [33].

Another essential hardware component is the Modbus module–MAX 485 module (Arduino), which serves as an interface for serial communication in Modbus networks. This module converts data between the Modbus protocol and RS-485 signals, enabling seamless communication between Modbus-compatible devices. Known for its efficiency and reliability, the MAX 485 module is frequently employed in industrial automation, energy management, and building automation systems where robust data transmission over long distances is required [34].

Next, the Arduino UNO R3 serves as the system's main microcontroller, offering both the flexibility and functionality required to manage and control various hardware components. Built around the ATmega328P microcontroller, the Arduino UNO R3 provides 14 digital input/output pins, six analog inputs, a 16 MHz quartz crystal, a Universal Serial Bus (USB) connection, and a power jack. This board plays a central role in the system, as it is where essential code is uploaded to control system operations. It directly interfaces with the NPK sensor, enabling seamless communication and data collection from the sensor and facilitating the integration of additional sensors and modules as needed [35].

E. Dataset Source

The researchers utilize a dataset from Kaggle, sourced from Kumar [36], which consists of 2,200 rows of soil nutrient data and originally includes data for multiple crops. However, for the purposes of the research, the dataset is trimmed to focus exclusively on rice as the crop of interest. The original dataset contains eight attributes. However, only four key attributes are retained due to their direct relevance to soil fertility for rice cultivation.

Each row contains values for nitrogen (N), phosphorus (P), and potassium (K), along with a corresponding classification of fertility level. To effectively train and evaluate machine learning models, the researchers divide the dataset into three subsets: 70% (1,540 samples) for training, enabling the models to learn patterns from the data; 15% (330 samples) for the validation set for model tuning and optimization; and the remaining 15% (330 samples) for the testing set to assess the final model's performance. Additionally, for reference and as an example, an optimal NPK level for rice cultivation includes nitrogen at 90, phosphorus at 42, and potassium at 43, or alternatively, nitrogen at 85, phosphorus at 58, and potassium at 41.

The four selected attributes included nitrogen, phosphorus, potassium, and fertility level. The fertility level attribute is added based on the original dataset. It provides recommendations for soil nutrient levels across various crops, including rice. By narrowing the dataset to rice, the fertility levels are refined to only those corresponding specifically to rice cultivation, ensuring the analysis remains relevant to the research objectives. The nitrogen, phosphorus, and potassium attributes, measured in parts per million (ppm), are essential for determining soil fertility, as they directly impact rice growth. The fertility level, classified based on the specific nutrient content for rice, serves as the classification label for the dataset. Meanwhile, attributes related to other crops are excluded as they are considered irrelevant to the specific focus on rice soil fertility. The data refinement ensures a more targeted and precise approach to predicting soil fertility and enhancing the model's performance about rice cultivation.

Upon obtaining the dataset, the next step in the research process involves meticulous data preprocessing and cleansing. These steps are crucial for ensuring the quality and suitability of the data for subsequent analysis and model development. The dataset initially contains information about various crops, each characterized by specific features. However, given the specific research focus on rice, the researchers opt to streamline the dataset by excluding information related to other crops. This curation process involves removing irrelevant crop entries. It results in a refined dataset that exclusively pertains to the features and characteristics of rice.

By narrowing the dataset to focus solely on rice-related information, the researchers aim to enhance the relevance and specificity of the data for the research objectives. This strategic data curation allows for a more targeted and meaningful analysis. So, the subsequent stages of the research are dedicated to developing a model specifically tailored for predicting soil fertility and recommending the appropriate fertilizer to be applied to the soil.

On the other hand, the RS485 Soil NPK Sensor employs three specific inquiry frames to retrieve values for nitrogen (N), phosphorus (P), and potassium (K) levels in the soil. Each frame, detailed in the sensor's instruction manual, is uniquely designed to capture data for a particular nutrient. The inquiry frame for nitrogen to retrieve soil nitrogen levels is coded as 0×01, 0×03, 0×00, 0×1e, 0×00, 0×01, 0×e4, 0×0c. Similarly, the frame for phosphorus retrieval is structured as 0×01, 0×03, 0×00, 0×1f, 0×00, 0×01, 0×b5, 0×cc, and the frame designated for potassium measurement is 0×01, 0×03, 0×00, 0×20, 0×00, 0×01, 0×85, 0×c0. These inquiry frames enable the sensor to communicate efficiently and deliver accurate data on soil nutrients for each element. Upon sending these inquiry frames, the sensor responds with values that are used for analysis. After receiving a response, the soil nitrogen, phosphorus, or potassium values can be calculated by converting the hexadecimal value (e.g., 0×HH) to decimal (e.g., DD), where DD represents the nutrient concentration in milligrams per kilogram (mg/kg). This analysis phase is pivotal for interpreting and utilizing the received data for accurate nutrient concentration estimations. Recommendations for fertilization or soil management can then be derived based on these calculated values. For the utmost precision, sending soil samples to a laboratory for analysis is still recommended.

Moreover, online datasets have been widely utilized in machine learning studies, demonstrating their reliability in various applications. Notable examples include the previous studies [25], which have leveraged soil nutrient data to enhance classification accuracy using machine learning techniques, and [15], which has integrated an online dataset into an IoT-enabled soil nutrient classification system for precision agriculture. These studies validate the use of publicly available datasets as a credible resource for machine learning-based soil fertility prediction, reinforcing their applicability in precision agriculture research.

F. System Design

The research utilizes two complementary data sources to enhance the accuracy and applicability of the machine learning model. The first source is a publicly available dataset from Kaggle for training and validating the model. The second source involves real-time sensor data, collected using an RS485 Integrated Soil NPK Sensor during system deployment, providing live soil nutrient readings. The trained machine learning model processes these real-time NPK values,

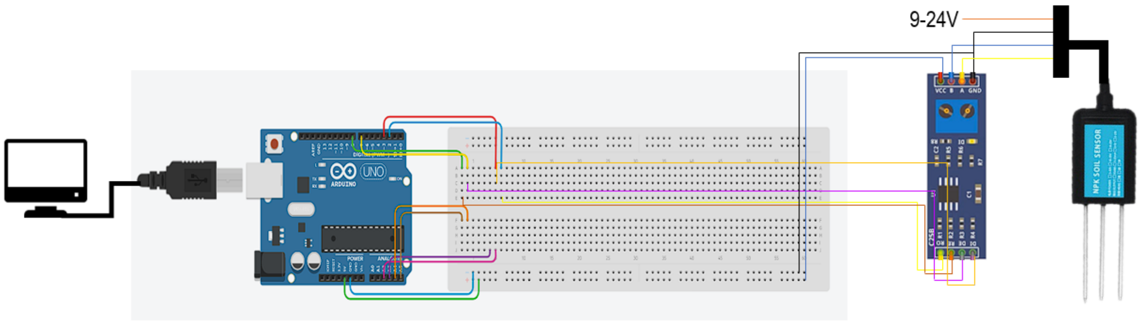

Fig. 3. Hardware architectural design.

classifies the soil fertility level, and offers fertilizer recommendations based on predicted deficiencies. This integration ensures that the model is trained on a diverse dataset while remaining applicable to real-world field conditions.

In the context of incorporating Arduino-based IoT devices and machine learning algorithms into rice farming, the adoption of an Agile Software Development Life Cycle (SDLC) emerges as an empirically validated and fitting choice. This methodology is chosen due to its inherent compatibility with the intricacies of the project [37], providing an iterative and incremental framework that aligns well with the dynamic nature of the agricultural domain. The adoption of an Agile SDLC for the integration of Arduino-based IoT devices and machine learning algorithms in rice farming is rooted in its ability to provide a flexible, collaborative, and iterative framework.

The Agile SDLC represents a flexible and adaptive approach to software development, emphasizing collaboration, rapid prototyping, and continuous improvement throughout the project's lifecycle. This iterative nature allows for the incorporation of feedback at various stages, enabling adjustments and refinements based on evolving requirements and insights gained during the development process. By embracing Agile principles, the proponent aims to create a responsive and resilient framework that can efficiently accommodate changes, a particularly valuable attribute in the dynamic and evolving landscape of agriculture.

The collaborative aspect of Agile is crucial in a project that involves the integration of Arduino-based IoT devices and machine learning algorithms. The iterative development cycles allow for the seamless integration of these components, ensuring that the hardware and software evolve cohesively and are aligned with the evolving needs of rice farming practices [38–40]. Rapid prototyping within the Agile framework enables quick experimentation and validation of concepts, fostering innovation and enabling the researchers to address challenges that may arise during the development process swiftly.

G. Procedures for the Different Phases

Figure 3 illustrates the architectural design, offering a comprehensive depiction of the hardware components and their respective functions in the system. The hardware setup and integration process involves the strategic assembly of core components, with the RS485 Integrated Sensor Soil NPK Sensor assuming a critical role in real-time nutrient monitoring. This sensor is directly connected to the soil, enabling continuous monitoring of essential elements, including nitrogen, phosphorus, and potassium. The integration of the Modbus module (MAX 485) with the Arduino UNO R3 microcontroller establishes a seamless communication channel. The Arduino UNO R3, serving as the central processing unit, is programmed through the Arduino Software IDE 2.0, adhering closely to user guidelines for optimal functionality.

On the software side, a dual approach is implemented. The Arduino IDE 2.0 facilitates code development, enabling the microcontroller to read, process, and transmit data from soil sensors. Concurrently, Microsoft Visual Studio 2019 is used to create a user-friendly interface that displays the real-time NPK levels. It fosters a seamless integration of software components and ensures an intuitive user experience.

Data utilization and model training are integral components of the system, particularly concerning soil fertility and crop recommendation. The dataset undergoes stringent preprocessing, where key parameters are filtered to align with the model's requirements. This curated dataset is foundational for training the machine learning model, enabling it to provide precise recommendations based on soil nutrient content.

During real-time operations, the Arduino processes and transmits soil NPK readings to a computer via the Modbus module. These values are presented within the Visual Studio interface, where Google Colab is employed to host the machine learning model, specifically the KNN algorithm. This cloud-based platform enhances the model's operational efficiency, enabling robust predictions for soil fertility that are crucial to the system's agricultural utility.

Calibration, optimization, and iterative enhancement are central to maintaining the system's accuracy and performance. The RS485 Integrated Sensor is calibrated regularly, ensuring ongoing precision in nutrient readings. Iterative improvements to the software interface and the KNN model are guided by real-world data from rice farming scenarios, establishing a dynamic feedback loop that continuously refines the system's predictive capabilities.

Then, monitoring and maintenance are essential for the long-term effectiveness of the system. Comprehensive checks are conducted to guarantee optimal hardware performance and accurate software readings. Identified issues are swiftly addressed, contributing to a cycle of continuous improvement that supports the reliability and longevity of this agriculture-focused soil monitoring and fertility prediction system.

H. Evaluation

The assessment of the machine learning-based soil nutrient recommendation system involves a meticulous testing process using Google Colab as the execution platform for training data. Google Colab is a cloud-based platform renowned for its scalability and collaborative features. This powerful platform facilitates the parallel execution of tasks by optimizing computational resources and expediting the evaluation process across a variety of diverse machine learning algorithms [41–43]. Leveraging the parallel execution capability significantly reduces testing time while maximizing computational efficiency. The collaborative nature of Google Colab enhances real-time interaction, enabling shared access to the testing framework, datasets, and results. Furthermore, the platform's scalability plays a pivotal role in accommodating the diverse computational requirements of testing multiple machine learning algorithms. In summary, Google Colab not only streamlines the testing process but also fosters collaboration and scalability, providing the researchers with comprehensive insights into the performance of each machine learning model in predicting soil nutrient levels.

The selected machine learning algorithms, Naive Bayes, SVM, KNN, Linear Regression, and Logistic Regression, undergo a comprehensive evaluation to gauge their efficacy in predicting soil nutrient levels. The evaluation results obtained from this systematic testing will provide valuable insights into the performance of each machine learning algorithm. These insights include an understanding of their predictive capabilities, accuracy, and potential limitations when applied to the task of soil fertility prediction within the agricultural context. The assessment process allows for a nuanced comparison of the algorithms, enabling the identification of their respective strengths and weaknesses.

The empirical evaluation outcomes serve as a foundational basis for making informed decisions regarding the selection of the most suitable machine learning algorithm for integration into the operational soil monitoring system. By considering the performance metrics and characteristics revealed during the testing phase, the researchers can make informed choices that align with the specific requirements and objectives of the soil nutrient recommendation system. This iterative and data-driven evaluation process ensures that the chosen algorithm aligns optimally with the demands of accurate and reliable soil fertility prediction, contributing to the effectiveness of the overall agricultural monitoring and management system.

III. Results and Discussion

A. Evaluation of the Performance of the Machine Learning Algorithm Used

The machine learning implementation in the research follows a structured pipeline to ensure accurate and reliable predictions of soil fertility levels. The process begins with dataset preprocessing, where a dataset of 2,200 soil samples is cleaned and normalized to remove inconsistencies. The dataset consists of key features, including Nitrogen (N), Phosphorus (P), Potassium (K), and Fertility Level classification. To ensure a balanced evaluation, the dataset is split into 70% for training, 15% for validation, and 15% for testing. Following preprocessing, feature selection and standardization are applied. Standardization is crucial for improving model convergence and preventing bias toward high-magnitude values, ensuring a fair comparison between features. Additionally, the fertility level is encoded as a categorical variable to facilitate classification tasks.

The model training phase involves implementing and comparing five machine learning algorithms: KNN, SVM, Naive Bayes, Logistic Regression, and Linear Regression. These models are trained using Google Colab, which provides computational resources for

TABLE I
MACHINE LEARNING ALGORITHM ACCURACY RESULTS.

Machine Learning Algorithm	Accuracy (%)
K-Nearest Neighbors (KNN)	98.00
Naive bayes	97.17
Support Vector Machine (SVM)	97.00
Logistic Regression	91.83
Linear Regression	60.53

efficient execution. The scikit-learn library is also utilized for model implementation and evaluation.

The researchers delve into a comprehensive examination of the research findings pertaining to the predictive performance of various machine learning algorithms in the context of soil fertility prediction within agricultural systems. A detailed analysis is presented, focusing on the accuracy metrics derived from the experimental deployment of the Naive Bayes, SVM, KNN, Linear Regression, and Logistic Regression algorithms. The findings are encapsulated in Table I. It provides a quantitative overview of each algorithm's efficacy in predicting soil fertility based on the assessment of a meticulously curated dataset. Following the presentation of the accuracy of the algorithms, a nuanced discussion ensues, elucidating the distinctive strengths and limitations of each algorithm, culminating in informed insights that contribute to the broader discourse on machine learning applications in research.

The empirical assessment of machine learning algorithms for soil fertility prediction, as presented in Table I, identifies the KNN algorithm as the most effective, achieving a remarkable accuracy rate of 98.00%. The research highlights the substantial predictive capacity of KNN, positioning it as a premier algorithm for precision agriculture applications. The observed high accuracy of KNN underscores its aptitude for discerning intricate patterns and complex interdependencies within the training dataset, contributing to its superior predictive performance. This effectiveness is attributed to KNN's utilization of localized data proximity. It is enabled to capture subtle variations within the dataset, thereby enhancing its capacity for precise soil fertility predictions.

In the comparative analysis, Naive Bayes and SVM also exhibit robust predictive capabilities, with accuracies of 97.17% and 97.00%, respectively. However, the marginal difference in accuracy relative to KNN underscores the latter's superior ability to capture the nuanced relationships inherent in soil fertility dynamics. In contrast, Linear Regression demonstrates a significantly lower accuracy of 60.53%, indicating limitations in its precision. It also suggests that its linear assumptions are inadequate for modeling soil fertility within this agricultural context. Meanwhile, Logistic Regression achieves a commendable accuracy of 91.83%, suggesting it as a viable alternative that balances predictive efficacy with interpretability in soil monitoring and fertility prediction.

For model parameter selection, each algorithm is fine-tuned to achieve optimal performance. KNN, which emerges as the best-performing model with an accuracy of 98%, is optimized with k = 5 (determined through cross-validation), Euclidean distance metric, and uniform weighting. SVM, achieving 97% accuracy, uses a Radial Basis Function (RBF) kernel with a regularization parameter (C) of 1.0. The Naive Bayes model, which achieves 97.17% accuracy, relies on probability estimation based on Gaussian distributions. Logistic Regression, which reaches 91.83% accuracy, uses the lbfgs solver, while Linear Regression, with the lowest accuracy at 60.53% employs the ordinary least squares method.

To determine the most suitable model, each is assessed using accuracy as the primary evaluation metric, which provides a reliable measure of overall model performance in this balanced dataset. The KNN model outperforms the others, achieving the highest accuracy of 98%, making it the optimal choice for soil fertility classification. The high accuracy of KNN confirms its effectiveness in capturing patterns in soil nutrient data, enabling accurate predictions for precision agriculture applications. This empirical evidence establishes the KNN algorithm as the optimal choice for soil fertility prediction in agriculture-specific applications. These findings contribute valuable insights for the research community and practitioners, informing algorithm selection and optimization within the advancing field of precision agriculture.

B. Recommendation for the Appropriate Fertilizer for the Soil

Upon the deployment of the IoT devices, the research initiative proceeds to deploy the soil NPK sensor across diverse soil classifications. To facilitate a comprehensive analysis, the researchers systematically collect two distinct soil samples from non-cultivated areas, while concurrently obtaining a specific soil sample enriched with organic fertilizers such as sawdust, chicken manure, and biodegradable fruit and vegetable waste. The subsequent soil fertility testing, detailed in Table II, encapsulates the results obtained from these IoT device-enabled assessments. The first two soil samples, taken from uncultivated areas, reveal relatively low nitrogen values, which lead the system to classify them as deficient and to recommend

TABLE II
SOIL FERTILITY TESTING USING INTERNET OF THINGS (IOT) DEVICE.

Test No.	Nitrogen	Phosphorus	Potassium
Test 1	41	36	42
Test 2	53	44	45
Test 3	68	57	41

the application of additional nitrogen fertilizer. This outcome demonstrates the system's ability to identify nutrient shortages in soils that have not received prior enrichment. In contrast, the third soil sample, which has been treated with organic fertilizers such as sawdust, chicken manure, and biodegradable waste, registers higher levels of nitrogen, phosphorus, and potassium. The model accordingly classifies this sample as having optimal fertility, confirming its capacity to recognize adequately balanced nutrient conditions. These results highlight how the system can both detect deficiencies and validate improvements resulting from organic amendments, thereby underscoring its practical utility for real-time decision-making in fertilizer management. Tests 1 and 2 involve the random acquisition of soil samples from uncultivated areas, while Test 3 represents a soil sample enriched with organic fertilizers. Subsequently, the recorded nitrogen, phosphorus, and potassium readings from each test are presented in real-time through a user-friendly C#-based Windows Form application.

Following the generation of the KNN model, the researchers incorporate this model into a web-based system designed for soil fertility prediction and optimized fertilizer recommendations. This integration marks a crucial step in operationalizing the KNN algorithm's predictive capabilities within an accessible and user-friendly platform. By leveraging the model's real-time assessment capabilities, the web-based system empowers end-users to make data-driven decisions regarding soil management practices. This approach exemplifies the convergence of advanced machine learning techniques with user-centered web interfaces, aligning with contemporary precision agriculture practices and facilitating the transition from research to real-world application in soil monitoring and fertility management.

The developed application establishes a direct connection with the Arduino board via a serial communication (COM) port, providing users with a streamlined and intuitive interface for monitoring and interpreting soil nutrient levels. The integration of a Windows Form application significantly enhances the user experience in soil monitoring and fertility optimization, offering immediate access to the recorded NPK readings in an accessible and efficient manner. Through rigorous testing, soil NPK data readings are captured and presented in a purpose-built Windows-based application. This application communicates with the Arduino board in real time via a serial COM port, enabling seamless data retrieval and integration. The extracted NPK values are subsequently embedded into a web-based platform. It integrates a sophisticated machine learning model that performs essential predictive analytics on soil fertility based on the recorded nutrient content.

Beyond soil fertility prediction, the machine learning model provides actionable fertilizer recommendations, calculating specific nutrient requirements to address the detected NPK deficiencies. This guidance, displayed as a percentage breakdown, offers precise recommendations to end-users on the appropriate nutrient proportions necessary for effective soil enrichment. This integration exemplifies the seamless fusion of hardware and software components, leveraging the strengths of the Windows-based application, Arduino board, and web-based system. This user-centric design for soil monitoring and fertility optimization represents a substantial advancement in the practical deployment of machine learning within precision agriculture.

As shown in Fig. 4, the web-based system, incorporating the machine learning model, analyzes soil NPK readings from three soil samples. The system plays a pivotal role in forecasting soil fertility and provides precise recommendations for optimal fertilizer application by fostering informed decision-making and sustainable soil management practices. Figure 4 presents the soil testing process using a Windows-based application that connects directly with the Arduino board. The interface displays the real-time NPK readings transmitted from the IoT-enabled sensor, allowing users to monitor soil nutrient levels immediately after data collection. This visualization demonstrates the system's ability to capture nutrient concentrations and present them in a user-friendly format that farmers can easily interpret. By showing the recorded values on screen, the application bridges the gap between raw sensor output and actionable insights, ensuring that nutrient deficiencies or balanced conditions are readily identifiable. Therefore, it highlights the practical usability of the developed platform, underscoring how the integration of IoT devices and machine learning can simplify soil testing and support evidence-based fertilizer management in actual farming contexts.

Figure 5 presents soil fertility prediction and fertilizer recommendations based on NPK values. For soil sample 1 (N: 41, P: 36, K: 42), the system's machine learning model predicts "Needs Additional Nitrogen Fertilizer," recommending a 61% nitrogen supplement. This result indicates that while phosphorus

Cite this article as: R. M. Sumacot and J. R. Paragas, "IoT-Enabled K-Nearest Neighbors (KNN)-Based Soil Nutrient Recommendation System for Rice (*Oryza Sativa L.*) Cultivation", CommIT Journal 19(2), 151–165, 2025.

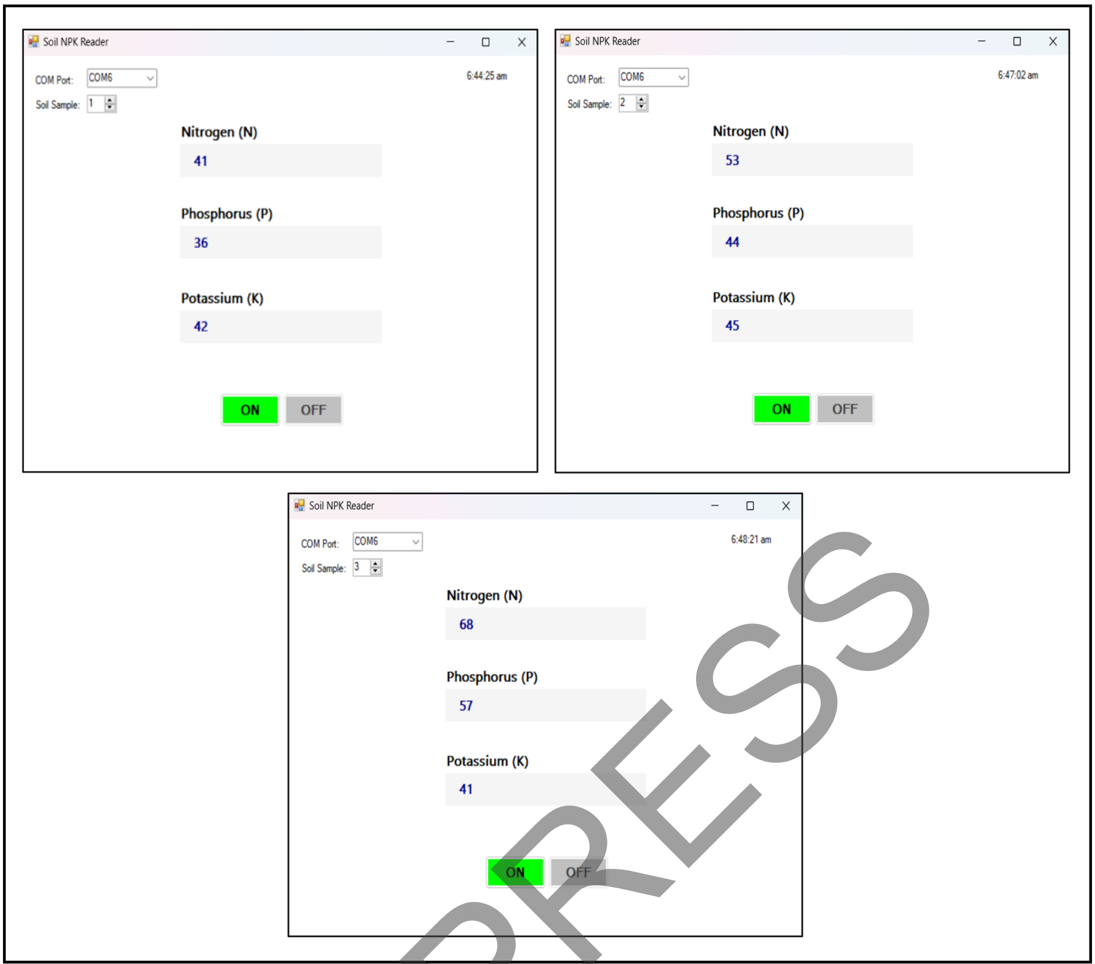

Fig. 4. Soil testing using a Windows-based application.

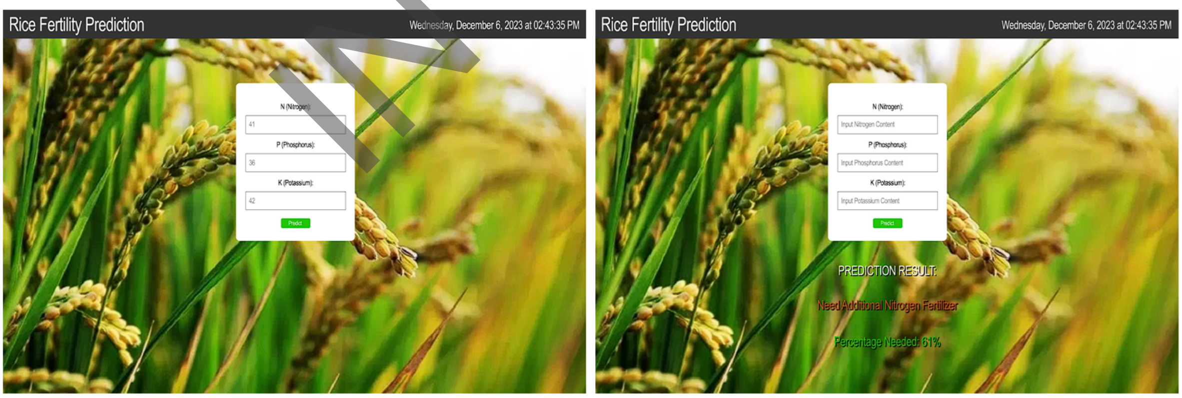

Fig. 5. Prediction of Soil Sample 1.

and potassium levels are within acceptable ranges, nitrogen is significantly below the optimal threshold required for rice cultivation. The recommendation aligns with agronomic practices that prioritize nitrogen management, given its crucial role in promoting tillering and grain yield in rice plants. By quantifying the required supplement, the system not only detects nutrient deficiency but also translates it into

Cite this article as: R. M. Sumacot and J. R. Paragas, "IoT-Enabled K-Nearest Neighbors (KNN)-Based Soil Nutrient Recommendation System for Rice (*Oryza Sativa L.*) Cultivation", CommIT Journal 19(2), 151–165, 2025.

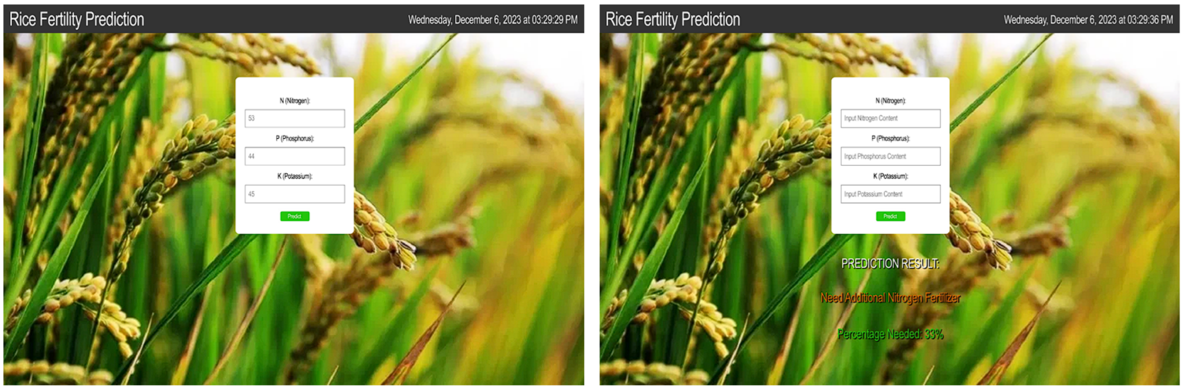

Fig. 6. Prediction of Soil Sample 2.

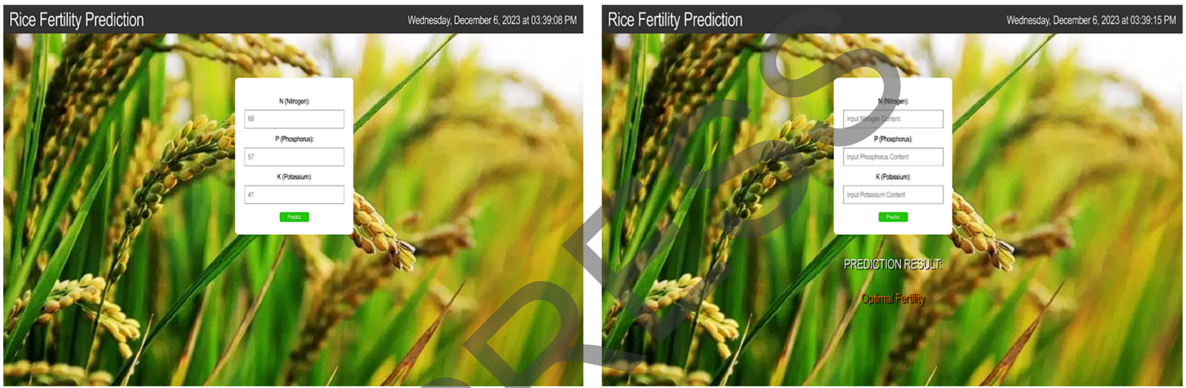

Fig. 7. Prediction of Soil Sample 3.

an actionable input for farmers, ensuring efficient and targeted fertilizer application. The result demonstrates the practical utility of the IoT and ML integration in reducing both under- and over-application of fertilizers, which supports higher productivity while minimizing environmental impact.

Figure 6 also shows soil fertility prediction and fertilizer recommendations based on NPK values. For soil sample 2 (N: 53, P: 44, K: 45), the system's machine learning model predicts "Needs Additional Nitrogen Fertilizer," recommending a 33% nitrogen supplement. This outcome demonstrates that even when phosphorus and potassium levels are sufficient, the model can isolate nitrogen as the limiting factor for fertility. Compared with soil sample 1, the nitrogen deficiency here is less severe, as reflected in the smaller recommended supplement. It highlights the system's ability to calibrate fertilizer recommendations according to the degree of deficiency, rather than offering a generic prescription. In practical terms, such tailored guidance allows farmers to optimize input costs by applying only the required nutrient in the necessary amount, supporting both productivity and sustainable fertilizer management.

Figure 7 presents soil fertility prediction and fertilizer recommendations based on NPK values. For Soil Sample 3 (N: 68, P: 57, K: 41), the system's machine learning model predicts "Optimal Fertility". It indicates that the nutrient levels fall within the desired range for rice cultivation. Unlike the previous two samples, no additional fertilizer input is recommended, confirming that the soil is already balanced and suitable for planting. This result validates the system's ability not only to detect nutrient deficiencies but also to confirm adequacy when fertility conditions are sufficient. Such outcomes are crucial for sustainable farming, as they help to prevent unnecessary fertilizer use, reduce production costs, and mitigate the environmental impacts of over-application. By accurately identifying both deficient and optimal soils, the system demonstrates its potential as a reliable decision-support tool for efficient nutrient management in rice production.

C. Comparison with Related Studies

To evaluate the effectiveness of the IoT-enabled KNN-based soil nutrient recommendation system, the researchers conduct a comparative analysis with previous studies on machine learning-based soil fertility prediction. Several studies have explored machine learning techniques for soil analysis. However, they exhibit limitations in real-time adaptability, computational efficiency, and field deployment, which the research addresses.

Previous research has applied machine learning techniques to assess soil nutrients in *Dacrydium Pectinatum* forests, demonstrating high accuracy in nutrient evaluation [39]. However, the approach lacks real-time IoT integration, which limits its practical use in precision agriculture. In contrast, this research incorporates IoT sensors for continuous and real-time data acquisition, allowing for dynamic soil nutrient monitoring and immediate decision-making.

Similarly, another previous research has optimized Extreme Learning Machine (ELM) parameters to improve soil classification accuracy [40]. While the approach enhances classification performance, it relies on static datasets, making it less adaptable for real-world agricultural applications. In comparison, this research utilizes a live data-driven approach that integrates real-time soil nutrient readings from IoT sensors, ensuring practical field deployment and adaptability to changing soil conditions.

In another research, ensemble deep learning techniques have been introduced for soil nutrient and pH classification [41]. While the approach achieves high accuracy, deep learning models typically demand significant computational resources, making them less feasible for small-scale farming applications. In contrast, this research employs KNN, a computationally efficient model. It maintains high accuracy (98%) while remaining lightweight and adaptable for real-time implementation.

Similarly, deep learning models have been developed for predicting soil nutrients in cabbage cultivation [42]. While the system delivers high classification accuracy, it depends on large and labeled datasets and complex computational infrastructure. It may not be suitable for resource-constrained agricultural environments. The approach in this research, in contrast, achieves a balance between accuracy and computational efficiency, making it a cost-effective and scalable solution for real-world precision agriculture.

This comparative analysis highlights the significance of this research in bridging the gap between high-accuracy machine learning models and real-time field applicability. By integrating IoT technology with an optimized KNN model, the system ensures both real-time monitoring and practical usability, making it a valuable innovation for precision agriculture and sustainable farming practices. Beyond demonstrating high predictive accuracy, the system’s ability to generate timely and actionable fertilizer recommendations directly addresses a common limitation in previous studies, where models are validated only in simulation or offline environments. The seamless flow of data from soil sensors to the machine learning framework and user-friendly recommendations shows how advanced algorithms can be translated into tools that farmers can adopt in day-to-day operations. In this way, the research not only contributes to academic discussions on agricultural data science but also provides a concrete and scalable solution that can improve crop productivity, reduce unnecessary input costs, and promote environmentally responsible nutrient management in rice cultivation.

IV. Conclusion

The research successfully develops an IoT-enabled KNN-based soil nutrient recommendation system for rice cultivation, integrating real-time soil monitoring with machine learning for precision agriculture. The system, leveraging an RS485 Integrated Soil NPK Sensor and an Arduino-based IoT framework, provides continuous soil nutrient analysis and automated fertilizer recommendations, significantly enhancing soil fertility management. Among the tested machine learning models, KNN demonstrates the highest accuracy (98%), proving its effectiveness in predictive soil fertility assessment. The research highlights the transformative potential of IoT and machine learning in optimizing fertilizer use, improving crop yields, and promoting sustainable farming practices.

Beyond algorithm evaluation, the research includes rigorous testing of IoT devices, wherein soil samples are systematically collected for real-time NPK analysis through a C#-based Windows Form application. It is followed by the integration of the KNN model within a web-based system, reflecting modern precision agriculture trends and showcasing a sophisticated, user-centric approach to soil monitoring and fertility optimization. The KNN model not only accurately predicts soil fertility but also provides targeted recommendations with detailed percentage breakdowns, representing a substantial advancement in the application of machine learning within precision agriculture.

While the sensor’s sub-second response time enables near-instantaneous nutrient readings, the research does not include the system’s wide performance metrics, such as latency, throughput, or power efficiency. Future research should explore these parameters to further validate and optimize the real-time capabilities and scalability of the system in diverse field environments. Additionally, future research should focus on expanding the system's adaptability to various crops and environmental conditions, integrating additional soil health indicators and enhancing scalability for broader agricultural applications. This innovation represents a significant step towards data-driven and sustainable agricultural solutions that support global food security and environmental conservation.

Author Contribution

Conceived and designed the analysis, R. M. S.; Collected the data, R. M. S.; Contributed data or analysis tools, R. M. S.; Performed the analysis, R. M. S.; Wrote the entire paper, R. M. S.; and Wrote and revised some parts of the paper, J. R. P.

Data Availability

The data that support the findings of the research are openly available in Zenodo at https://doi.org/10.5281/zenodo.15307297.

References

[1] J. MacPherson, A. Voglhuber-Slavinsky, M. Olbrisch, P. Schöbel, E. Dönitz, I. Mouratiadou, and K. Helming, "Future agricultural systems and the role of digitalization for achieving sustainability goals. A review," *Agronomy for Sustainable Development*, vol. 42, no. 4, pp. 1–18, 2022.

[2] N. Khan, R. L. Ray, G. R. Sargani, M. Ihtisham, M. Khayyam, and S. Ismail, "Current progress and future prospects of agriculture technology: Gateway to sustainable agriculture," *Sustainability*, vol. 13, no. 9, pp. 1–31, 2021.

[3] E. Taer and A. Taer, "Cultivating tomorrow: A comprehensive review of agricultural innovations in the Philippines (2018–2023)," *Agricultural Research*, vol. 13, no. 4, pp. 623–635, 2024.

[4] P. Mthiyane, M. Aycan, and T. Mitsui, "Strategic advancements in rice cultivation: Combating heat stress through genetic innovation and sustainable practices—A review," *Stresses*, vol. 4, no. 3, pp. 452–480, 2024.

[5] M. F. Stuecker, M. Tigchelaar, and M. B. Kantar, "Climate variability impacts on rice production in the philippines," *PLOS ONE*, vol. 13, no. 8, pp. 1–17, 2018.

[6] Ricelytics, "Rice industry in the Philippines." [Online]. Available: https://ricelytics.philrice.gov.ph/rice_industry

[7] R. J. Mamiit, J. Yanagida, and T. Miura, "Productivity hot spots and cold spots: Setting geographic priorities for achieving food production targets," *Frontiers in Sustainable Food Systems*, vol. 5, pp. 1–15, 2021.

[8] R. J. C. Rollon, J. M. Golis, and E. Salas, "Impacts of soil nutrient management practices on soil fertility, nutrient uptake, rice (Oryza Sativa L.) productivity, and profitability," *Journal of Applied Biology and Biotechnology*, vol. 9, no. 2, pp. 75–84, 2021.

[9] M. Becker, R. Clavero, Z. N. Maung, O. M. Khin, S. Kong, P. Men, M. J. C. Regalado, S. Ro, K. K. Win, and S. Pariyar, "Pathways and determinants of changing nutrient management in lowland rice-based systems of Southeast Asia," *Agronomy for Sustainable Development*, vol. 43, no. 6, pp. 1–17, 2023.

[10] S. Jain, D. Sethia, and K. C. Tiwari, "A critical systematic review on spectral-based soil nutrient prediction using machine learning," *Environmental Monitoring and Assessment*, vol. 196, no. 8, 2024.

[11] M. Guo, "Soil health assessment and management: Recent development in science and practices," *Soil Systems*, vol. 5, no. 4, pp. 1–20, 2021.

[12] J. A. Reijneveld, M. J. van Oostrum, K. M. Brolsma, D. Fletcher, and O. Oenema, "Empower innovations in routine soil testing," *Agronomy*, vol. 12, no. 1, pp. 1–20, 2022.

[13] R. S. Lalitha, R. Srinivas, C. V. Raghavendran, K. Kavitha, P. S. V. V. S. R. Kumar, and P. S. L. Sravanthi, "Real time Nitrogen, Phosphorus, Potassium (NPK) detection in soil using IoT," in *International Conference on Emerging Applications of Information Technology*. West Bengal, India: Springer, November 2021, pp. 408–416.

[14] M. A. Mohammed and S. Sanjay, "IoT based soil NPK monitor via asynchronousweb-server using NodeMCU," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 10, no. VII, pp. 525–530, 2022.

[15] M. K. Senapaty, A. Ray, and N. Padhy, "IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture," *Computers*, vol. 12, no. 3, pp. 1–34, 2023.

[16] C. Wu, Y. Chen, X. Hong, Z. Liu, and C. Peng, "Evaluating soil nutrients of Dacrydium Pectinatum in China using machine learning techniques," *Forest Ecosystems*, vol. 7, pp. 1–14, 2020.

[17] M. S. Suchithra and M. L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine pa-

rameters," *Information Processing in Agriculture*, vol. 7, no. 1, pp. 72–82, 2020.

[18] J. Escorcia-Gutierrez, M. Gamarra, R. Soto-Diaz, M. Pérez, N. Madera, and R. F. Mansour, "Intelligent agricultural modelling of soil nutrients and pH classification using ensemble deep learning techniques," *Agriculture*, vol. 12, no. 7, pp. 1–16, 2022.

[19] H. Sajindra, T. Abekoon, J. A. D. C. A. Jayakody, and U. Rathnayake, "A novel deep learning model to predict the soil nutrient levels (N, P, and K) in cabbage cultivation," *Smart Agricultural Technology*, vol. 7, pp. 1–10, 2024.

[20] I. Wickramasinghe and H. Kalutarage, "Naive Bayes: Applications, variations and vulnerabilities: A review of literature with code snippets for implementation," *Soft Computing*, vol. 25, no. 3, pp. 2277–2293, 2021.

[21] Q. Liu, W. Zhu, Z. Dai, and Z. Ma, "Multi-task Support Vector Machine Classifier with Generalized Huber Loss," *Journal of Classification*, vol. 42, no. 1, pp. 221–252, 2025.

[22] M. S. Ahmed, M. N'diaye, M. K. Attouch, and S. Dabo-Niange, "K-Nearest Neighbors prediction and classification for spatial data," *Journal of Spatial Econometrics*, vol. 4, no. 1, 2023.

[23] A. F. Castro Torres and A. Akbaritabar, "The use of linear models in quantitative research," *Quantitative Science Studies*, vol. 5, no. 2, pp. 426–446, 2024.

[24] M. Seyedan and F. Mafakheri, "Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities," *Journal of Big Data*, vol. 7, no. 1, pp. 1–22, 2020.

[25] V. Jayaraman, S. Parthasarathy, A. R. Lakshminarayanan, and S. Sridevi, "Crop recommendation by analysing the soil nutrients using machine learning techniques: A study," in *International Conference on Computational Intelligence in Data Science*. Chennai, India: Springer, March 18–20, 2021, pp. 15–26.

[26] United Nations, "The 17 goals — Sustainable development." [Online]. Available: https://sdgs.un.org/goals

[27] V. Varbanova and P. Beutels, "Recent quantitative research on determinants of health in high income countries: A scoping review," *PLOS ONE*, vol. 15, no. 9, pp. 1–21, 2020.

[28] F. Asadi, "Introduction to Arduino boards," in *Essentials of Arduino™ boards programming. Maker innovations series*. Apress, 2023.

[29] B. Johnson, *Essential Visual Studio 2019: Boosting development productivity with Containers, Git, and Azure tools*. Apress, 2020.

[30] D. Strauss, *Getting started with Visual Studio 2019: Learning and implementing new features*. Apress, 2019.

[31] S. R. Sukhdeve and S. S. Sukhdeve, "Google Colaboratory," in *Google Cloud platform for data science: A crash course on big data, machine learning, and data analytics services*. Springer, 2023, pp. 11–34.

[32] iMicon System, "RS-NPK-N01-TR soil nitrogen, phosphorus,potassium three-in-one fertility sensor (Type 485) user's guide," 2021. [Online]. Available: http://bit.ly/4mVnYdS

[33] S. Ameer, H. Ibrahim, F. N. U. Kulsoom, G. Ameer, and M. Sher, "Real-time detection and measurements of nitrogen, phosphorous & potassium from soil samples: A comprehensive review," *Journal of Soils and Sediments*, vol. 24, no. 7, pp. 2565–2583, 2024.

[34] S. Vennela and A. Velumani, "Design and fabrication of sensor based soil test analyzer to test the quality of sandy soil in comparison with experimental analysis," in *AIP Conference Proceedings*, vol. 3193, no. 1. AIP Publishing, 2024.

[35] A. N. Sharkawy, M. Hasanin, M. Sharf, M. Mohamed, and A. Elsheikh, "Development of smart home applications based on Arduino and android platforms: An experimental work," *Automation*, vol. 3, no. 4, pp. 579–595, 2022.

[36] A. Kumar, "Crop growth recommendations: Optimal conditions for higher yields," 2023. [Online]. Available: https://www.kaggle.com/datasets/aksahaha/crop-recommendation/data

[37] P. d. O. Santos and M. M. de Carvalho, "Exploring the challenges and benefits for scaling agile project management to large projects: A review," *Requirements Engineering*, vol. 27, no. 1, pp. 117–134, 2022.

[38] H. Dong, N. Dacre, and A. Bailey, "Sustainability in complex agriculture projects: A study of agile agricultural co-operative institutions," 2021. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3879454

[39] G. Mgendi, "Unlocking the potential of precision agriculture for sustainable farming," *Discover Agriculture*, vol. 2, pp. 1–24, 2024.

[40] World Economic Forum, "Agritech: Shaping agriculture in emerging economies, today and tomorrow," 2024. [Online]. Available: https://www3.weforum.org/docs/WEF_Agritech_2024.pdf

[41] M. Luppichini, M. Bini, and R. Giannecchini, "CleverRiver: An open source and free Google

Colab toolkit for deep-learning river-flow models," *Earth Science Informatics*, vol. 16, pp. 1119–1130, 2023.

[42] M. J. Nelson and A. K. Hoover, "Notes on using Google Colaboratory in AI education," in *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, Trondheim, Norway, June 15–19, 2020, pp. 533–534.

[43] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: Diagnosis to detection," *Multimedia Tools and Applications*, vol. 82, pp. 26 731–26 769, 2023.