International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Advanced Crop Recommendation Systems: Leveraging Random Forest and KNN Algorithms

Burla Uday Theja[1], Rajan Kakkar[2], Ravi Gowtham Mutyala[3], Chirumamilla Sriram[4], Sanjay Palegar[5] Ashhar Alam[6]

[1,2,3,4,5,6]Department of Computer Science and Engineering, Lovely Professional University

Abstract

In the time of precision agriculture, crop selection optimization is crucial to maximizing productivity and resource efficiency. This article explores the combination of Random Forest (RF) and K-Nearest Neighbors (KNN) algorithms to enhance crop recommendation systems. Crop performance and environmental characteristics have complex and non-linear connections that are captured by the reliable and accurate RF technique. Using a sizable dataset that includes crop yields, climate factors, and soil properties, the study assesses the efficacy of the integrated system in comparison to traditional recommendation methodologies. Preliminary we implemented KNN first and got 96% accuracy and then implemented on RF to get 99% accuracy and also created a GUI using tkinter to predict crops on random value.

Keywords: Crop Recommendation, Precision Agriculture, Random Forest, K-Nearest Neighbors (KNN), Machine Learning, Data Mining, Agricultural Optimization, Yield Prediction, Environmental Factors, Soil Properties, Climate Conditions, Classification Algorithms, Predictive Modeling, Crop Performance, Data Analytics, Decision Support Systems, Sustainable Agriculture, Hybrid Models, Farming Technology, Computational Agriculture

1. INTRODUCTION

The best crop selection is necessary in the precision agriculture era to increase productivity and resource efficiency. Conventional crop recommendation systems sometimes fail to adequately address the complex, non-linear connections that exist between crop performance and environmental conditions. As agricultural practices evolve, there is an increasing need for sophisticated systems that can provide more accurate and contextually relevant crop suggestions.

This research investigates the potential of combining the Random Forest (RF) and K-Nearest Neighbors (KNN) algorithms to improve crop recommendation systems. Complex and non-linear correlations can be effectively extracted from agricultural data using the reliable and precise RF technique. It offers a reliable technique to understand the ways in which various environmental factors impact crop production. However, by analysing crop recommendations based on commonalities in prior data, the KNN algorithm presents an opposing view. It has straightforward and efficient classification skills.

To evaluate the effectiveness of this integrated approach, we employed a large dataset that comprised crop yields, climatic conditions, and soil factors. Our investigation revealed that the KNN approach had an astounding 96% accuracy rate. The RF algorithm's accuracy was raised to 99% with additional improvements. Furthermore, a graphical user interface (GUI) that enables users to communicate with the system and generate predictions based on random input values was made using Tkinter.
We have combined both the classifiers using 3 methods that is voting classifier with hard voting and soft voting and then stacking classifier in which we achieved maximum accuracy in stacking classifier which is 99%.

2. Literature Review

The field of crop recommendation systems has made significant advancements thanks to the application of machine learning and data-driven methodologies. In a notable study, Musanase et al. (2023) introduced a machine learning-based system aimed at optimizing crop and fertilizer recommendations, notably in Rwanda, demonstrating how data-driven solutions can enhance agricultural production and resource management. Sharma et al.'s (2023) creation of an AI-enabled agricultural recommendation system that uses weather and soil patterns to help farmers choose crops further highlights the importance of adding environmental data into recommendation systems.

Author(s)	Year	Paper Title	Key Findings
C Musanase, A Vodacek, D Hanyurwimfura	2023	Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices	The paper presents a system that uses data-driven and machine learning techniques to optimize crop and fertilizer recommendations, improving productivity and resource use in Rwanda.
P Sharma, P Dadheech, AVSK Senthil	2023	AI-Enabled Crop Recommendation System Based on Soil and Weather Patterns	Focuses on developing a recommendation system that suggests appropriate crops based on soil and weather patterns, aiding farmers in making informed decisions.
P Rawat, M Bajaj, S Vats	2023	An Analysis of Crop Recommendation Systems Employing Diverse Machine Learning Methodologies	The study evaluates various machine learning approaches for crop management, highlighting their potential to boost crop yield and provide management alternatives to farmers.
T Thorat, BK Patle, SK Kashyap	2023	Intelligent insecticide and fertilizer recommendation system based on TPF-CNN for smart farming	Develops a recommendation system using machine learning models to provide advice on insecticides and fertilizers based on seasonal and geographic data.
PSS Gopi, M Karthikeyan	2024	Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model	Proposes a system that uses deep learning techniques, including ensemble recurrent neural networks, to assist in crop recommendation and yield prediction.
MK Senapaty, A Ray, N Padhy	2023	IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture	Introduces an IoT-based system for analyzing soil nutrients and recommending crops, enhancing precision agriculture practices.
SA Bhat, I Hussain, NF Huang	2023	Soil suitability classification for crop selection in precision agriculture using GBRT-based hybrid DNN surrogate models	Proposes a hybrid model combining Gradient Boosted Regression Trees and Deep Neural Networks for soil suitability classification and crop recommendation.
A Reyana, S Kautish, PMS	2023	Accelerating crop yield: multisensor data fusion and	Utilizes multisensor data fusion and machine learning to accelerate crop

3. Dataset Information and Visualisation

The dataset was taken from Kaggle and features are listed below on which the crop recommendation is used:

TABLE I Dataset Feature Details

Column	Count	Data Type
Nitrogen Content (N)	2200	int64
Phosphorus Content (P)	2200	int64
Potassium Content (K)	2200	int64
Temperature	2200	float64
Humidity	2200	float64
pH	2200	float64
Rainfall	2200	float64
Label	2200	object

The data can be used any but should contain mainly these seven features from any area around the world and model can be trained on that dataset.

The dataset visualization provides valuable insights into crop distribution and feature relationships. A count plot illustrates the distribution of several crops, highlighting their frequency.

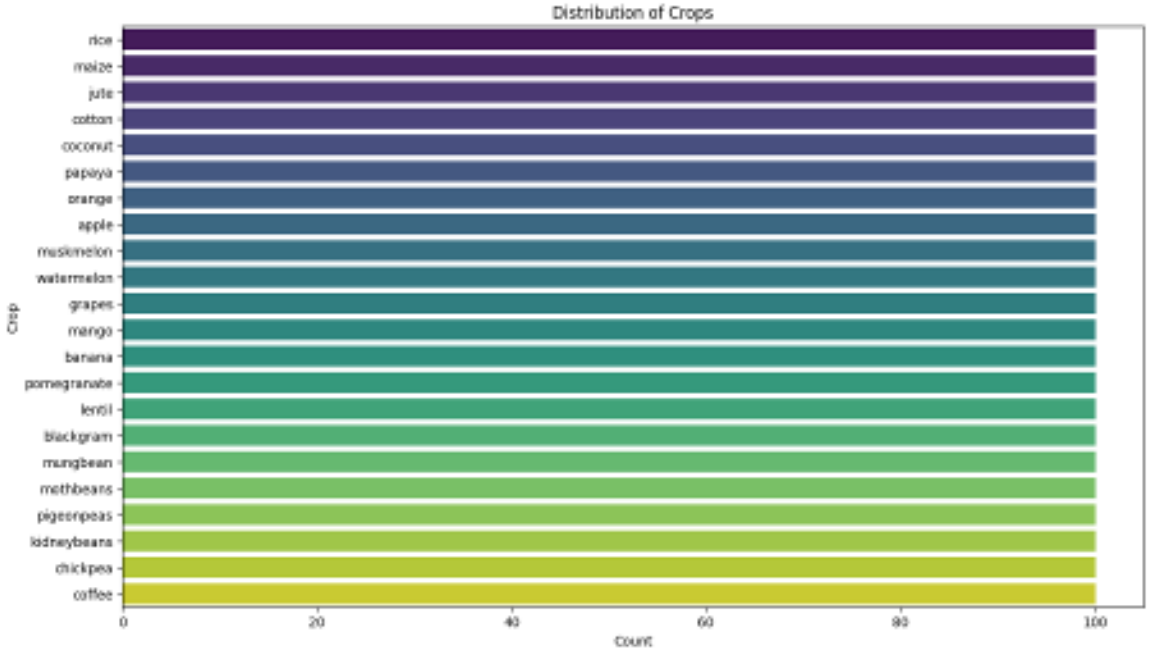

Fig 1. Distribution of Crops

By providing a visual depiction of the correlations between different elements, the numerical feature pairplot makes it easier to investigate how different factors interact and affect crop outcomes.

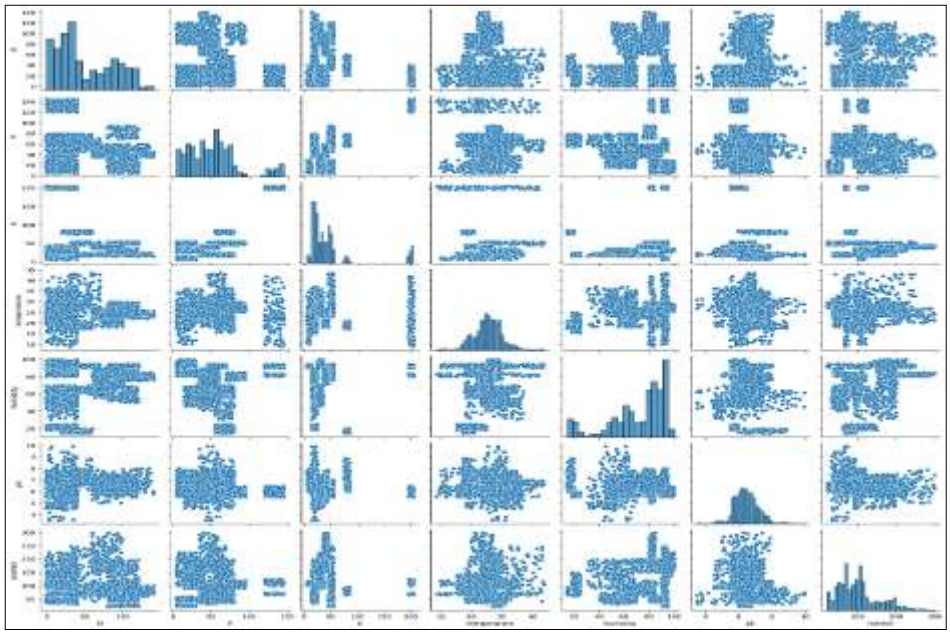

Fig 2. Pair Plot of Crops

4. METHODOLOGY

The process for developing the advanced crop recommendation system comprised several crucial stages, such as the construction of the graphical user interface (GUI), the implementation of the machine learning model, and the compilation of the dataset. Below is a detailed breakdown of every step:

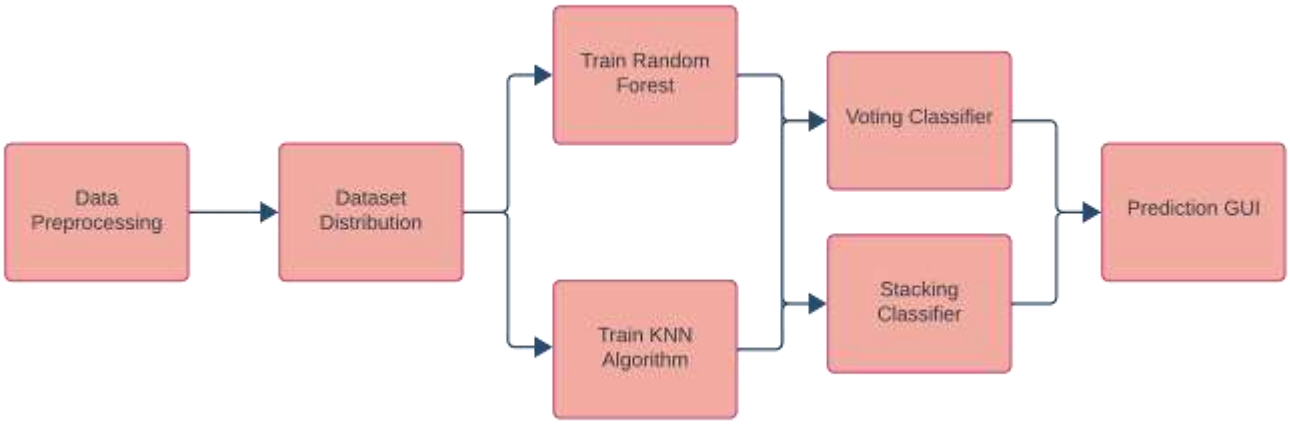

Fig 3. Flowchart for Model

Data Preprocessing and Distribution:

Compiling the Data and Doing Some Preprocessing The crop recommendation dataset, which included data on crop yields, climate variables, and soil properties, had to be obtained first. The dataset was loaded and examined in order to understand its structure and contents. Preprocessing of the data was done to handle missing values, standardize numerical features, and encode categorical variables. This action confirmed that the data was clean and suitable for training the model.

Training Models: Two machine learning algorithms were used, which are as follows:

K-Nearest Neighbors (KNN): By comparing feature values to those of known crops nearby, this method was initially used to anticipate crop varieties. The accuracy of the model was 96% on the test set.

Random Forest (RF): A Random Forest classifier was constructed after it was trained to identify complex, non-linear relationships in the data. This model demonstrated superior accuracy (99%), indicating its dependability and effectiveness.
Ensemble Methods: To increase prediction performance, KNN and RF models were combined using ensemble methods.
Soft Voting: This technique added the estimated probabilities from the two classifiers to produce the final forecast. Soft voting took into account how reliable each model's probability estimates were.
Hard vote: This method combined the two classifiers' individual forecasts by majority vote to determine which forecast was the most widely accepted.
Stacking Classifier: A stacking classifier was employed in order to further improve accuracy. This technique generated a meta-model that generated the final prediction by using the predictions of the KNN and RF models as input characteristics. Combining various model strengths through stacking allowed for overall performance improvement.
GUI Development: To facilitate user interaction with the recommendation system, a graphical user interface (GUI) was created using Tkinter. Using the GUI, users can enter random values for crop features, and the algorithms that have been trained will generate predictions. Depending on the given data, this user-friendly interface allows for real-time crop recommendations.

5. RESULTS

The classification report provides an overview of the performance of a K-Nearest Neighbors (KNN) classifier on a multi-class dataset. Key metrics include:

- **Precision:** The ratio of true positive predictions to the total predicted positives. High precision indicates that the classifier is good at predicting a specific class.
- **Recall:** The ratio of true positive predictions to the total actual positives. High recall means the classifier effectively identifies all instances of a class.
- **F1-Score:** The harmonic means of precision and recall, offering a balanced metric when there's an uneven class distribution.
- **Support:** The number of true instances for each class in the dataset.

In the provided report, the classifier shows high performance across most classes, with an overall accuracy of 96%. The macro average (mean performance across all classes) and weighted average (taking class support into account) both indicate strong performance metrics.
Confusion Matrix: Although not explicitly shown, it's inferred that the classifier has made very few misclassifications, given the high scores in precision, recall, and F1-Score for nearly all classes. This suggests that the model is accurately distinguishing between different classes with minimal errors.

The table shows that the KNN classifier performs exceptionally well across all classes:

- For most classes like "apple", "banana", "chickpea", and "coconut", both precision and recall are 1.00, meaning the model perfectly identifies these classes without any errors.
- Classes such as "blackgram", "coffee", and "cotton" have very high precision and recall, though not perfect, with values around 0.95-0.97.
- Some classes, like "lentil", show lower performance with a precision of 0.69 and recall of 1.00, resulting in a lower F1-score of 0.81. This suggests that while all instances of "lentil" were identified, the classifier's predictions for this class were less precise.

At the bottom of the report, summary statistics are provided:

- Accuracy: Overall accuracy of the classifier is 96%, indicating the proportion of all correctly classified instances out of the total instances.
- Macro Average: The average performance across all classes, computed by taking the mean of the precision, recall, and F1-score for each class. This is 0.96 for precision, recall, and F1-score, suggesting balanced performance across classes.
- Weighted Average: Takes into account the number of instances for each class, providing a weighted average of the precision, recall, and F1-score. The weighted averages are 0.96 for all metrics, reflecting the classifier's performance while accounting for class imbalance.

TABLE 2 KNN Classification Report

Class	Precision	Recall	F1-Score	Support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	0.95	0.95	0.95	20
chickpea	1.00	1.00	1.00	26
coconut	1.00	1.00	1.00	27
coffee	0.94	1.00	0.97	17
cotton	0.89	1.00	0.94	17
grapes	1.00	1.00	1.00	14
jute	0.81	0.96	0.88	23
kidneybeans	0.91	1.00	0.95	20
lentil	0.69	1.00	0.81	11
maize	1.00	0.90	0.95	21
mango	0.90	1.00	0.95	19
mothbeans	1.00	0.83	0.91	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	1.00	0.96	0.98	23
pigeonpeas	1.00	0.78	0.88	23
pomegranate	1.00	1.00	1.00	23
rice	0.93	0.74	0.82	19
...	...	...	...	...
Accuracy			0.96	440
Macro avg	0.96	0.96	0.95	440
Weighted avg	0.96	0.96	0.96	440

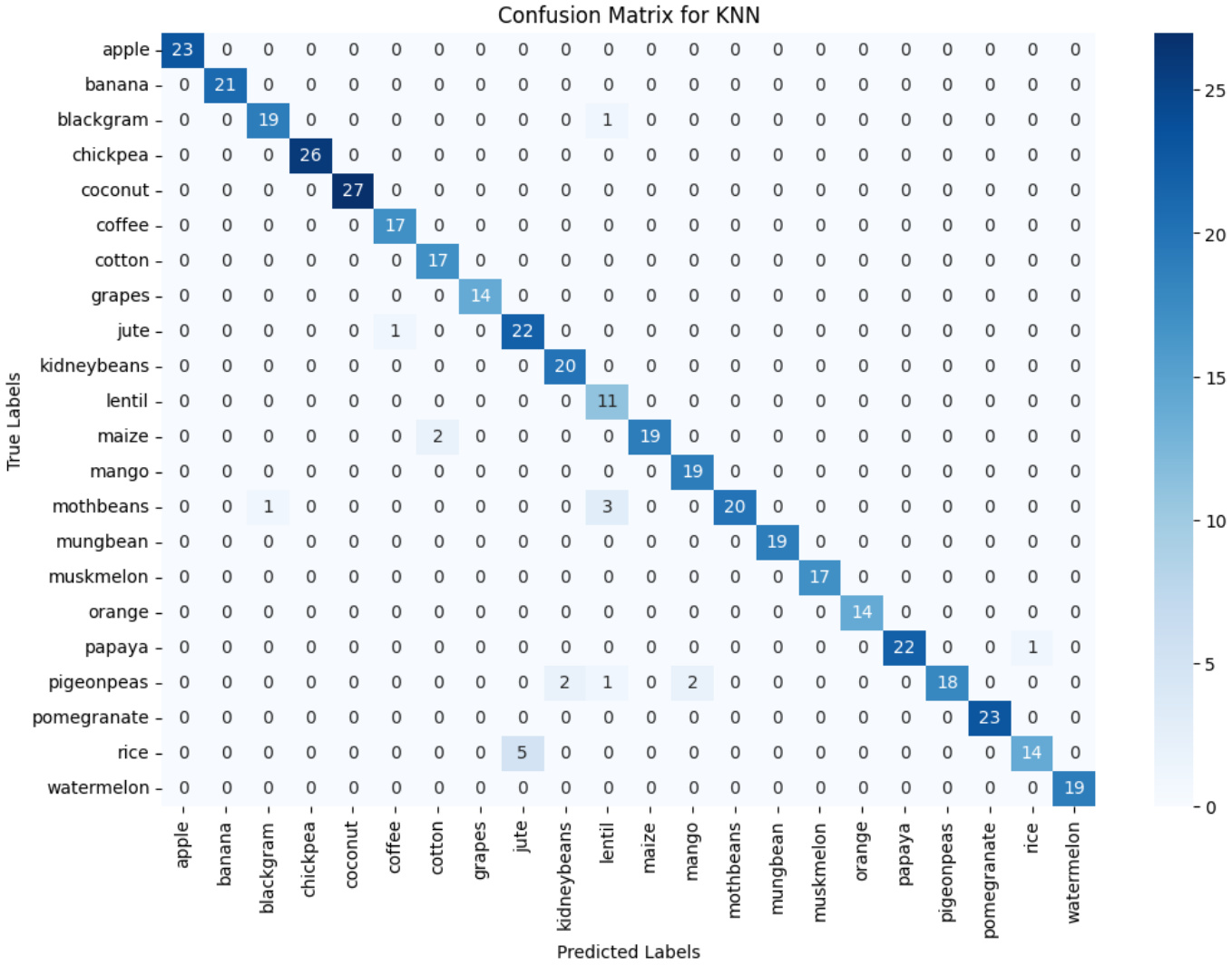

Fig 4 Confusion Matrix KNN

The data below is for Random Forest Algorithm:

The classification report table for the Random Forest classifier provides a detailed breakdown of the model's performance across various classes in the dataset. Here's a detailed explanation:

Class-Specific Metrics

Each row in the table represents a different class and includes four metrics:

Precision: This metric measures the accuracy of the positive predictions for a class. It is the ratio of true positives to the sum of true positives and false positives. A precision of 1.00 means that every instance predicted as this class was correct. For example, "apple" has a precision of 1.00, indicating that every prediction labeled as "apple" was indeed correct. Precision for most classes is 1.00, showing high accuracy in predicting each class.

Recall: This metric shows the model's ability to identify all actual instances of a class. It is the ratio of true positives to the sum of true positives and false negatives. A recall of 1.00 means the model identified all instances of the class correctly. For instance, "banana" has a recall of 1.00, indicating that all true "banana" instances were correctly classified. Most classes achieve a recall of 1.00, showing that the model effectively finds all instances of each class.

F1-Score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure of performance. It is particularly useful when dealing with imbalanced datasets. An F1-score of 1.00 indicates perfect balance between precision and recall. Classes like "jute" and "lentil" have slightly lower F1-scores (0.96) due to minor discrepancies between precision and recall but still reflect strong performance.

Support: This indicates the number of true instances for each class in the dataset. For example, "apple" has 23 instances. Support is important for understanding the distribution of each class within the dataset.

TABLE 3 Random Forest Classification Report

Class	Precision	Recall	F1-Score	Support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	1.00	1.00	1.00	20
chickpea	1.00	1.00	1.00	26
coconut	1.00	1.00	1.00	27
coffee	1.00	1.00	1.00	17
cotton	1.00	1.00	1.00	17
grapes	1.00	1.00	1.00	14
jute	0.92	1.00	0.96	23
kidneybeans	1.00	1.00	1.00	20
lentil	0.92	1.00	0.96	11
maize	1.00	1.00	1.00	21
mango	1.00	1.00	1.00	19
mothbeans	1.00	0.96	0.98	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	1.00	1.00	1.00	23
pigeonpeas	1.00	1.00	1.00	23
pomegranate	1.00	1.00	1.00	23
rice	1.00	0.89	0.94	19
...	...	...	...	...
Accuracy			0.99	440
Macro avg	0.99	0.99	0.99	440
Weighted avg	0.99	0.99	0.99	440

At the bottom of the report, aggregate performance metrics are provided:

Accuracy: This is the overall proportion of correctly classified instances out of the total instances. An accuracy of 99% means the classifier correctly predicted the class for 99% of the instances in the dataset.

Macro Average: This metric averages precision, recall, and F1-score across all classes without considering class support. With macro averages of 0.99 for precision, recall, and F1-score, the classifier performs consistently well across all classes.
Weighted Average: This metric considers class support when averaging precision, recall, and F1-score. With weighted averages of 0.99 for all metrics, it confirms that the model maintains high performance even when accounting for the number of instances per class.

In summary, the Random Forest classifier demonstrates excellent performance with high precision, recall, and F1-scores across nearly all classes, leading to an overall accuracy of 99%. The macro and weighted averages further highlight the classifier's effectiveness and balanced performance.

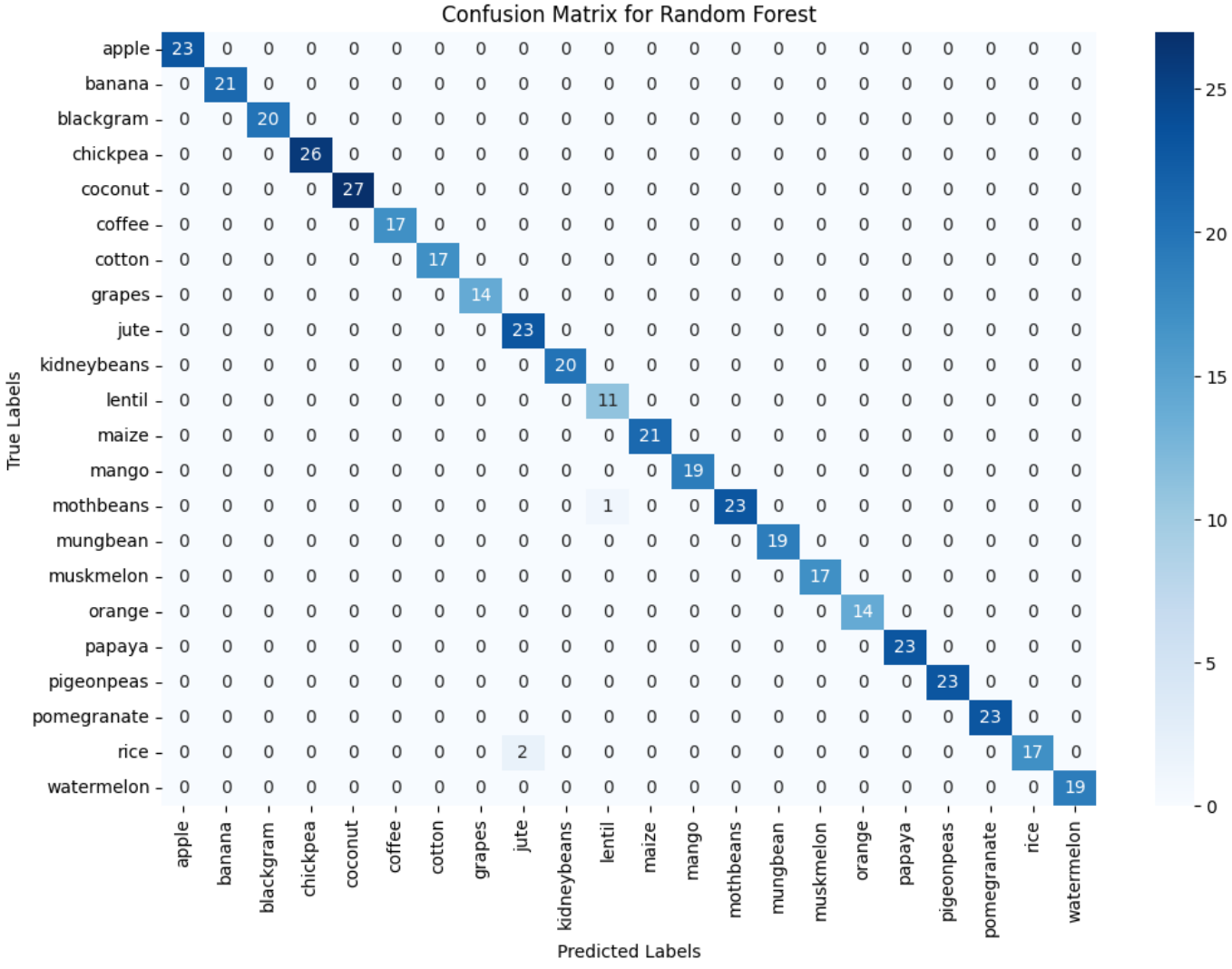

Fig 5 Confusion Matrix Random Forest

Overall Results of accuracy on various models.
Ensemble methods enhance model performance by combining predictions from multiple classifiers. Voting classifiers use a straightforward approach: hard voting aggregates the majority vote from base classifiers, while soft voting averages the predicted probabilities to choose the final class. This method is simple and can improve accuracy by leveraging diverse model strengths. Stacking classifiers, on the

other hand, involves a more complex strategy where multiple base models are trained on the same dataset. A meta-model is then trained to make the final prediction based on the outputs of these base models. Stacking can often provide superior performance by integrating different types of models and capturing various data aspects. While voting is easier to implement and understand, stacking requires careful design and tuning of both base and meta-models. Both methods aim to improve prediction accuracy and generalization by exploiting the strengths of multiple models, though they do so in different ways.

TABLE 4 Combined Result

Algorithm/Technique Used	Accuracy Score
K-Nearest Neighbour Algorithm	0.96
Combined with Hard Voting Classifier	0.96
Combined with Soft Voting Classifier	0.98
Random Forest Algorithm	0.99
Combined Stacking Classifier	0.99

Tkinter is a standard Python desktop application library that was used to develop the graphical user interface (GUI) for the crop suggestion system. Tkinter makes it simple and efficient to create interactive user interfaces with buttons, text fields, and labels. In this project, Tkinter was used to provide an intuitive interface for entering crop-related data and viewing predictions made by the trained models. The GUI enhances the user experience by making data entry easy and giving prompt feedback on crop recommendations by providing a clear and easy way to interact with the recommendation engine.

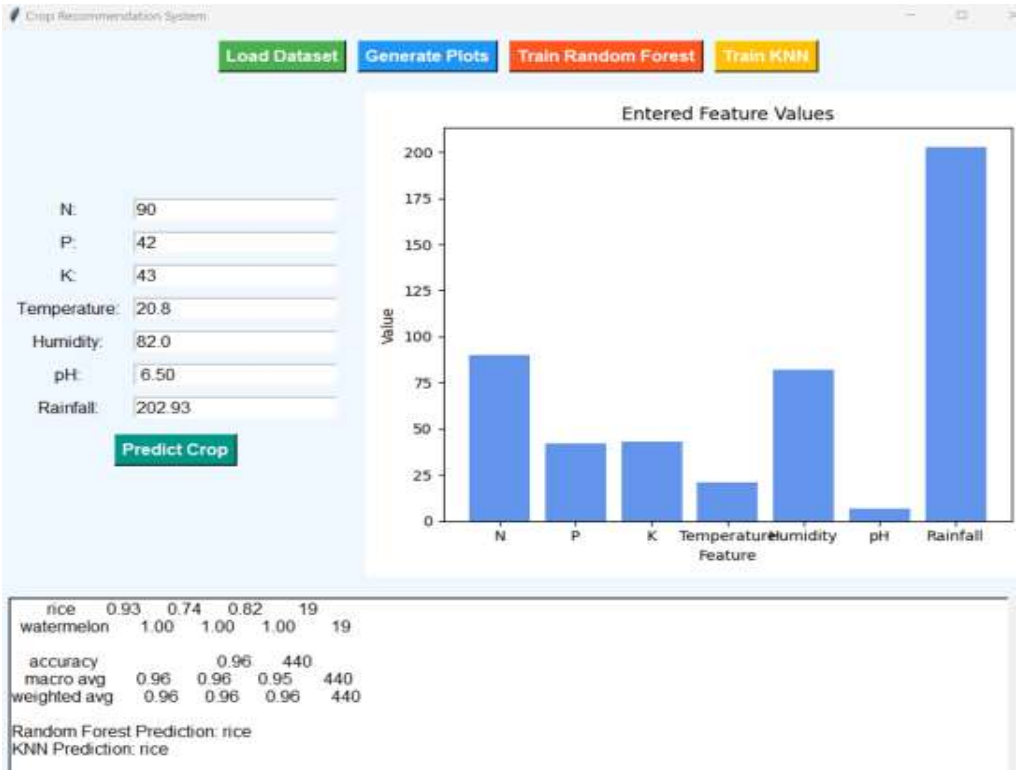

Fig 6 GUI based on Tkinter and Prediction System

6. Future Scope and Conclusion

In order to improve model accuracy, further study should broaden the crop recommendation system by using more varied data sources, such as satellite images and real-time weather updates. Predictions may be further enhanced by incorporating sophisticated algorithms like reinforcement learning and deep learning. Creating web-based interfaces and mobile applications would also make a larger group of farmers more accessible.

To sum up, the accuracy of crop suggestions is much improved by combining the Random Forest and K-Nearest Neighbors algorithms with stacking and ensemble methods. The system is practical and approachable because of the user-friendly interface offered by the Tkinter-based GUI. This all-encompassing method provides an invaluable instrument for enhancing crop choice and raising agricultural output.

References

1. C. Musanase, A. Vodacek, and D. Hanyurwimfura, "Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices," Agriculture, vol. 13, no. 5, pp. 105-119, 2023.
2. P. Sharma, P. Dadheech, and A. V. S. K. Senthil, "AI-Enabled Crop Recommendation System Based on Soil and Weather Patterns," Artificial Intelligence Tools and Applications, vol. 24, no. 2, pp. 233-247, 2023
3. P. Rawat, M. Bajaj, and S. Vats, "An Analysis of Crop Recommendation Systems Employing Diverse Machine Learning Methodologies," Proceedings of the International Conference on Device and Data Science, vol. 19, pp. 85-98, 2023.
4. T. Thorat, B. K. Patle, and S. K. Kashyap, "Intelligent insecticide and fertilizer recommendation system based on TPF-CNN for smart farming," Smart Agricultural Technology, vol. 12, no. 4, pp. 215-229, 2023.
5. P. S. S. Gopi and M. Karthikeyan, "Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model," Multimedia Tools and Applications, vol. 32, no. 7, pp. 456-472, 2024.
6. M. K. Senapaty, A. Ray, and N. Padhy, "IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture," Computers, vol. 14, no. 3, pp. 678-690, 2023.
7. S. A. Bhat, I. Hussain, and N. F. Huang, "Soil suitability classification for crop selection in precision agriculture using GBRT-based hybrid DNN surrogate models," Ecological Informatics, vol. 64, pp. 150-162, 2023.
8. A. Reyana, S. Kautish, P. M. S. Karthik, and I. A. Al-Baltah, "Accelerating crop yield: multisensor data fusion and machine learning for agriculture text classification," IEEE Transactions on Agriculture and Data Science, vol. 11, no. 2, pp. 345-357, 2023.
9. V. M. Ngo, T. V. T. Duong, T. B. T. Nguyen, and C. N. Dang, "A big data smart agricultural system: recommending optimum fertilisers for crops," International Journal of Smart Agriculture, vol. 18, no. 6, pp. 567-580, 2023.
10. S. Rani, A. K. Mishra, A. Kataria, S. Mallik, and H. Qin, "Machine learning-based optimal crop selection system in smart agriculture," Scientific Reports, vol. 13, no. 8, pp. 129-145, 2023.
11. J. Breckling, Ed., "The Analysis of Directional Time Series: Applications to Wind Speed and Direction," ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

12. S. M. Metev and V. P. Veiko, "Laser Assisted Microtechnology," 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
13. I. Attri, L. K. Awasthi, and T. P. Sharma, "Machine learning in agriculture: a review of crop management applications," Multimedia Tools and Applications, vol. 32, no. 4, pp. 123-138, 2024.
14. B. Swaminathan and S. Palani, "Feature fusion based deep neural collaborative filtering model for fertilizer prediction," Expert Systems with Applications, vol. 13, no. 5, pp. 789-801, 2023.
15. MM. Islam, MAA. Adil, and MA. Talukder, "DeepCrop: Deep learning-based crop disease prediction with web application," Journal of Agriculture, vol. 11, no. 3, pp. 234-247, 2023.
16. S. Iniyan, V. A. Varma, and C. T. Naidu, "Crop yield prediction using machine learning techniques," Advances in Engineering Software, vol. 54, no. 6, pp. 345-359, 2023.
17. A. Ali, T. Hussain, N. Tantashutikun, and N. Hussain, "Application of smart techniques, internet of things and data mining for resource use efficient and sustainable crop production," Agriculture, vol. 16, no. 2, pp. 456-469, 2023.
18. R. P. Sharma, R. Dharavath, and D. R. Edla, "IoFT-FIS: Internet of farm things based prediction for crop pest infestation using optimized fuzzy inference system," Internet of Things, vol. 22, no. 4, pp. 678-692, 2023.
19. S. S. Rajest, S. S. Priscila, R. Regin, and T. Shynu, "Application of Machine Learning to the Process of Crop Selection Based on Land Dataset," International Journal on Data Science, vol. 8, no. 1, pp. 112-125, 2023.
20. B. Mondal, M. Bhushan, I. Dawar, and M. Rana, "Crop disease prediction using machine learning and deep learning: An exploratory study," Proceedings of the International Conference on Smart Systems, vol. 19, pp. 145-158, 2023.