International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 4, July Year 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Intelligent Agriculture: Enhancing Crop Choice with Machine Learning and Predictive Analytics

Saranya R[1], Nishanth Kumar A[2] , Albert Samuel I[3]

*(*Computer Science and Engineering, Amrita College of Engineering and Technology, Nagercoil, Tamil Nadu, India*

Email: saranyamvasan@gmail.com)

** (*Computer Science and Engineering, Amrita College of Engineering and Technology, Nagercoil, Tamil Nadu, India*

Email: nishanthkumar3101@gmail.com)

(Computer Science and Engineering, Amrita College of Engineering and Technology, Nagercoil, Tamil Nadu, India

Email: albertsamuel347@gmail.com)

----------**********----------

Abstract:

Agriculture is a vital sector in the economy of India, providing employment to a large population. However, many farmers face productivity challenges due to improper crop selection that does not align with soil requirements. Precision agriculture addresses this issue by analyzing characteristics of a soil, types of soil, and information of crop yield to recommend the most suitable crops. This approach enhances productivity by reducing the cultivation of non-suitable crops and improving resource efficiency. Additionally, predicting agricultural productivity is essential in forecasting agriculture output using historical data, including temperature, relative humidity, soil pH, rainfall, and cultivated area. A system of recommendations employing an ensemble technique with voting methods, utilizing K-Nearest Neighbor (KNN) and Random Forest (RF), enhances accuracy and efficiency in crop selection. This method ensures data-driven decision-making, optimized resource utilization, and improved agricultural outcomes, contributing to sustainable farming practices. A comparison of both methods indicates the approach's resilience and dependability, with Random Forest outperforming in managing complex, non-linear interactions within the data. The suggested system intends to provide farmers with actionable insights, reduce crop failure risks, and enhance precision agriculture by implementing AI-driven methodologies.

Keywords — Random Forest, K-Nearest Neighbor, Agriculture, crops.

----------**********----------

I. INTRODUCTION

This document is a template. An electronic copy Agriculture is a crucial industry for global food security, but farmers sometimes struggle to make informed crop selection selections due to unexpected weather conditions, soil variation, and a lack of expert guidance [1]. The advent of digital technology, particularly in Artificial Intelligence (AI) and data analytics, presents an increasing possibility of transforming conventional farming into a more accurate and data-driven practice [2]. Intelligent agriculture uses machine learning and predictive analytics to help farmers make reliable, timely choices to enhance crop output, eliminate wasteful use of resources, and adjust to varying circumstances in the environment [3]. Many countries are still experiencing hunger as a result of food shortages or scarcity due to increasing population. Increasing production in food is a convincing procedure for eliminating starvation. As a result, protecting crops, land evaluation, and forecasting crop yields are increasingly important to global food production [4].

A. *Problem Statement*

1. Agricultural production in India is unstable due to its reliance on the monsoon, which causes periodic swings in crop yield, prices, revenue, and employment.
2. Cropping patterns in India's agricultural sector have shifted, with not food-related crop yields declining and rabi production becoming nearly as important as kharif, indicating a structural change.
3. Land ownership within India is uneven, with a minority of wealthy landowners and numerous tiny, unprofitable properties, resulting in low productivity, high expenses, and extensive rural poverty.
4. Due to population increase and family division, India's land holdings are constantly subdivided and

fragmented, making cultivation uneconomic and impeding efficient land use, resulting in poor agricultural production.

5. The remainder of the work follows this format. In Section II, many innovations and methods for enhancing crop yields are reviewed. In Section III, the materials and methods used are described in depth, along with the suggested framework and the method of execution procedure in Section IV, the outcomes of the experiments are displayed in Section V, the conclusions and future work are succinctly explained.

II. RECENT WORKS

The Internet of Things (IoT) produces enormous volumes of information in streams, recognized as "big data," that creates fresh opportunities regarding food and agriculture process surveillance. In addition to indicators, the food business is increasingly relying on big data from social media. We provide an overview of IoT, big data, and AI in this analysis, along to their disrupting potential to influence agri-food systems in future generations.

Monica Dutta, Deepali Gupta, and Sumegh Tharewali[5] investigate the transformational potential of IoT technology in modern farming, namely soilless production systems like hydroponics and aeroponics. This underline how IoT-enabled precision farming may assist to tackle important worldwide dietary issues by optimising resource utilization, increasing crop output, and lowering environmental impact. The article emphasizes several major prospects, such as real-time nutritional monitoring, agricultural system automated processes, and selecting choices determined by data.

Abid Badshah et al. [6] gives a research on the utilization of sophisticated methods of ML to enhance the productivity of agriculture. The researchers investigate various farming databases using robust techniques like RF, Support Vector Machines (SVM), and gradient boosting. Their outcome shows that those models could reliably categorize crop types and anticipate yield according to variables such as soil qualities, weather conditions, and historical data, resulting in improved sustainability and sophisticated farming methods.

Hulam Mohyuddin et al. [7] highlights critical difficulties such as data heterogeneity, a lack of standardized processes, and the necessity for analytics in real time. Overall, the research is an invaluable tool for academics and professionals who want to incorporate intelligent technology into agriculture to improve making choices and efficiency in operation [8]. Using the Rice Seedling and WeedNet datasets, the study presents a five-stage method that includes data collecting, feature extraction, classification, segmentation, and prediction [9].

Experimental findings show that the AgriChainSync infrastructure enhances memory efficiency, shortens processed time, and optimizes network efficiency, rendering it a promising solution for large-scale smart farming operations. This article emphasizes the possibilities for integrating IoT, blockchain, and AI technology to develop a safe, scalable, and efficient system for precision agriculture [10]. The authors suggest combining these AI approaches with environmental and climatic data to improve prediction accuracy. Furthermore, the paper discusses the issues of data quality, model generalization, and processing needs [11].

This research focuses on creating a system for decision-making that employs ML algorithms, notably Random Forest and K-Nearest Neighbor (KNN), to examine numerous agricultural elements like properties of a soil, climatic condition, and historical information of a crop. The system's goal is to improve agricultural production as well as sustainability by finding patterns and anticipating the best crops for certain conditions, eventually assisting in agriculture's modernisation using smart technology.

III. PROPOSED METHODOLOGY

With the growth of IoT devices and the collecting of massive volumes of data, worries about security and confidentiality of information have grown [12]. Agricultural and food firms have to address potential weaknesses in their IoT networks and big data platforms to secure private information from cyberattacks, illegal access, and data thefts. The system architecture for AI-assisted software for farmers on agricultural practices and suggested crops includes numerous critical components and features that are adapted to farmers' various demands [13]-[14]. To begin, the app will have a simple interface that farmers may use on their tablets or smartphones to browse through many functions and capabilities. The interface has been naturally built to suit users of various technology levels of literacy, guaranteeing widespread acceptance and usage among farmers.

Second, the app's backend architecture shall be backed by advanced AI algorithms and data analysis abilities. The algorithms involved will analyze massive volumes of farming data, elements like parameters of soil condition, climate trends, crop information about performance, and historical yield data. The software utilizes ML to create specific suggestions for farmers, such as optimal crop selection, planting schedules, irrigation tactics, pest control methods, and harvests practices [15].

The AI-assisted app for farmers has numerous levels, including a fundamental backend system driven by AI and ML. This system analyzes agricultural information such as condition of a soil, climate trends, crop performance, and yield history to make specific crop recommendations. A user-friendly interface allows farmers to interact with the system, gain real-time data, and make informed decisions to enhance farming outcomes. The architecture also includes a user interface layer in a form of an application for mobile devices that could be reached using cellphones or tablets. This user-friendly interface allows farmers to acquire individualized crop suggestions, obtain real-time warnings on weather or soil conditions, and share insights

or opinions with others in the agricultural sector, providing seamless user involvement and making decisions support. Fig. 1 shows the suggested method's architecture diagram.

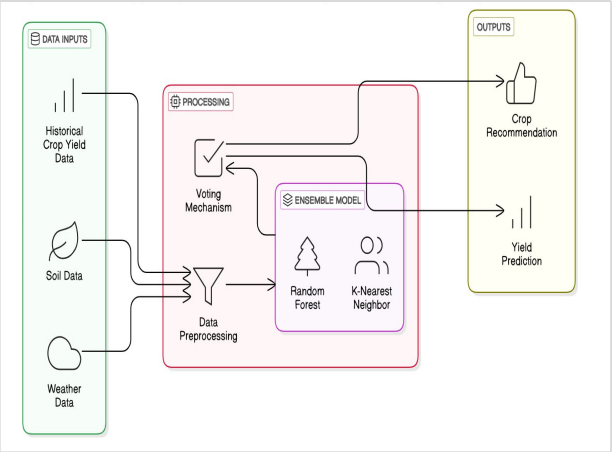

Fig. 1 Architecture diagram

A. Random forest

RF is a supervised ML technique for classification and regression problems. Its cornerstone is the concepts of ensemble learning, which combines several classifiers to solve a challenging issue and increase the performance of the model. A classifier called Random Forest increases the database's predicted accuracy by averaging several decision trees on that data set. The ultimate result is predicted by the random rainforest as opposed to depending just on one tree of decisions, based upon the predictions' majority vote.

Ley Y be the feature vector of a soil, pH, nitrogen, temperature, humidity, Nitrogen, Phosphorus, Potassium and rainfall with Y=[y1, y2,y3….Yn]. T(x) is the decision tree output and k is the total number of trees. Then the predicted crop y is represented as:

$y=\text{mode}(\{T_1(y), T_2(y), T_3(y)\ldots. T_k(y))$ (1)

For example $T_1(y)$=wheat $T_2(y)$= mothbeans, $T_3(y)$= Rice.
Then y= mode({Wheat,Rice, mothbeans})= mothbeans

B. k-nearest neighbors' algorithm (K-NN)

The identification of patterns uses a nonparametric classification technique called K-NN. Membership in the class is the outcome. An item is categorized by its neighbors' vote majority, and when k is a positive integer, usually small, it belongs to the class that is more common between its KNN. The object's single nearest neighbor's class is simply applied if k = 1.

Compute the **distance** among a new input point Y and all training points Yi. Let $Y=[y_1,y_2,y_3,\ldots y_n]$ be the input features and $y_i=[y_{i1},y_{i2},y_{i3},\ldots y_{in}]$ be the training samples and d be the euclidean distance between input and training samples.

Let the k closest points be

Y=[y1, y2,y3….Yn]. Then the predicted crop y is represented as:

$y=\text{mode}(\{T_1(y), T_2(y), T_3(y)\ldots. T_k(y))$ …… (2)

y= mode({Wheat,Rice, mothbeans})= mothbeans

IV. RESULT AND DISCUSSION

Using ML methods like Random Forest and KNN, the suggested intelligent agricultural system was created and evaluated on a standard setup. A 500 GB hard disk, 2 GB of RAM, and a Pentium i3 processors, and a 15-inch LED monitor are all part of the hardware configuration. For interaction, simple input devices such a keyboard and mouse were utilized. The Windows 7 operating system powers the system's apps. The Python programming language and development tools like PyCharm and Visual Studio Code were used to carry out the implementation. Libraries such as OpenCV were used to improve visual analysis and feature extraction for image and data processing workloads.

A. Dataset Used

The database utilized in this researc was sourced primarily from Kaggle and other online platforms, containing soil-specific attributes and general crop data. It includes crops like rice, maize, lentil, banana, mango, coffee, and more. Key features are Nitrogen (N), Phosphorus (P), Potassium (K), Humidity, temperature, pH, and rainfall. These factors influence plant growth, nutrient uptake, and overall yield. For example, nitrogen aids leaf growth, phosphorus supports roots and fruiting, and potassium regulates plant functions. Temperature, pH, humidity, and rainfall further impact germination, development, and harvesting timelines. Three components may be distinguished in the output of the Python-prepared categorization model: Pre-model data visualization, model information visualization, and post-model data visualization.

All data visualizations, including those made using the real database to comprehend the properties and allocation of the dataset, are included in the pre-model visualization of data section. At least one factor's distribution is monitored using a density plot. When retrieving new information, it is crucial to independently verify how the elements are being disseminated. It offers a multitude of data. Density of the seven parameters—nitrogen, phosphorus, potassium, humidity, temperature, rainfall and pH are contrasted in Fig. 2

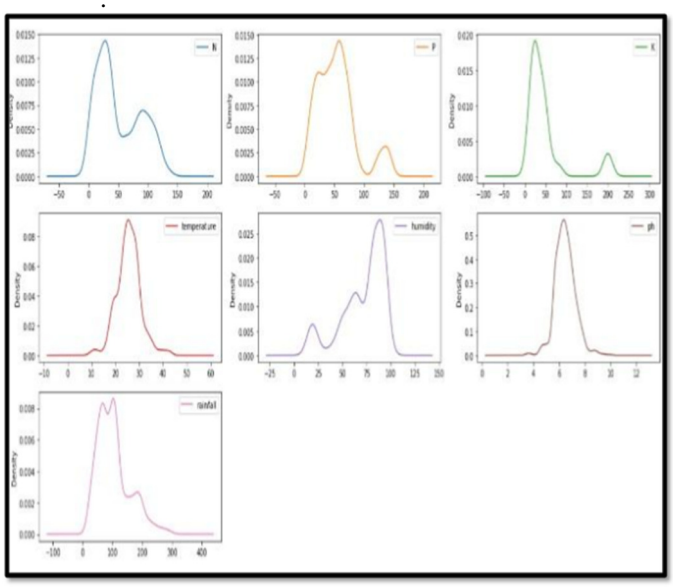

Fig. 2 Density Diagram

The correct and incorrect categories are depicted in Fig. 3 and 4. Based on the 22 training data examples, it incorrectly classified 22 times. The bias which was mentioned earlier is undoubtedly to blame for this. The integrity and dependability of the model are demonstrated by the absence of any additional misclassifications.

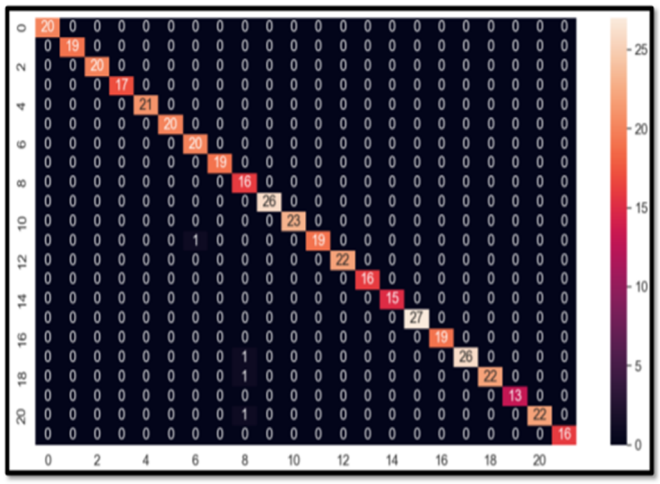

Fig. 3 Confusion Matrix for Random Forest

Fig. 5 illustrates how the relationship between variables enables us to ascertain exactly how strongly or weakly these are linked to one other. The number 1 indicates that there is a positive correlation among the variables. The more negatively the variables are related, the deeper the color of the numerical values 0.

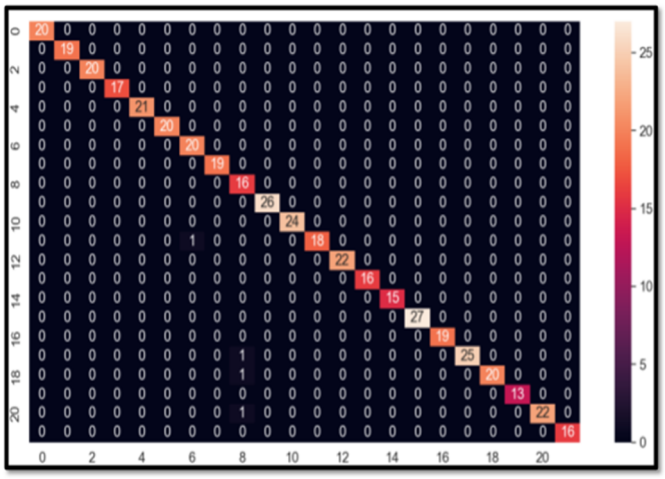

Fig. 4 Confusion Matrix for KNN

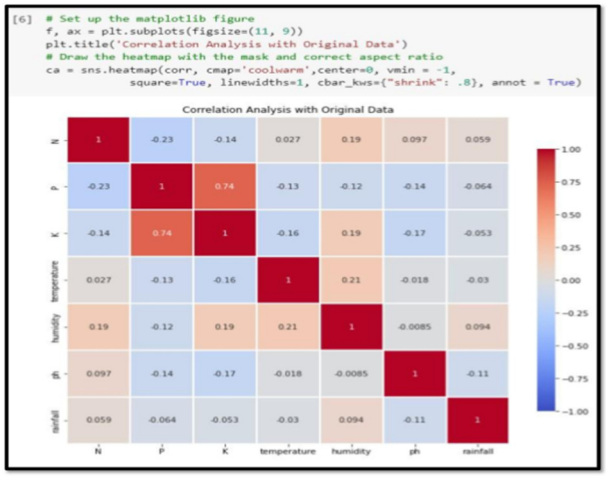

Fig. 5 Correlation Diagram

B. Analysis

For farmers, crop yield data is incredibly useful. Crop yields were once anticipated by seasoned farmers. The suggested system functions similarly. It forecasts future yields by applying the historical data. Crop production is mostly impacted by fertilizers and weather. The effectiveness of this prediction depends on how accurate the information is. Consequently, the proposed method minimizes loss and predicts yield. As an experienced farmer, the intended system assumes a function. On the other hand, it is more accurate and considers numerous other factors. Considerations include soil conditions, yield, pH, humidity, and weather forecasts.

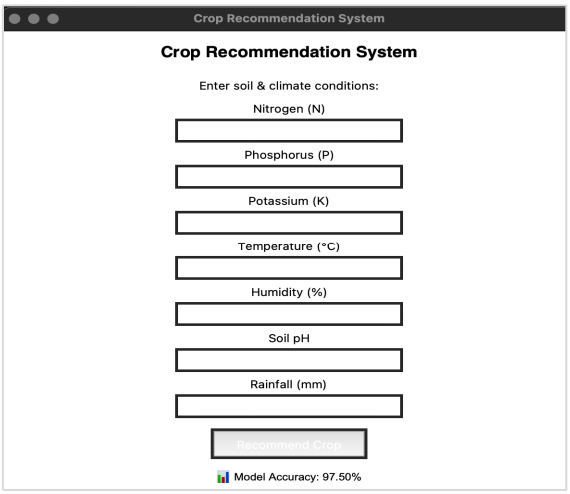

Fig. 6 Preview

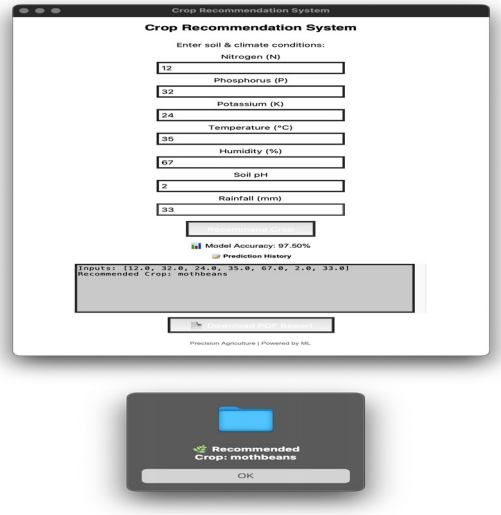

Fig. 7 Crop Prediction

Fig. 6 and 7 illustrate the way to use seven distinct variables to find out: heat, temperature, pH, precipitation, phosphorus, potassium, and nitrogen. Right now, you want to get a good harvest in order to get a good harvest. We achieved 9% more on KNN than we did on the previous task.

C. Performance Evaluation

The following measures were employed to evaluate the efficacy of the suggested system: F1-score, accuracy, recall or sensitivity, precision or specificity, and error rate. The following is a definition of these performance metrics. Table 1 and Fig. 8 shows the performance criteria of all models.

Accuracy

A classification problem statistic called accuracy shows the proportion of precise forecasts. By dividing the complete quantity of guesses by the total quantity of precise forecasts.

$$\text{Accuracy} = \frac{TN+TP}{(TN+TP+FN+FP)} \quad (3)$$

Where TN is True Negative, TP is True Positive, FN is False Negative and FP is False Positive.

Precision

It is evaluated by dividing the correctly estimated negative values by the sum of all negative values.

$$\text{Specificity or Precision} = \frac{TN}{TN+FP} \quad (4)$$

Recall

It's evaluated by splitting the total size of correct data by the size of correctly estimated information.

$$\text{Sensitivity or recall} = \frac{TP}{TP+FN} \quad (5)$$

F1-score

$$F1\text{-}score = \frac{2*Precision*Recall}{Precision+Recall} \quad (6)$$

This is described as the recall and precision harmonic mean. Other names for it include the F1 Score and the F1 Measure. Stated differently, the F1 score represents the equilibrium among recall and precision. It is believed to represent a more accurate metric than precision and recall alone since it is challenging to strike a balance among the two.

TABLE I
Algorithm Comparison

	RANDOM FOREST	KNN
ACCURACY	0.995	0.988
PRECISION	0.996	0.996
RECALL	0.995	0.989
F1-SCORE	0.995	0.989

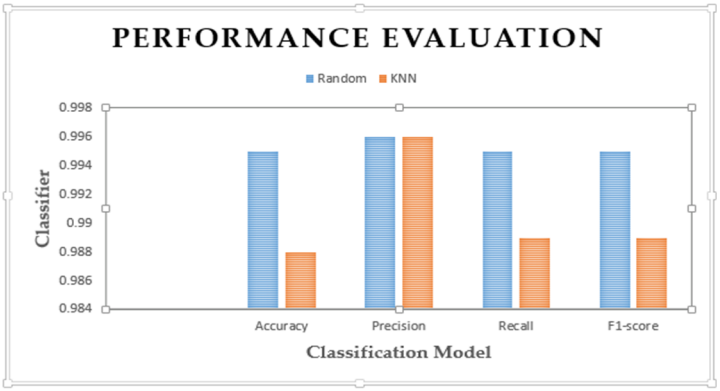

Fig. 8. Algorithm Comparison

Both Random Forest and KNN models were evaluated and RF achieved an accuracy of 99.5% and F1-score of 99.5%, slightly outperforming KNN, which had an accuracy of 98.8% and F1-score of 98.9%. Precision remained consistent at 99.6%

for both models. These results confirm the efficiency of ML in improving crop recommendation accuracy.

V. CONCLUSION AND FUTURE WORK

A variety of agricultural fields use ML methods. System that uses historical data gathering to forecast crop production. Execute a plan to speculate on widely accepted rural creation. Data handling methods are prepared for plant creation. These days, an interesting yield is described as a true harvest using random forest. It is generally in line with expected yields in agribusiness. The application of ML algorithms such as **RF** and **KNN** demonstrates significant potential in optimizing crop selection according to ecological and soil factors. By analyzing key attributes like Nitrogen, Phosphorus, Potassium, temperature, humidity, pH, and rainfall, the model provides precise crop suggestions based on particular conditions. Random Forest offers high accuracy and robustness by aggregating decisions from multiple decision trees, while KNN provides intuitive, distance-based predictions that adapt well to local variations in the data. Together, these techniques contribute to smarter, data-driven farming decisions, enabling improved contemporary farming, yields for crops, effectiveness of resources, and environmental sustainability. Future research will concentrate on growing the database with more attributes and incorporating crop yield prediction for greater accuracy and impact.

REFERENCE

[1] Mohyuddin, G., Khan, M. A., Haseeb, A., Mahpara, S., Waseem, M., & Saleh, A. M. (2024). Evaluation of Machine Learning approaches for precision farming in Smart Agriculture System-A comprehensive Review. *IEEE Access*.

[2] Rani, S., Mishra, A. K., Kataria, A., Mallik, S., & Qin, H. (2023). Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports*, *13*(1), 15997.

[3] Akhter, R., & Sofi, S. A. (2022). Precision agriculture using IoT data analytics and machine learning. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 5602-5618.

[4] Shaikh, T. A., Rasool, T., & Lone, F. R. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture*, *198*, 107119.

[5] Shaikh, T. A., Rasool, T., & Lone, F. R. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture*, *198*, 107119.

[6] Badshah, A., Alkazemi, B. Y., Din, F., Zamli, K. Z., & Haris, M. (2024). Crop classification and yield prediction using robust machine learning models for agricultural sustainability. *IEEE Access*.

[7] Mohyuddin, G., Khan, M. A., Haseeb, A., Mahpara, S., Waseem, M., & Saleh, A. M. (2024). Evaluation of Machine Learning approaches for precision farming in Smart Agriculture System-A comprehensive Review. *IEEE Access*.

[8] Mohyuddin, G., Khan, M. A., Haseeb, A., Mahpara, S., Waseem, M., & Saleh, A. M. (2024). Evaluation of Machine Learning approaches for precision farming in Smart Agriculture System-A comprehensive Review. *IEEE Access*.

[9] Anwar, B., Morsey, M. M., Hegazy, I., Fayed, Z. T., & El-Arif, T. (2024). Towards Precise Agriculture: Integrating Machine Learning Techniques for Smart Farming Systems. *IEEE Access*.

[10] Tahir, H. A., Alayed, W., Hassan, W. U., Nabi, F., & Tran, T. X. (2024). AgriChainSync: A Scalable and Secure Blockchain-Enabled Framework for IoT-Driven Precision Agriculture. *IEEE Access*.

[11] Goel, M., & Pandey, M. (2024, March). Crop Yield Prediction Using AI: A Review. In *2024 2nd International Conference on Disruptive Technologies (ICDT)* (pp. 1547-1553). IEEE.

[12] AlZubi, A. A., & Galyna, K. (2023). Artificial intelligence and internet of things for sustainable farming and smart agriculture. *Ieee Access*, *11*, 78686-78692.

[13] Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. *IEEE access*, *9*, 63406-63439.

[14] Bhat, S. A., & Huang, N. F. (2021). Big data and ai revolution in precision agriculture: Survey and challenges. *Ieee Access*, *9*, 110209-110222.

[15] Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., & Aggoune, E. H. M. (2019). Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. *IEEE access*, *7*, 129551-129583.

