Enhancing Sustainable Agriculture through Machine Learning-Based Crop Recommendation and NPK Sensor Integration

Moghazy Refaat[1*], Moaz Mohamed[2],
Ahmed Elsayed Ali Abubakr[3], Fawzy Ibrahim Eissa[4],
Omar mohamed Eid kamel[5], Elsaeed Mohamed[6],
Omar alaa kamel[7], Mohamed Omar[8], Zainab H.Ali[9†],
Third Author[1,2†]

[1*] Faculty of Artificial Intelligence, Kafrelsheikh University, El-Gish St, Kafrelsheikh, 113355, Kafrelsheikh, Egypt.
[2-8] Faculty of Artificial Intelligence, Kafrelsheikh University, El-Gish St, Kafrelsheikh, 113355, Kafrelsheikh, Egypt.
[9]Department of Embedded Network Systems and Technology, Faculty of Artificial Intelligence, Kafrelsheikh University, El-Gish St, Kafrelsheikh, 113355, Kafrelsheikh, Egypt.

*Contributingauthors: mogazeymashaly@gmail.com ;
moazeldsoky8@gmail.com ; aboelkhaira054@gmail.com ;
fawzyeissa18@gmail.com ; tthkamel0@gmail.com ;
alsydalrhmany7@gmail.com; 3amo10ry@gmail.com ;
mhmdmrw837@gmail.com ; zainabhassan@ai.kfs.edu.eg ;
iiiauthor@gmail.com ; ahmedelsayedabubakr@gmail.com ;

†These authors contributed equally to this work.

Abstract

Artificial intelligence (AI) is a strong method for evaluating agricultural data, particularly in understanding the links between crop output, soil nutrient levels, and climate variability. AI algorithms can use such data to make informed suggestions about optimal crop selection and applying supplemental fertilizers to increase agricultural yield. The main objective of this study is to introduce a strong crop recommendation system based on Machine Learning (ML) approach to help farmers in resource-limited areas in avoiding abuse of fertilizers and pesticides by proposing crops that correspond to natural nutrient availability, thereby reducing environmental degradation and soil depletion. The dataset covers crucial agricultural metrics such as nitrogen (N), phosphorus (P), and potassium (K) content, soil pH, and three critical climate variables: temperature, precipitation, and humidity. The models examined in this study are as follows: The Support Vector Classifier (SVC), Light Gradient Boosting Machine (LGBM) Classifier, Random Forest (FR), Logistic Regression, and Decision Tree Classifier were trained using yield data from 22 crops. The models had respective accuracy scores of 0.8750, 0.9682, 0.9159, 0.9727, and 0.9159. These findings suggest that ML-driven systems can improve customer decision-making by accurately identifying product quality. Field trials validated the system's practical usefulness by comparing model-based recommendations to traditional agricultural approaches. The findings underscore the system's capacity to increase crop productivity, improve resource efficiency, and support sustainable farming practices. Furthermore, this research contributes to water footprint analysis by demonstrating how ML techniques can advance soil nutrient optimization, inform fertilizer application decisions, and improve crop quality assessment, thereby promoting food security and environmental sustainability.

Keywords: Crop recommendation, Water Footprint, NPK Sensor, Machine Learning, Soil nutrients, Agriculture

1 Introduction

The country's economy heavily depends on farming, the sector helps ensure that all Egyptians have enough to eat, and increases economic growth. But today, agriculture problems include less land available for farming, insufficient water supply, and damage caused by climate change [1]. People worldwide still struggle with food shortages, as pests, diseases, and changing weather make the problem more acute. AI and ML are increasingly recognized as practical solutions to improve agriculture operations and reduce losses [2][3]. Soil analyses and crop advice in traditional farming tend to be laborious and sometimes inaccurate. They allow for more efficient and well-thought-out steps when allocating resources. Having fertile soil is very important for the cultivation of good yields of crops [4] [5]. The soil needs to remain productive so that there is a proper combination of organic and inorganic nutrients.

Using AI technologies, farmers can better manage soil nutrients, and thus promote sustainability [6]. Agronomists often grow crops where productivity is reduced. With AI added to precision agriculture, the most suitable types of crop can be selected based on environmental conditions. Forecasting soil moisture levels, crop yields, and possible disease outbreaks is done with Support Vector Machines (SVM) [7], Decision Tree Classifier (DTC) [8], Random Forest (RF) [9], and Light gradient Boost Machine Classifier(LGBM)[10]. They are designed to improve the way decisions are made and the way risks are managed in the agricultural domain. Notable applications of AI include Agro Consultant, which provides crop recommendations tailored to specific soil and climatic conditions, and ML models integrated with the Internet of Things (IoT) that deliver real-time strategic guidance for agricultural practices [11].

Despite the advantages offered by these technologies, challenges persist in the refinement of ML models to enhance their accuracy and operational efficiency. Diligent data collection, comprehensive training, and effective implementation are critical to fully maximizing the potential of AI in the agricultural sector [12]-[13]. Artificial Intelligence (AI) must be used in Egyptian farming to cope with the increasing demand for food, protect the environment, and support economic success [36,39]. Developing AI-centered farming methods will help maintain the future of the country's farms. To produce accurate forecasts and advice considering what each crop needs in nutrients, its current stage of growth, and the current climate. It also allows users to see the most valuable soil elements for plant growth and calculates how much water the farm uses [14].

This research effort could completely change the way farmers care for the soil and make plans to grow crops. With its ability to give accurate, up-to-date information on soil and advice on the most suitable crops, the Crop Filtration Machine for each soil powered by ML works to help farmers become more productive, save resources, ensure food security, and uphold sustainable agriculture practices. Contributions to this work include building and creating a crop recommendation device with integrated machine learning to constantly monitor soil moisture, humidity, temperature, and NPK levels using a five-container NPK sensor. It allows for up-to-date data, helping farmers make the right and timely choices. Look over the data collected with ML methods and tell farmers what crops to plant from the results. Doing this can help to prosper crops, produce a larger harvest, and make agriculture more economical. Test the device in practical situations and evaluate how it compares to the previous methods. The assessment shows that the device is better than traditional methods and can change agriculture.

The structure of this paper is as follows: Section 2 studies significant research on the recommendation corps topic. Section 3 covers how the study is performed and the tools or materials used, like crop taxonomy sensors, data transmission and analysis, machine-learning models, and the suggested framework. Section 4 discusses results, gathering and pre-processing data, evaluation measures, applying the models, and assessing their performances. Section 5 focuses on reviewing the findings of this study and its limitations.

2 Related Works

Soil nutrient management and figuring out the right crops to use are essential for keeping farming healthy and productive. Typically, farmers use old methods that involve self-analysis and a lot of labor to decide which crops to grow and what soil nutrition to use. As a result, they do not achieve the best results or high crop yields. However, new ML systems have helped us find better solutions to these problems. This literature review looks at what research and new ideas are out there that use machine learning to help keep track of soil nutrients and suggest the right crops to grow. Systems for monitoring soils using machine learning are now capturing much interest. These systems have sensors that check for how much moisture is in the soil, how acidic it is, or what nutrients there are, and help gather data about the soil in real time [15]. The data collected goes to an office in the middle, where it's checked and used to make decisions. For example, an IoT-based system with sensors for soil moisture and nutrients can check the condition of the soil as it happens in real time [16]. The collected data was then used to help figure out better ways to water the plants, give them nutrients, and save money on water and fertilizer.

ML techniques have been found to work well when looking at soil data and making guesses about how much nutrients are in the land. Machine learning algorithms can see patterns and connections in big data sets, which helps them make better predictions and help people make good, forward-thinking choices. [17] established an ML model that looked at past data, nearby weather, and what crops were needed to determine how much specific nutrients were likely to be in the soil. The model did a good job of deciding whether plants were low or too high on particular nutrients, making it easier to change how much fertilizer was used when needed. Crop recommendation systems help farmers choose the best crops for their fields based on the soil type and what people are most likely to want to buy. ML algorithms have been used to make crop suggestions that check soil's nutrients, the weather, and what farm activities are popular in the market [18]. [53] suggested a system that uses ML to analyze past harvest records, local weather trends, and nutrients in the soil. Through the system, advisors could recommend the most suitable crops, leading to higher yields and income for farmers. A system with an NPK sensor, an Arduino panel, a Raspberry Pi, an LCD screen, and a battery was introduced. ML methods were used with this proposed method to check how many minerals are in the soil and determine the appropriate recipe for the use of fertilizer.

The recommendation of fertilizers is essential because they help farmers get better harvests and ensure that farming methods don't harm the environment in the long run. Most of the time, recommending fertilizers is done by experts who analyze the data with their experience. It is known to take considerable time and may not always be accurate. Thanks to more advanced ML techniques, there has been a rise in using them to help improve the methods for recommending fertilizers [19]. Represents a way of using ML to help suggest which crops and fertilizers are best to plant in each area. They applied various approaches from ML. Many studies show that people have progressed greatly with devices that help farmers watch soil nutrients and pick which crops to grow. Internet of Things technology and ML help offer real-time soil information and recommend individual solutions for farmers. They offer a great deal of potential to increase the number of crops grown, make better use of resources, and support sustainable ways of farming. Future research should work on making ML systems more accurate and able to work for larger groups, investigate how to keep food safe for everyone once it leaves the farm, and find new data sources to help the technology work well. Below is a simple table that shows the main points of the latest studies in Table 1.

Table 1 Summary of Agricultural Applications Using Data-Driven Techniques

Application	Method	Dataset	Key Findings	Limitation
An automated remote field monitoring system	LoRaWAN	time series data	Visualization in the cloud	Data analysis technique is not mentioned
Soil monitoring system	IoT	time series data	Efficient identification of soil type with real-time display	Data analysis technique is not mentioned
Soil classification based on micronutrients	ELM	Private	Achieved 94% accuracy	Dataset is limited to the Tamil Nadu region
Crop recommendation	MLP	Kaggle Dataset	Accuracy of 98.22%	Lacks detailed analytical evaluation
Crop recommendation platform for farmers	RF	Kaggle Dataset	Random Forest achieved 97.18% accuracy	No platform was developed for user deployment
Ongoing crop and field information support	MSVM-DAG-FFO	Own dataset	Achieved 97.3% accuracy	No interactive platform available for farmers
Correct selection of crop	SCS	Dataset from Pakistan	Accuracy of 97.4%	Dataset covers only two soil types; limited crop applicability

3 Methods and Materials

This section provides a comprehensive overview of the materials and methods utilized in the study, encompassing the sensing hardware in Section 3.1, the architecture of the ML model in Section 3.2, and the proposed system is discussed in Section 3.3. The research integrates various hardware components from international manufacturers to construct a robust and scalable crop recommendation system. The NPK five-pin sensor, imported from China, was selected as the primary soil nutrient detection device. This sensor is capable of accurately measuring the concentrations of nitrogen (N), phosphorus (P), and potassium (K) in the soil as key indicators of soil fertility. Its real-time sensing capability enables high-resolution monitoring of nutrient levels, which is critical for precise crop recommendation. The sensor's compatibility with digital interfaces and its resistance to environmental stressors, such as moisture and salinity, make it suitable for field deployment. The Raspberry Pi 3 Model B was employed for data processing and system control. The UK Electronic Technology Company manufactures this single-board computer with a powerful yet compact platform, 1.2 GHz 64-bit quad-core ARM Cortex-A53 processor, 1 GB RAM, built-in Wi-Fi, and Bluetooth capabilities. It was the central unit for integrating sensor data, running the ML algorithms, and facilitating user interaction via a connected display or remote interface.

3.1 Sensors and Hardware

3.1.1 NPK five-pin sensor

The NPK sensor is significant in this study because it allows for precise evaluation of soil nutrient levels, notably N, P, and K concentrations, the three key macronutrients required for plant growth. This sensor collects and analyzes data by measuring soil factors such as temperature, moisture, salinity, pH, and nutrient content. The NPK sensor, designed specifically for agricultural applications, provides speedy and precise readings, allowing real-time soil health and fertility monitoring [20].

Integrating the NPK sensor into the research framework enhances the ability to accurately monitor soil conditions, thereby supporting intelligent algorithms for nutrient analysis and crop recommendation development. This technology is widely applicable in scientific experiments, water-efficient irrigation systems, greenhouses, horticulture, pastures, rapid soil testing, plant cultivation, wastewater treatment, grain storage, and soil water content and temperature measurement. Its compatibility with other monitoring instruments and its responsiveness to various agronomic practices further affirms its value in advancing precision agriculture and sustainable farming strategies.

3.1.2 Max RS485 TTL module

To connect the Soil NPK Sensor with the Arduino, the MAX485 TTL to RS-485 module is used. Max485 is commonly found in industry since it is great for sending data longer distances or across areas with a lot of electrical noise. It allows up to 2.5MBit/Sec of data to be sent, yet at longer distances, its maximum speed decays. A total of 32 devices can communicate on the same Bus/cable using the RS-485, if it is configured as master and slave. An article describing how to communicate with several controllers using the MAX485 interface module with Arduino has already been written. The term Max RS485-TTL is used for a type of communication standard that allows data to be sent over far distances, mostly in industrial and farm sensor systems. RS485 is a form of bus that allows multiple gadgets to connect to one network. TTL (Transistor-Transistor Logic) is a type of logic to make communication at a lower voltage simpler. For soil sensors and IoT applications, using this standard enables stable and long-distance communication of data to other devices.

3.2 ML Models

3.2.1 Support Vector Classifier(SVC)

The SVC algorithm [21] is used in agriculture to classify crops by looking at things such as NPK fertilizer levels, soil pH, and the climate. The process works by selecting the appropriate ‘super-level’ to make different crop classes easier to tell apart in a high-dimensional space. It performs well when there are nonlinear links in the data, and input features are made higher by kernel functions, resulting in easy separation of the data. To train the decision tree model, the settings were kernel “rbf”, best estimator, c=1.0, gamma = “scale”, random state 42, and probability = True .

3.2.2 Decision Tree Classifier (DTC)

The DTC [22] is a way of training computers to help sort out crops by looking at things like how much fertilizer they get, how acidic or basic their soil is, and the weather they face. The model works by splitting the training data into smaller groups using different tests for each attribute, and then finding out which attribute helps you learn the most at each stage. This creates a tree-shaped outline where each part of the tree stands for a choice made based on a certain trait, and the end parts, or leaf nodes, show the different kinds of crops. Specifically, the tree was trained using settings such as maximum features set to ”auto”, using the best estimator, setting alpha to 0.001, setting the standard method as ”entropy”, using random state 42, and using a maximum depth of 5 .

3.2.3 Light gradient Boost Machine Classifier(LGBM)

The LGBM [23] can quickly and correctly tell what crops are by noting their amounts of fertilizer, type of soil, and levels of rain. Since LGBM adds new splits to its branches thoughtfully, it can form very deep trees quickly even when working with a lot of data. During training the decision tree model, num_leavels = 31, estimators = 100, learning_rate = 0.05, and random state = 42 were included .

3.2.4 Logistic Regression

Logistic regression [24] is a common statistical method used for crop classification that looks at things like how much NPK fertilizer is used, the soil pH, and the weather to figure out which type of crop it is. Rather than trees, logistic regression works under the assumption that there is a straight-line connection between features and the probability of a crop class, which is why it is easy to use for different types of classification problems. The decision tree model was trained using estimators=100 and random state = 42 .

3.2.5 Random Forest (RF)

A combination of multiple decision trees in RF helps improve the way crops are identified and reduces over handling of the data [25]. RF builds several decision trees while training and then uses the average predictions to form a reliable and general model. With this technology, RF can deal with nonlinear combinations and features, which is why it is ideal for agricultural data that reports NPK levels, soil pH, and climate. The values estimators=100 and random state = 42 were set during the training stage of the decision tree model.

3.2.6 Proposed Framework

The framework includes the use of a five-pin sensor system in agriculture to monitor different aspects, like soil nutrient levels (NPK), how moist the soil is, its pH, and the current temperature. The gathered information goes straight to ML algorithms to identify crops that work well on this soil. As this sensor collects real-time data, farmers can observe soil conditions and decide how to irrigate and fertilize their fields, as well as which crops will grow well in the soil. This helps farmers pick which crops they want to grow and often increases the number of crops they can harvest. Additionally, it checks the nutritional value of the crops, amounts of pesticides, and for anything that might make the food harmful by studying information from all the sensors. Because it uses ML, the system can quickly analyze much data and decide how to best care for agricultural fields, which leads to greater production and less risk of contamination. The proposed framework looks like what you can see in Figure 1.

Fig. 1 A simple program that helps farmers manage their soil by letting them see how many nutrients their soil needs and what kinds of crops might do well with that amount of nutrients.

4 Experimental Results

This section presents a comprehensive overview of the experimental results derived from implementing our proposed system. The study was successfully carried out, demonstrating the approach's feasibility and effectiveness. This section begins by detailing the experimental setup, including the selection of the test environment, data sources, and the criteria used for evaluating system performance. The methodology encompasses field-based data collection and the application of ML models to analyze soil and climatic conditions relevant to crop yield optimization.

4.1 Experimental Setup

The study was conducted using the eighth generation Intel Core (TM) i5 with a 100200H processor registered at a speed of up to 2.4 GHz and 8 GB of RAM. Many tests were performed using different ML techniques using the Visual Studio tool. To collect various data from the crop field, NPK sensor that explained in Table **??**.

Table 2 Sensor Measurement Ranges and Descriptions

Sensor Parameter	Range	Description
NPK (N, P, K)	0–1999	Computes the relative concentration of nitrogen, phosphorus, and potassium in the soil
Temperature	±0.5°C	Measures soil temperature to assess environmental suitability for crops
Conductivity	0–1000 µS/cm	Evaluates the electrical conductivity of the soil, indicative of nutrient and salt levels
pH	3–9 pH	Measures soil acidity or alkalinity, important for nutrient availability
Humidity	0–100%	Determines the volumetric water content of the soil

4.2 Dataset Description

The Indian Chamber of Food and Agriculture gathered the dataset from the Kaggle repository [26], which is used for this research. There is a total of 2,200 data points, including 22 types of agricultural crops, the effect of NPK fertilizer, the pH level of the soil, and rainfall, temperature, and humidity. The average values in the dataset for external N, P and K fertilizers in agriculture are 50.55, 53.36 and 48.14 kg / hector, respectively, for the given environmental conditions. The researchers kept track of a temperature of 25.62 ° C ± 5.06 ° C, a relative humidity (RH) of 71.48% ± 22.26%, a pH level of 6.47 ± 0.77, and a precipitation amount of 103.46 ± 54.96 mm during the study. However, horticultural crops usually received an average of 50.55 kg/ha of nitrogen fertilizer, 53.36 kg/ha of phosphorus fertilizer, and 48.15 kg/ha of potassium fertilizer.

To evaluate the effectiveness of the proposed models, it is essential to demonstrate their performance in diverse geographical regions and various crop species. The system developed in this study exhibits strong generalizability, indicating its potential applicability in areas with similar environmental conditions. The methodology involved training and validating the model using a dataset comprising 2,200 records collected from different crop-growing regions. The dataset includes information on 22 distinct crop types, such as apple, banana, black gram, grape, beans, chickpea, coconut, coffee, cotton, jute, lentil, maize, moth bean, mung bean, pomegranate, pea, pigeon pea, watermelon, muskmelon, orange, papaya, and rice. This data was utilized to construct a robust crop recommendation system. The system generates recommendations based on seven key environmental and soil parameters: N, P, K, temperature, humidity, pH, and soil moisture. The overall process of the model's operation and decision-making workflow is illustrated in Figure 2, highlighting its structured approach to accurate crop prediction. Figure 3 illustrates the correlation matrix of the features within the Crop Recommendation training dataset X_train using a heatmap. This visualization facilitates the identification of linear relationships among variables and highlights potential multicollinearity issues, which are critical to consider during feature selection and model optimization.

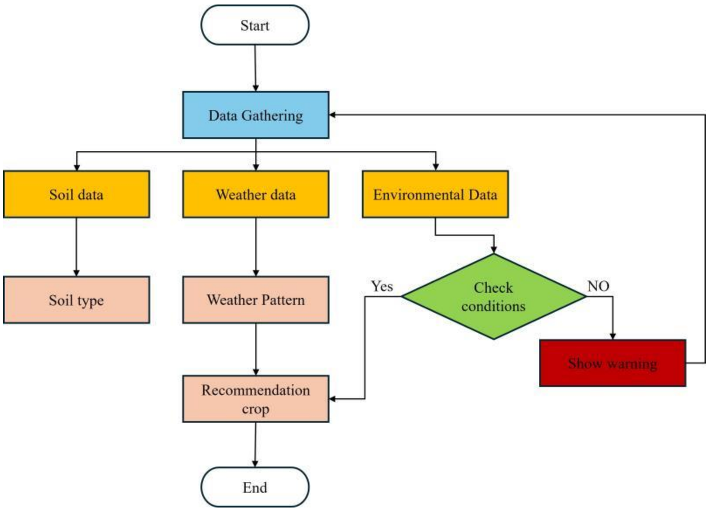

Fig. 2 A flowchart of recommending crops based on gathered soil.

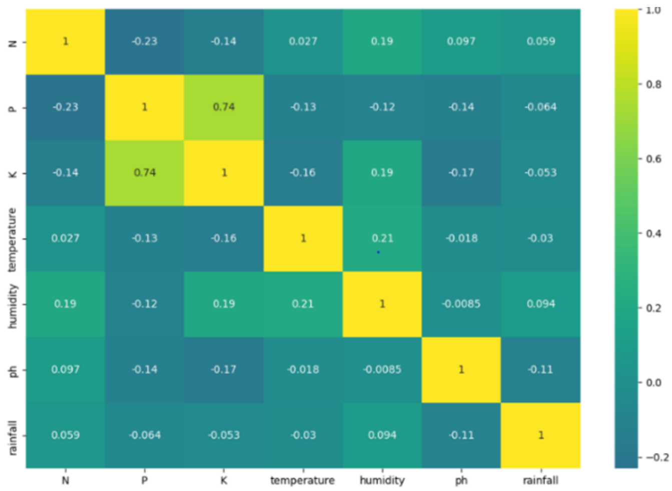

Fig. 3 The correlation between different features in the training dataset using a heatmap.

4.3 Evaluation Metrics

Before deploying the model in real-world simulations, conducting a thorough performance evaluation and refinement procedure is critical. Relying simply on accuracy as an evaluation metric may result in incorrect findings, especially if the dataset is uneven or specific errors are given more weight than others. As a result, a complete evaluation method should include a variety of performance indicators to guarantee that the model is robust and generalizable when applied to new data. To effectively test the suggested model's performance and dependability, this study employs a variety of evaluation metrics such as accuracy, precision, recall, and F1 score. These measures, taken together, provide a more comprehensive view of the model's predictive capabilities and help to reduce the danger of erroneous or biased results in practical applications (see Table 3).

Table 3 Performance Evaluation Metrics Used in the Study

Metric	Definition and Formula
Accuracy	Indicates the proportion of correctly classified samples in relation to the total number of samples in the dataset. It reflects the overall effectiveness of the model. $$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$
Precision	Measures the proportion of true positive predictions out of all positive predictions made by the model. It reflects the accuracy of positive classifications. $$\text{Precision} = \frac{TP}{TP + FP}$$
Recall (Sensitivity)	Also referred to as the True Positive Rate (TPR), recall measures the proportion of actual positive instances correctly identified by the model. It evaluates the model's ability to retrieve relevant results. $$\text{Recall} = \frac{TP}{TP + FN}$$
F1	Represents the harmonic mean of precision and recall, providing a balanced metric that considers both false positives and false negatives. It is especially useful in cases of imbalanced datasets. $$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.4 Discussion

The performance of the proposed crop recommendation system is quantitatively assessed using standard evaluation metrics, including accuracy, precision, recall, and the F1 score. These metrics offer a comprehensive understanding of how effectively the model recommends appropriate crops based on environmental and soil conditions. Accuracy indicates the general correctness of the model by measuring the proportion of correctly recommended crops out of the total number of predictions. Precision reflects the model's ability to avoid false positives, showing how many crops it recommends suitably. Accuracy, Recall evaluates the system's ability to identify all relevant crop options, thus indicating how effectively the model retrieves suitable crops from the dataset. The F1 score, the harmonic mean of precision and Recall, is a balanced metric that accounts for false positives and false negatives. This is especially important in agricultural decision-making, where recommending an unsuitable crop or missing a viable one can have significant economic or environmental consequences. The RF classifier was tested under five distinct experimental setups to validate the system's robustness.

As shown in Figure 4, the results present the comparative performance of these test cases. Each metric is visualized to highlight the consistency and reliability of the classifier in terms of both predictive accuracy and generalization capability. These evaluations confirm that the system can deliver precise, reliable, and context-sensitive crop recommendations, significantly helping farmers optimize crop selection and improving agricultural productivity. RF got the best accuracy out of all the ways of finding shapes, getting a result of 97.3% correct, finding 97.6% of what was there, and correctly picking out the right shape 97.3% of the time. Other models also work well. The support bus classifier had an accuracy of 87.5%, 86.8% accuracy, recall of 87.5%, grade F1 of 85.8%, while the logistic regression classifier got an accuracy of 91.5%, accuracy of 92.3%, recall of 91.5%, F1 score of 91.5%, LGBM classifier got a score of 96.3% for accuracy, accuracy was 96.6%, recall was 96.6%, F1 score was 96.9%, and the decision tree classifier ended up with a score of 96.8% for accuracy, accuracy was 97.0%, recall was 96.8%, and F1 score was 96.9%.

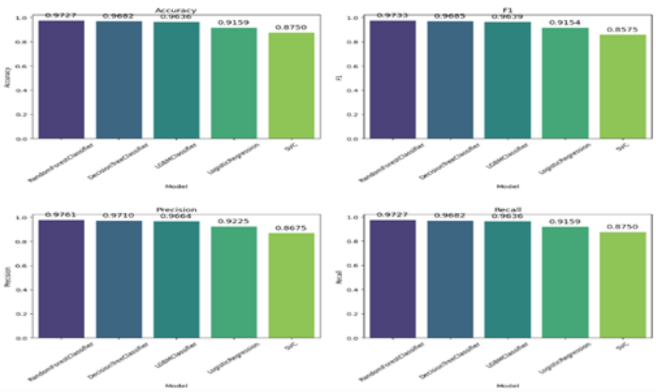

Fig. 4 Comparisons of 5 different model's evaluation metrics.

In addition, the confusion matrix was analyzed to gain insight into the specific types of classification errors committed by the system presented in Figure 5. When analyzing the results, the model displays many accurate positive and negative predictions across various categories. However, there are cases where the model generates incorrect output. Specifically, there are 50 iterations in which the model misclassified certain categories. Such errors may have the potential to influence crop selection decisions in the agricultural field. Table 4 represents a comparative performance analysis of various ML algorithms used in the modern literature to recommend the right crop for cultivation in agricultural fields. Indicates that the random forest classifier shows promising performance with the precision grade 97% in accurately recommending crops based on specific input parameters. This study shows that the device and

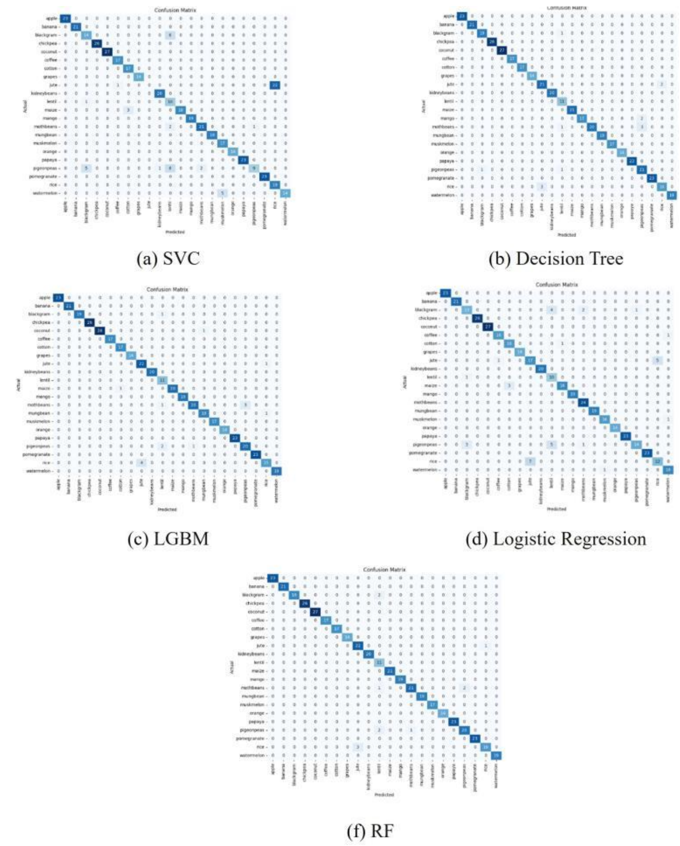

Fig. 5 Confusion matrix of five model classifiers..

framework developed are highly promising in the field of precision agriculture. An important outcome of using ML methods in our project was the ability to monitor soil nutrients in real time. Using the NPK Five Pins sensor in the crop field, we gathered data on NPK, humidity, air temperature, and relative soil moisture. The information gathered was transmitted to the decision-making form. ML algorithms and in particular the LGMB classifier and RF played a crucial part in selecting the most appropriate crops for a given environment. Both algorithms were trained and fine-tuned using the inbuilt datasets before achieving precision in diagnosing the best type of crop and the optimal fertilizer. Crop types and moisture levels in general. LGBM performed well in determining the highest yield by using variables including N, P, K, atmospheric temperature, humidity, soil pH, soil moisture and rainfall.

RF algorithm was better at identifying the best fertilizer based on factors consisting of N, P, K, temperature, humidity, soil type, crop type, and soil moisture.

Our evaluation demonstrated that the models we developed were highly accurate in suggesting the right crops and fertilizers in most cases. It should be considered that occasionally the models might make incorrect predictions, which may affect farmers' choices regarding crops and fertilizers. Continued efforts in research and development are critical to increasing the accuracy and usability of the recommendation system for farmers. Several validation steps were incorporated to ensure the reliability of the results, including cross-validation techniques and performance metric evaluations such as accuracy, precision, and recall. The system was tested under various environmental scenarios to assess its robustness and adaptability. Furthermore, we compared the predictive performance of five distinct ML algorithms SVC, LGBM Classifier, RF, Logistic Regression, and Decision Tree Classifier across multiple test cases.

Table 4 Performance Comparison of Different Classifiers

Classifier	Precision	Accuracy	F1 Score	Recall
Decision Tree	97.0%	96.8%	96.9%	96.8%
SVC	86.8%	87.5%	85.8%	87.5%
Logistic Regression	92.3%	91.6%	91.5%	91.6%
LGBM	96.6%	96.4%	96.4%	96.4%
RF	97.6%	97.3%	97.3%	97.3%

5 Conclusion

The research showed that AI, namely ML approaches, can improve agricultural decision-making by assessing complicated connections between crop yields, soil nutrient levels, and climate variability. Using critical factors such as N, P, K, soil pH, temperature, precipitation, and humidity, the proposed ML-based framework makes data-driven recommendations for optimal crop selection and nutrient management in various geographic locales. Five classifiers (SVC, LGBM, RF, Logistic Regression, and decision tree) were trained using yield data from 22 crops and achieved accuracy scores of 87.50%, 96.82%, 91.59%, 97.27%, and 91.59%, respectively. These findings demonstrate the ability of machine learning algorithms to deliver reliable and precise recommendations, boosting agricultural output and decision support. Experimental results demonstrated the proposed crop recommendation system's practical applicability by comparing model projections to traditional agricultural approaches. The findings show the system's ability to raise agricultural yields, eliminate input waste, and optimize resource utilization. Furthermore, the study adds to sustainable farming practices by addressing water footprint reduction, soil nutrient optimization, and informed fertilizer use. The proposed system is essential to integrating intelligent systems into agriculture, with obvious advantages for environmental sustainability and global food security.

References

[1] Csathó, P., Arendas, T., Fodor, N., N´emeth, T.: Evaluation of different fertilizer recommendation systems on various soils and crops in hungary. Communications in Soil Science and Plant Analysis **40**(11-12), 1689–1711 (2009)

[2] Wu LiangQuan, W.L., Wu Liang, W.L., Cui ZhenLing, C.Z., Chen XinPing, C.X., Zhang FuSuo, Z.F.: Basic npk fertilizer recommendation and fertilizer formula for maize production regions in china. (2015)

[3] Singh, P., Agrawal, M., Agrawal, S.B.: Evaluation of physiological, growth and yield responses of a tropical oil crop (brassica campestris l. var. kranti) under ambient ozone pollution at varying npk levels. Environmental Pollution **157**(3), 871–880 (2009)

[4] Shinde, M., Ekbote, K., Ghorpade, S., Pawar, S., Mone, S.: Crop recommendation and fertilizer purchase system. International Journal of Computer Science and Information Technologies **7**(2), 665–667 (2016)

[5] Cao HanBing, C.H., Wang ZhaoHui, W.Z., He Gang, H.G., Dai Jian, D.J., Huang Ming, H.M., Wang Sen, W.S., Luo LaiChao, L.L., Sadras, V., Hoogmoed, M., Malhi, S.: Tailoring npk fertilizer application to precipitation for dryland winter wheat in the loess plateau. (2017)

[6] El-Marsafawy, S.M., Mohamed, A.I.: Water footprint of egyptian crops and its economics. Alexandria Engineering Journal **60**(5), 4711–4721 (2021)

[7] Awad, M., Khanna, R.: Support vector machines for classification. In: Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, pp. 39–66. Springer, ??? (2015)

[8] Safavian, S.R., Landgrebe, D.: A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics **21**(3), 660–674 (1991)

[9] Pal, M.: Random forest classifier for remote sensing classification. International journal of remote sensing **26**(1), 217–222 (2005)

[10] Sajid, S.W., Hasan, M., Rabbi, M.F., Abedin, M.Z.: An ensemble lgbm (light gradient boosting machine) approach for crude oil price prediction. In: Novel Financial Applications of Machine Learning and Deep Learning: Algorithms, Product Modeling, and Applications, pp. 153–165. Springer, ??? (2023)

[11] Usman, M., Nangere, M., Musa, I.: Effect of three levels of npk fertilizer on growth parameters and yield of maize-soybean intercrop. International Journal of Scientific and Research Publications **5**(9), 1–6 (2015)

[12] Zhang, Y., Xia, C., Zhang, X., Sha, Y., Feng, G., Gao, Q.: Quantifying the relationships of soil properties and crop growth with yield in a npk fertilizer

application maize field. Computers and Electronics in Agriculture **198**, 107011 (2022)

[13] Yadav, R., Dwivedi, B., Pandey, P.: Rice-wheat cropping system: assessment of sustainability under green manuring and chemical fertilizer inputs. Field crops research **65**(1), 15–30 (2000)

[14] Ahmad, M., Afzal, M., Ahmad, A., Ahmad, A., Azeem, M.: Role of organic and inorganic nutrient sources in improving wheat crop production (2012)

[15] Misbah, K., Laamrani, A., Khechba, K., Dhiba, D., Chehbouni, A.: Multi-sensors remote sensing applications for assessing, monitoring, and mapping npk content in soil and crops in african agricultural land. Remote Sensing **14**(1), 81 (2021)

[16] Gottemukkala, L., Jajala, S.T.R., Thalari, A., Vootkuri, S.R., Kumar, V., Naidu, G.M.: Sustainable crop recommendation system using soil npk sensor. In: E3S Web of Conferences, vol. 430, p. 01100 (2023). EDP Sciences

[17] Singh, P., Agrawal, M., Agrawal, S.B.: Evaluation of physiological, growth and yield responses of a tropical oil crop (brassica campestris l. var. kranti) under ambient ozone pollution at varying npk levels. Environmental Pollution **157**(3), 871–880 (2009)

[18] Baligar, V., Bennett, O.: Npk-fertilizer efficiency—a situation analysis for the tropics. Fertilizer research **10**, 147–164 (1986)

[19] Dey, B., Ferdous, J., Ahmed, R.: Machine learning based recommendation of agricultural and horticultural crop farming in india under the regime of npk, soil ph and three climatic variables. Heliyon **10**(3) (2024)

[20] Praneeth, N., Kalyan, K., et al.: Iot based smart monitoring system for enhancing precision agriculture and environmental farming system (2024)

[21] Guido, R., Ferrisi, S., Lofaro, D., Conforti, D.: An overview on the advancements of support vector machine models in healthcare applications: a review. Information **15**(4), 235 (2024)

[22] Sai, G.V., Thinakaran, K.: Efficient music genre prediction system using novel enhanced decision tree classifier over naive bayes with improved accuracy. In: AIP Conference Proceedings, vol. 3252, p. 020064 (2025). AIP Publishing LLC

[23] Mahmoudzadeh, A., Amiri-Ramsheh, B., Atashrouz, S., Abedi, A., Abuswer, M.A., Ostadhassan, M., Mohaddespour, A., Hemmati-Sarapardeh, A.: Modeling co2 solubility in water using gradient boosting and light gradient boosting machine. Scientific Reports **14**(1), 13511 (2024)

[24] Das, A.: Logistic regression. In: Encyclopedia of Quality of Life and Well-being

Research, pp. 3985–3986. Springer, ??? (2024)

[25] Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., Liang, X.: An improved random forest based on the classification accuracy and correlation measurement of decision trees. Expert Systems with Applications **237**, 121549 (2024)

[26] Dataset: Dataset. Accessed: May 30, 2025 (2025). https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset