Enhancing Crop Management: Ensemble Machine Learning for Real-Time Crop Recommendation System from Sensor Data

Sadia Hossain
Internet of Things and Robotics Engineering
Bangabandhu Sheikh Mujibur Rahman Digital University
Gazipur, Bangladesh
sadiaupoma540@gmail.com

Nuzhat Noor Islam Prova
Pace University, Seidenberg School of CSIS
One Pace Plaza, New York City, NY 10038, USA
nuzhatnsu@gmail.com

Md Rezwane Sadik*
University of South Dakota
Vermillion, SD 57069, USA
mdrezwane.sadik@coyotes.usd.edu

Abdullah Al Maruf
Department of Computer Science & Engineering
Bangladesh University of Business & Technology (BUBT)
Dhaka, Bangladesh
marufbubt822@gmail.com

***Abstract*—The agricultural industry is essential to the world's food production, and it is critical to use cutting-edge technologies to increase crop productivity. We provide a revolutionary Crop Recommendation System (CRS) that utilizes cutting-edge technology to maximize crop output in response to the pressing need for improvement. Our study incorporates real-time monitoring of soil conditions, made possible by a custom hardware configuration that includes sensors for temperature, humidity, phosphorus, potassium, nitrogen, and pH measurements. First, we assembled a large dataset with 22 kinds of agricultural production components. Using many machine learning models, such as ensemble methods and baseline classifiers, we were able to classify crops with an astounding 99% accuracy rate. With the application of these insights, the CRS provides customized recommendations through an easy-to-use user interface for appropriate crops under particular climatic conditions. Our system's innovative combination of hardware sensing capabilities and AI-driven decision-making promises to revolutionize crop management practices, offering actionable insights for agricultural stakeholders. Our system's novel integration of AI-driven decision-making and hardware sensing capabilities promises to transform crop management techniques and provide agricultural stakeholders with useful insights.**

***Index Terms*—Crop Recommendation, Agriculture, Machine Learning, Ensemble Model, Internet of Things (IoT)**

I. Introduction

The agricultural enterprise in Bangladesh is a pivotal driving force of the country's monetary enlargement and also supplying employment for a noteworthy 60% of the population and contributing a considerable 16% to GDP. Bangladesh additionally produces jute, tea, wheat, and sugar. In addition to an extensive collection of culmination and greens. Organic fertilizers and natural strategies of pest manipulation are examples of the way conventional agricultural practices in Bangladesh have cultivated a profound reference to the natural world. However, Bangladesh has encountered limitations because of winning development patterns, such as it encompasses the improper application of insecticides and fertilizers, insufficient irrigation infrastructure and the imperative to benefit from a greater complete understanding of customer needs.

In the context of precision agriculture and crop yield prediction, researchers have investigated a wide range of methodologies, gaining valuable insights and addressing a number of obstacles along the way. As an illustration, Sangeeta et al. [1] utilized machine learning techniques such as random forest and decision trees for the purpose of conducting historical analysis, and the findings appeared to be encouraging. Miriyala et al. [2] made use of deep learning and remote sensing but emphasized the significance of visual quality throughout the presentation. Also, there was a lack of particular performance measurements, which prevented Masare et al. [3] from achieving the goal of optimizing agricultural productivity through the use of decision trees and KNN. Chehri et al. [4] investigated the possibilities of the Internet of Things and brought attention to difficulties such as the unpredictability of the weather. However, there are no performance standards for the Internet of Things system that Anand et al. [5] propose for direct farmer connectivity, and Gosh et al. [6] proposed an Internet of Things (IoT) and machine learning method for predicting solar radiation, although there some uncertainty on the approach's applicability. There were some studies [7] that emphasized the importance of soil fertility research being recognized by a wider audience, but they did not highlight the specifics of the practical situations that were involved. Our research introduces a revolutionary crop recommendation system with 99% accuracy using a voting-based ensemble. Our research aims to address key in agricultural production optimization by

*Corresponding author

pursuing the following objectives:

- To integrate real-time prediction model integration with IoT devices improves usability and gives farmers valuable insights.
- To focus on reducing over-fitting and assuring robust performance across varied agricultural contexts, we employ a machine learning ensemble model.
- To revolutionize precision agriculture and surpass prior works with its accuracy, adaptability and use and also increase the accuracy rate and optimization more through Our ensemble model.
- To prioritize the development of a simplified interface for agricultural production and usability optimization.
- To streamline data collection, analysis and decision-making through IoT devices and our user-friendly recommendation interface, which gives suggestions in the Bengali language.

The remaining section is organized as follows: Section II discusses recent state-of-the-art works, Section III elaborates on the materials and methods employed, Section IV delves into the experimental results and discussion, and finally, Section IV addresses the conclusion and future scope of our project.

II. Literature Review

Researchers have investigated a range of methodologies in the fields of precision agriculture and crop yield prediction, each providing distinctive insights and solving particular issues. The goal of Sangeeta at el. [1] was to analyze historical data using machine learning techniques like decision trees and random forest polynomial regression. This method showed encouraging results with a training accuracy of 80 and a testing accuracy of 85; nevertheless, difficulties were found in the selection of datasets and parameter tweaking Miriyala et al. [2] while admitting the need for better image clarity the book explored the challenges of employing deep learning remote sensing and wireless sensor networks to predict crop yields with a focus on the usage of drones and satellites masare et al. [3] clarified the goal of their system was to increase agricultural productivity in Indian agriculture by using Naive Bayes and Decision Tree classifier KNN although the study lacked detailed information on performance matrices. Cock et al. [8] and Waikar et al. [9] both employed distinctive approaches by combining operations research and machine learning by extracting valuable insights from farmers' knowledge and stressing the importance of precision agriculture. Chehri at el. [4] examined how IOT and wireless sensors could revolutionize agriculture, enabling more productive and smarter farming while noting the difficulties brought on by erratic weather patterns and conventional farming practices. Gupta et al. [10] envisioned an IOT-powered system that connected farmers directly to portable devices measuring real-time parameters with data seamlessly transferred to the thing speak cloud, a hybrid machine learning algorithm predicted suitable crops for specific environmental conditions promising timely decisions and optimized resource usage but the lack of performance metrics and a clear procedural explanation raise doubts about the reliability of their approach. Anand et al. [5] employed an IoT-powered system connecting farmers directly to portable devices measuring real-time parameters and facilitated with data seamlessly transferred to the ThingSpeak Cloud. Sheeba et al. [11] investigated soil fertility in specific districts, emphasizing the necessity for broader validation across diverse landscapes. Ghosh et al. [6] proposed a tech-oriented solution using IoT, Arduino, and machine learning to predict solar radiation while showcasing successful simulations and accurate predictions through a linear regression model. Zunzunwalacrop et al. [7] and AlZubi et al. [12] both utilized machine learning algorithms, such as Random Forest, Decision Tree, and Logistic Regression, for accurate crop predictions and real-time field insights. However, it lacked specific details on performance metrics and the overall procedure.

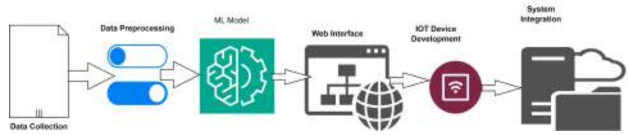

Fig. 1: The framework of the proposed research.

Examining earlier research makes it clear that real-time analysis methodology exploration is noticeably lacking. Current studies are primarily based on gathering sensor data and then making decisions manually or using only classification tools. These methods, however, are not able to offer recommendations that are based on dynamic, real-time datasets. As a result, there is a great chance to further research in this field and create intelligent systems that can use real-time data streams to make proactive decisions.

III. Methodology

A. Dataset Description & Processing

The dataset contains 8 columns such as Nitrogen(N), Phosphorus(P), Potassium(K), Temperature, Humidity, pH, Rainfall, and label(crop). The system predicts the best crop considering the above seven parameters. A total of 22 classes are in the dataset that we have used. Each class contains 100 instances, and the total instances are 2200. The unique 22 crops are there such as rice, maize, jute, cotton, coconut, papaya, orange, apple, muskmelon, watermelon, grapes, mango, banana, pomegranate, lentil, black gram, mung bean, moth beans, pigeon peas, kidney beans, chickpea, coffee in this data set. After collecting the raw dataset, we performed some steps to make the dataset a suitable format. We performed the cleaning and checked if there were any null values. We have also applied feature scaling in our processing steps. In the machine learning model, feature scaling normalizes and standardizes the dataset's range of independent variables or features. This makes sure that no feature outweighs the others and promotes faster and better algorithm convergence. In our research we have used Robust and Unit Vector scaling.

We used train_test_split for dividing the training and testing dataset .80% data were used for training and 20% testing.

B. Model

In our research, firstly, we applied 4 baseline classifiers, namely, SVM, KNN, DT, and NB. Later, with the help of the baseline classifiers, we applied the voting base ensemble model. We applied soft-voting approaches in our research. In order to identify whether data belongs to a specific class, our specially designed soft voting ensemble classifier combines several classifiers, each of which makes decisions based on its probability values. Predictions are merged and weighted according to the importance of the classifier in the soft voting ensemble method, yielding the total of weighted probabilities. Because it has the highest voting value, the target label with the largest sum of weighted probabilities is chosen. To determine the weighted average and assign greater significance and involvement to a particular learning model (base classifier), customized weights can also be employed. First, we use the following formula to determine the overall accuracy of all base classifiers:

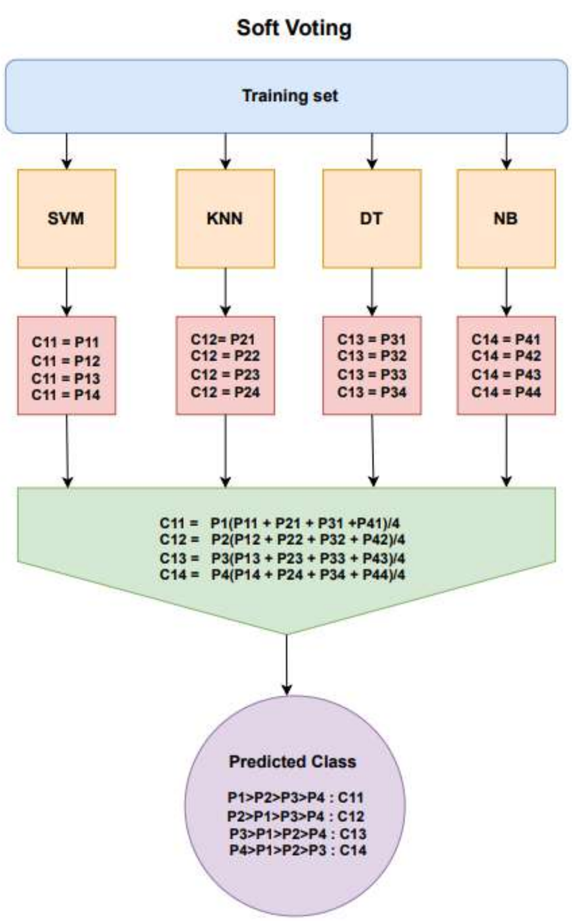

Fig. 2: Architecture of the proposed voting ensemble model

All of the baseline classifiers performed well in our research. So, we put equal weights of all the classifiers except the KNN. The soft voting classifier architecture is shown in Figure 2. Let's consider an example of how our base classifiers performed to classify the classes. Suppose our first classifier (SVM) predicts the class is Rice with 0.9 probability and Jute with 0.1 probability. Similarly, KNN, DT, and NB predict the class Rice with 0.85,0.89, and 0.88 probability, and they predict the class can be Jute with 0.15, 0.11, and 0.12 probability. Our ensemble model will predict the class as Rice because the total probability is bigger than the jute class. Every base classifier—SVM, KNN, DT, and NB—contributes its insights to the ensemble classification scenario by offering probability estimates for the classes ”Rice” and ”Jute.” The classifiers' collective confidence in their predictions is reflected in the harmonic aggregation of probabilities. Interestingly, the ensemble model chooses the class with the highest cumulative probability, thereby identifying the dominant class with great intelligence. In this particular case, the ensemble model chooses to classify the sample as ”Rice” because it has a higher total probability than the other class, ”Jute.” By cleverly combining the outputs of each classifier, the ensemble is able to make a strong and well-informed decision, utilizing the various viewpoints of its component models to improve prediction accuracy.

The base classifiers ere initialized with various parameters for improving the classification model. The parameters of the classifiers are shown in Table I.

TABLE I: Parameters of baseline classifiers.

Classifier	Parameter
SVM	Kernel ='Linear',probability =True ,C=0.5 ,gamma ='scale'
DT	criterion='entropy',max_depth=5, min_samples_split=5
KNN	n_neighbors=3, weights='distance', algorithm='auto', p=1

C. Hardware Interfacing

In this project, some hardware components are needed, Such as:

*a) **Arduino UNO:** :* The microcontroller board used for this project is the Arduino UNO, which has the ATmega328p at its heart. It has 14 digital input/output pins (6 of which enable PWM outputs), a USB connector, a power jack header, and a reset button. The Arduino UNO is user-friendly and versatile. We have used this as the central microcontroller of our hardware interfacing.

*b) **ESP8266/ NodeMCU:** :* NodeMCU is an open-source firmware and development kit for building any IOT product. The sensor data is sent to the NodeMCU (ESP8266) microcontroller board, which has Wi-Fi capabilities, allowing it to connect to the Internet and transmit the collected data to a dedicated server.

*c) **DHT11 Sensor:** :* The DHT11 temperature and humidity sensor features a complicated temperature and humidity sensor with a calibrated digital signal output. We have used a DHT11 Sensor to collect real-time temperature and humidity levels.

*d) **Soil-NPK Sensor:** :* The NPK sensor in the soil is critical for assessing the amounts of nitrogen, phosphorus, and potassium in the soil. This useful information enables farmers and researchers to understand the soil's nutrient content and make educated decisions about fertilization and crop management tactics to optimize plant development and output.

*e) **pH Sensor:** :* Soil pH meters assess the soil's acidity, neutrality (7), and alkalinity. This pH sensor gauges the activity of hydrogen ions and expresses it on a pH scale ranging from 0 to 14. A pH value of 7 denotes neutrality, values below 7 indicate acidity, and values above 7 indicate alkalinity. Maintaining an optimal soil pH is crucial for promoting healthy plant growth. So, a pH sensor can easily measure the current pH level and condition of the soil.

*f) **Rain Sensor:** :* The rain sensor measures rainfall in any atmosphere and setting. It provides real-time data on the current rainfall condition. We have used a rainfall sensor here. The integration of the hardware is shown in Figure 3.

D. Software Requirements and Implementations

In this project, some software components are needed, Such as:

*a) **Arduino IDE**:* The Arduino IDE is open-source software for constructing IoT projects. It contains a large collection of common input and output functions and is written in C and C++. It is used to build and upload programs to Arduino-comparable boards as well as other vendor development boards like the ESP8266. When we produce a sketch, it is processed and compiled into machine language.

*b) **Python IDE**:* Python IDEs are acronyms for python integrated development environments. It acts as a software platform providing programmers and developers with a full set of software development tools. Pycharm stands out as one of the best and most comprehensive full-featured python IDEs. We used PySimpleGUI in our research work, a Python tool that allows programmers of all levels to create graphical user interfaces (GUIs). Creating a user-friendly GUI that works consistently across several platforms may be difficult. However, by combining the power of Python and PySimpleGUI, we can create visually beautiful user interfaces that users will prefer more.

*c) **Data Flow from IOT Devices to User Interface**:* This system brings together different IoT sensors, like DHT for temperature and humidity, pH for soil acidity, NPK for soil nutrient levels, and a rain sensor for rainfall. These sensors link up with the NodeMCU (ESP8266) microcontroller, which can connect to Wi-Fi. The NodeMCU gathers real-time data on temperature, humidity, soil pH, NPK levels, and rainfall. After collecting data, the NodeMCU is set up to connect to the internet using Wi-Fi. It uses communication methods like HTTP to send the collected sensor data to a dedicated server. This server is prepared to handle and process incoming data, featuring a web server or API for efficiency. To secure the transmitted data, encryption measures are in place. On the other hand, there are programs written in Python using the PySimpleGUI framework to deploy the ensemble model and process the incoming sensor data. Our ensemble model comes into play for predicting crop conditions and providing recommendations for optimal crop growth. Users access this information through a user-friendly interface. The interface showcases real-time and historical sensor data, empowering users with insights for smarter decisions in crop management.

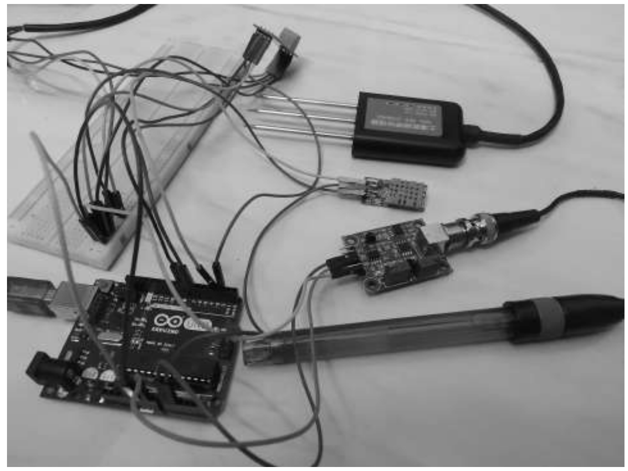

Fig. 3: Hardware Interfacing

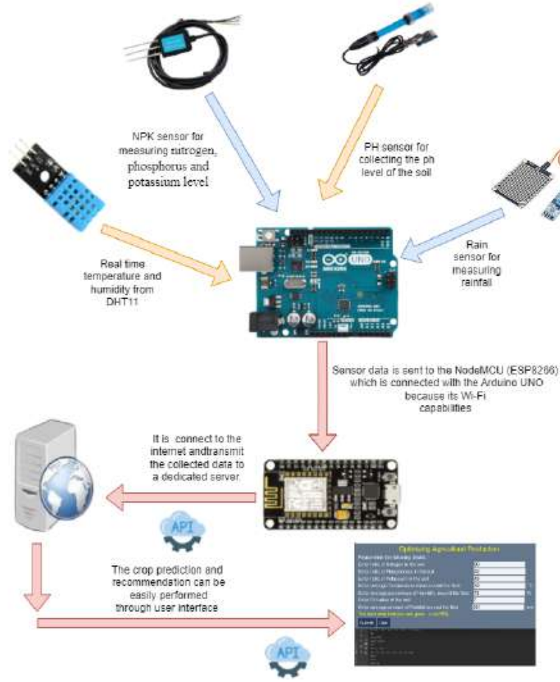

Fig. 4: Real-time data flow in the user interface.

IV. Result and Discussion

Several models have been used and assessed in the context of machine learning-based crop recommendation prediction, with key performance metrics like precision, recall, and F1-score being used as the basis for evaluation. For this task, several algorithms were considered: SVM, KNN, DT, NB, and a Voting Base Ensemble method. The performance is shown in Table II. The SVM model performed exceptionally well, achieving 98% for precision, recall, and F1-score. This demonstrates the model's capacity to correctly categorize crops that are advised for planting. Comparable results were also shown by the KNN and Decision Tree models, which achieved 97% precision, recall, and F1-score. Despite being marginally less accurate than SVM, these models nevertheless show good predictive power and offer trustworthy crop selection recommendations.

TABLE II: Performance metrics of different classifiers

Model	Precision	Recall	F1-Score
SVM	98%	98%	98%
KNN	97%	97%	97%
Decision Tree	98%	98%	98%
Naïve Bayes	98%	98%	98%
Ensemble	99%	99%	99%

TABLE III: Comparison our research with prior works

Paper Reference	Focus	Model	Accuracy
Modi et al. [13]	Focuses on the issue of poor crop selection and aims to enhance productivity by analyzing soil parameters.	SVM	98% .
Ray et al. [14]	Recommend crops accurately.	NB and ensemble	NB - 99.54%
Pande et al. [15]	Yield prediction of crops using mobile application connectivity for farmers.	SVM, ANN, RF, MLR, KNN	RF - 95%
Our Research Work	Real-time crop recommendation system.	Ensemble model.	99% accuracy.

Impressive results were also obtained by the NB model, which is well-known for its simplicity and efficiency. Its precision, recall, and F1-score were all 98%. This demonstrates how well-suited it is for agricultural tasks involving crop recommendation. Ultimately, surpassing the performance of the individual models, the Voting Base Ensemble model combination of all previous models achieved an astounding 99% precision, recall, and F1-score. By utilizing the advantages of various models, this ensemble approach improves the crop recommendation system's overall predictive accuracy and robustness. The training accuracy and testing accuracy of the classifiers are shown in Figure 5.SVM and DT, two of

the individual models showed good accuracy during training, attaining 97% and 98%, respectively. SVM performed well in testing, maintaining a 95% accuracy rate, while DT performed slightly better, with a 96% accuracy rate. With 97% accuracy during training and 94% accuracy during testing, KNN demonstrated consistency. NB performed well, hitting 98% accuracy in the testing and training phases. Interestingly, the Voting Ensemble model—a hybrid of multiple models—outperformed the others in terms of accuracy. It tested with an astounding 99% accuracy and completed training with a perfect 100% accuracy. The ability to combine different models for improved predictive performance in crop recommendation applications is demonstrated by this ensemble approach. It's critical to address potential overfitting issues in the assessment of machine learning models for crop recommendation prediction. When a model learns the training set too well—capturing details and noise—it becomes overfitted and finds it difficult to generalize to new, unobserved data. Even though specific models such as SVM, DT, and NB showed good accuracy in testing and training, it's important to consider any overfitting issues carefully. The Voting Ensemble's 100% training accuracy may be a sign that the training data is at risk of being overfitting. While the ensemble's testing accuracy of 99% is impressive, its perfect training accuracy begs the question of how well it can adapt to a variety of agricultural scenarios.

Figure 6 displays the confusion matrix for the ensemble model, which shows a very strong performance. The model predicts with minimal confusion, indicating a high degree of accuracy. To be more precise, the ensemble model only got confused twice: once when classifying pomegranate and jute and again when separating the classes chickpea and muskmelon. These rare occurrences of misunderstanding highlight how accurate the model is in the majority of classification situations. The ensemble model demonstrates an impressive error-avoiding capability, which strengthens its dependability in producing precise predictions across a range of crop classes.

Once our ideal model has been determined, a smooth integration into our user interface enables the dynamic recommendation of crops suited to particular soil types. The hardware elements utilized are crucial in measuring various soil parameters, and the collected information is easily sent to the server for handling. Our real-time process is shown in Figure 4, which shows an example of how our predictive model and IoT) devices interact intricately within the user interface. The interface facilitates the recommendations of our model, illustrating the interplay between cutting-edge data analytics and the useful application of IoT devices to mine valuable insights from the soil.

V. Conclusion

The diverse methodologies explored in precision agriculture and crop yield prediction, as highlighted by various researchers, have contributed valuable insights and solutions to specific challenges. Each approach, ranging from historical data analysis and deep learning to the integration of IoT and ML, has its strengths and limitations. Our motivation to revolutionize agriculture in Bangladesh stems from the gaps identified in existing work specific to the country. Through a meticulously planned project, we aim to provide a groundbreaking solution, leveraging continuous field surveillance and machine learning for precise data analysis. By addressing the unique challenges faced by Bangladeshi farmers and setting clear goals, we believe our initiative will not only transform agricultural practices locally but also serve as a

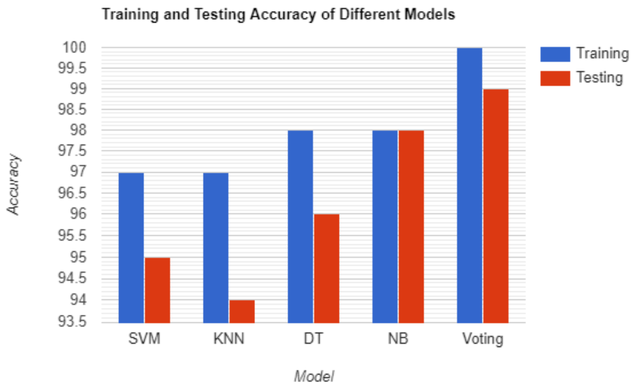

Fig. 5: Accuracy of the classifiers

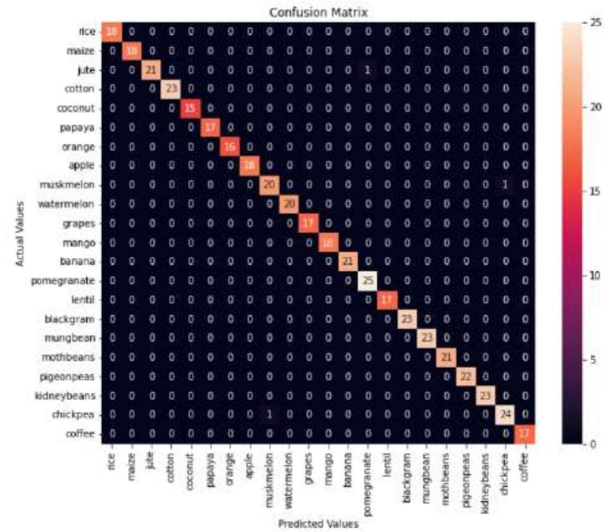

Fig. 6: Confusion matrix (Ensemble Model)

model for global advancements. The methodology, involving a well-defined dataset, processing steps, and a voting ensemble model, reflects our commitment to transparency and accuracy in predicting optimal crops. Our dedication extends beyond mere crop production optimization. It signifies a commitment to the prosperity and progress of the agricultural community in Bangladesh and potentially beyond. In the future, investigating how cutting-edge remote sensing tools like satellite imaging and unmanned aerial vehicles (UAVs) may be integrated could improve data collecting timeliness and accuracy. Through the use of these instruments, scholars can get a more profound understanding of crop well-being, soil attributes, and surrounding circumstances, thus improving prognostic models for optimal agricultural management.

REFERENCES

[1] S. G. Sangeeta, "Design and implementation of crop yield prediction model in agriculture," *International Journal of Scientific & Technology Research*, vol. 8, no. 1, pp. 544–549, 2020.

[2] G. Pratyusha and A. Sinha, *Prediction of Crop Yield Using Deep Learning Techniques: A Concise Review*, 06 2020, pp. 145–159.

[3] Y. Masare, S. Mahale, M. Kele, A. Upadhyay, and B. R. Nanwalkar, "The system for maximize the yielding rate of crops using machine learning algorithm," *Int J Eng Res Technol (IJERT)*, vol. 10, no. 6, pp. 453–458, 2021.

[4] A. Chehri, H. Chaibi, R. Saadane, N. Hakem, and M. Wahbi, "A framework of optimizing the deployment of iot for precision agriculture industry," *Procedia Computer Science*, vol. 176, pp. 2414–2422, 2020.

[5] R. Anand, D. Sethi, K. Sharma, and P. Gambhir, "Soil moisture and atmosphere components detection system using iot and machine learning," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2019, pp. 842–847.

[6] S. Ghosh, A. Sarkar, A. Mitra, and A. Das, "Smart cropping based on predicted solar radiation using iot and machine learning," in *2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI)*, 2020, pp. 1–5.

[7] C. Zunzunwala, S. Bisht, Y. Konghe, and M. Hasan, "Crop prediction for maximum yield using machine learning."

[8] J. Cock, D. Jiménez, H. Dorado, and T. Oberthür, "Operations research and machine learning to manage risk and optimize production practices in agriculture: good and bad experience," *Current Opinion in Environmental Sustainability*, vol. 62, p. 101278, 2023.

[9] V. C. Waikar, S. Y. Thorat, A. A. Ghute, P. P. Rajput, and M. S. Shinde, "Crop prediction based on soil classification using machine learning with classifier ensembling," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 05, pp. 4857–4861, 2020.

[10] A. Gupta, D. Nagda, P. Nikhare, and A. Sandbhor, "Smart crop prediction using iot and machine learning," *International Journal of Engineering Research & Technology (IJERT)*, pp. 2278–0181, 2021.

[11] T. Blesslin Sheeba, L. Anand, G. Manohar, S. Selvan, C. B. Wilfred, K. Muthukumar, S. Padmavathy, P. Ramesh Kumar, and B. T. Asfaw, "Machine learning algorithm for soil analysis and classification of micronutrients in iot-enabled automated farms," *Journal of Nanomaterials*, vol. 2022, 2022.

[12] A. Ali and K. Galyna, "Artificial intelligence and internet of things for sustainable farming and smart agriculture," *IEEE Access*, vol. PP, pp. 1–1, 01 2023.

[13] D. Modi, A. V. Sutagundar, V. Yalavigi, and A. Aravatagimath, "Crop recommendation using machine learning algorithm," in *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, 2021, pp. 1–5.

[14] R. K. Ray, S. K. Das, and S. Chakravarty, "Smart crop recommender system-a machine learning approach," in *2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, 2022, pp. 494–499.

[15] S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA, and K. SHAURYA, "Crop recommender system using machine learning approach," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1066–1071.