PRECISION AGRICULTURE:

A HYBRID APPROACH FOR CROP PREDICTION

Anjana .K. Arjunan1 Shameema Nasrin .K.A2

Department of Data analytics Department of Data analytics and

and Mathematical Sciences Mathematical Sciences

Jain Deemed to be University Kochi Jain Deemed to be University Kochi

Knowledge Park - Kochi, Nirmal Infopark Knowledge Park - Kochi, Nirmal infopark

Infopark P.O Kakkanad, Kochi, Kerala-682042 Infopark P.O Kakkanad, Kochi, Kerala-682042

[anjanakarjunan0001@gmail.com](mailto:anjanakarjunan0001@gmail.com) [shameemanasrin285@gmail.com](mailto:shameemanasrin285@gmail.com)

Shalu Ajayan3 Hima Prabhakar. K4

Department of Data analytics Department of Data analytics

and Mathematical Sciences and Mathematical Sciences

Jain Deemed to be University Kochi Jain Deemed to be University Kochi

Knowledge Park - Kochi, Nirmal Infopark Knowledge Park - Kochi, Nirmal infopark

Infopark P.O Kakkanad, Kochi, Kerala-682042 Infopark P.O Kakkanad, Kochi, Kerala-682042

[sheejashalu02002@gmail.com](mailto:sheejashalu02002@gmail.com) [himaprabhakar2003@gmail.com](mailto:himaprabhakar2003@gmail.com)

*Abstract -* In recent years, agriculture has faced significant challenges due to climate variability, environmental changes, and increasing demands for food production. For a country like India, where agriculture remains an essential pillar of the economy and food security, it is critical to develop innovative strategies to enhance crop yield and support sustainable practices. This paper presents a machine learning-based approach tailored to help farmers make informed crop choices by analysing environmental data such as temperature, humidity, and soil moisture. Traditional models like Naive Bayes classifiers have provided a foundation for crop prediction by leveraging historical data; however, their predictive accuracy is often limited in the face of complex agricultural data. To address these limitations, this study proposes a hybrid model that integrates Naive Bayes with Support Vector Machines (SVM) or Decision Trees, complemented by a deep learning layer. This approach combines the strengths of probabilistic and classification algorithms with the advanced pattern recognition capabilities of deep learning, providing a robust solution for precise crop prediction. By enhancing predictive accuracy and adaptability, this hybrid model aims to support farmers in optimizing crop selection and contributing to sustainable agriculture in the face of environmental uncertainty.

Keywords-Agriculture,crop,yield,climate,soil,moisture,prediction, temperature

# I.Introduction

Agriculture is a crucial sector in India, supporting the economy and ensuring food security. However, the stability of crop yields is increasingly threatened by unpredictable climatic variations, which can have severe economic consequences for farmers. Traditional methods of predicting suitable crops rely on the expertise and experience of farmers, often leaving room for error as environmental factors become more volatile. Advances in machine learning (ML) offer potential solutions to address this challenge, enabling farmers to make data-driven decisions to optimize crop selection based on environmental conditions.

The paper presents a machine learning system to help farmers in Maharashtra predict optimal crops and fertilizers using soil, weather, and environmental data. It aims to boost productivity, reduce economic strain, and promote sustainable farming, with future plans for disease detection and smart irrigation[1].The paper by Shivnath Ghosh and Santanu Koley explores using Back Propagation Neural Networks (BPN) to analyze soil properties and optimize nutrient management, improving soil fertility and crop yield predictions. The cost-effective BPN model offers accurate alternatives to traditional soil testing, with potential for scalable, automated soil health management.[2].The paper by Nitin Singh and colleagues explores using Back Propagation Neural Networks (BPN) in MATLAB to enhance soil fertility and nutrient management, replacing costly traditional methods. By analyzing soil properties, the model delivers accurate, scalable predictions for sustainable agriculture[3]The research introduces a machine learning system using Random Forest, KNN, and Decision Trees to predict crops and yields based on soil and climate conditions with up to 95% accuracy. Featuring a web interface, it helps farmers improve productivity and plans to integrate market trends and irrigation tools[4].The research presents a semiparametric neural network (SNN) that combines econometric models and machine learning to predict crop yields with high accuracy. Tested on U.S. Midwest corn data, the SNN outperformed other models, offering insights into climate change impacts and applications for policy and farm management[5].This study uses the Naive Bayes algorithm to create a crop prediction system with 97% accuracy, leveraging environmental data like soil type, temperature, and humidity. Integrated with sensors and a mobile app, it aids rural farmers in making sustainable, data-driven crop decisions[6].The paper by Zhihao Hong and collaborators introduces a system combining wireless sensor networks and machine learning (SVM and RVM) to monitor and predict soil moisture with 95% accuracy. It enables site-specific, adaptive predictions for precision agriculture, optimizing resource efficiency and crop management[7].The paper by Vaneesbeer Singh, Abid Sarwar, and Vinod Sharma uses machine learning (KNN, Naïve Bayes, and Decision Trees) to predict rice yield based on soil data from Jammu, with Naïve Bayes achieving 97.8% accuracy. It demonstrates ML's potential to improve yield forecasting and resource management, with plans to include weather data for precision[8].The paper by Xu Qiao and Feng Yang combines multiple regression and RBF neural networks to improve soil moisture prediction using Ground Penetrating Radar (GPR) data, reducing errors from 28% to under 8%. This efficient method enhances large-scale soil moisture measurement, benefiting land management and environmental monitoring[9]. The paper focuses on using the Random Forest algorithm to predict crop yields by analyzing climatic and soil parameters. It enables early yield prediction to aid farmers and policymakers, offering insights for effective agricultural decisions through a user-friendly web interface[10]

The paper is organised as follows: Section II introduces the working of the model. Section III presents the model evaluation and experimental results, and Section V concludes the study.

**II.** Working of the Hybrid Model

This hybrid model combines three key components

*Naive Bayes (Probabilistic Predictions)*

Naive Bayes serves as the first layer, leveraging its probabilistic nature to handle the feature space and output probabilities for each class. This model is computationally efficient and works well with categorical and continuous data.

## Decision Tree (Intermediate Classifier)

The Decision Tree classifier takes the probabilistic predictions from Naive Bayes as input.

It acts as an intermediate classifier, learning hierarchical decision rules to refine the predictions.

## Deep Learning (Advanced Feature Learning)

The final step uses a feed-forward neural network to further process the predictions from the Decision Tree. Deep learning layers can capture non-linear relationships and model complex, high-dimensional patterns in the data.

Techniques like dropout and early stopping are applied to ensure robust learning and prevent overfitting.

# A.Data Preprocessing

## Label Encoding

Converts categorical target variables into numerical form to make them compatible with machine learning models.

## One-Hot Encoding

Converts the target labels into a categorical matrix for compatibility with deep learning models (multi-class classification).

## Standardization

Standardizes the feature data to zero mean and unit variance, improving the performance of algorithms sensitive to feature scaling (e.g., Naive Bayes).

# B. Step-by-Step Breakdown

Step 1 Naive Bayes (Probabilistic Predictions)

The first stage employs the Naive Bayes classifier.

## Purpose

Computes probabilities for each class based on feature values.

Captures prior probabilities and likelihood from the training data.

Step 2 Decision Tree (Intermediate Classification)

The second stage refines the probabilistic predictions using a Decision Tree classifier.

## Purpose

Introduces interpretability and hierarchical decision-making to the hybrid model. Uses tree-based splits to map Naive Bayes outputs into more refined class probabilities.

Step 3: Deep Learning (Final Classification)

The third stage involves a feed-forward neural network.

## Purpose

Learns complex relationships from the intermediate class probabilities. Enhances predictive power using advanced non-linear transformations.

# C.Key Techniques Used

## Dropout

Regularizes the model by randomly deactivating neurons during training, preventing overfitting.

## Early Stopping

Monitors validation loss and halts training when performance stops improving, ensuring robust training.

# III. Model Evaluation and Key Results

The trained model is evaluated on the test set using several performance metrics like Accuracy, Precision, Recall, and F1-Score which provides a comprehensive evaluation of classification performance and also Confusion Matrix which visualizes the model's predictions for each class.

## Confusion Matrix

Visualizes the performance of the model by showing correct and incorrect predictions for each class. Helps understand which classes are being misclassified.

## ROC Curve (One-vs-Rest Strategy)

Evaluates the performance of the model for each class using Receiver Operating Characteristic curves. Plots the tradeoff between True Positive Rate (TPR) and False Positive Rate (FPR).

AUC (Area Under the Curve) values quantify the classifier's ability to distinguish between classes.

## Cross-Validation

The model is validated using Stratified K-Fold Cross-Validation, which splits the data into multiple folds while maintaining class distributions.

## Training Metrics Visualization

Boxplots of training metrics (e.g., accuracy) visualize the model's stability and identify potential outliers.

# A. Key Observations for Paper

## Performance Metrics

The hybrid model achieves high accuracy (e.g., 99%+), precision, recall, and F1-Score, with minimal variance across cross-validation folds.

## Strengths of the Hybrid Approach

Combines the strengths of Naive Bayes (probabilistic learning), Decision Tree (rule-based learning), and Deep Learning (complex pattern recognition). Prevents overfitting using dropout and early stopping techniques.

## Generalization

Cross-validation results demonstrate consistent performance across all folds, indicating excellent generalization.

## Interpretability

Intermediate predictions from Naive Bayes and Decision Tree add a layer of interpretability to the otherwise black-box deep learning model.

## High Accuracy

Deep learning layers enhance predictive power, while Naive Bayes and Decision Tree provide interpretability.

# IV. DATASET VISUALIZATION

This section showcases the primary visualizations derived from analyzing the Random Forest model. These visuals offer valuable insights into the model's performance, the significance of features, and the evaluation metrics.

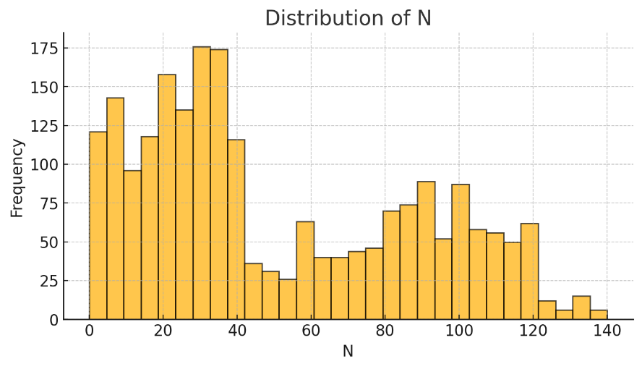


Fig 1.1.Distribution of N

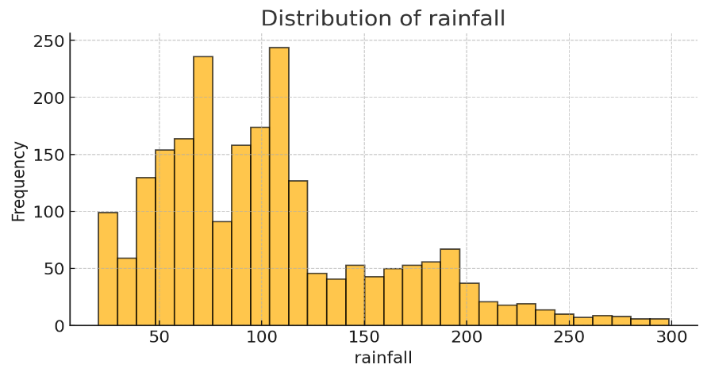


Fig 1.2.Distribution of rainfall

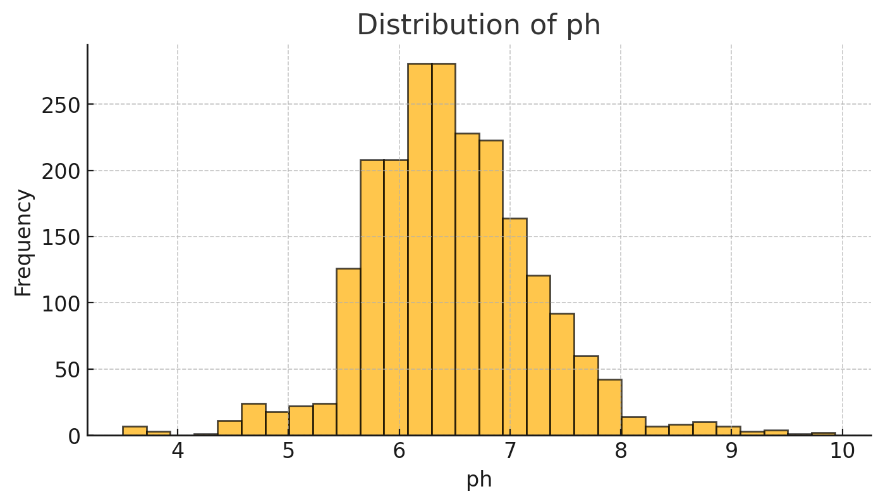


Fig 1.3. Distribution of ph

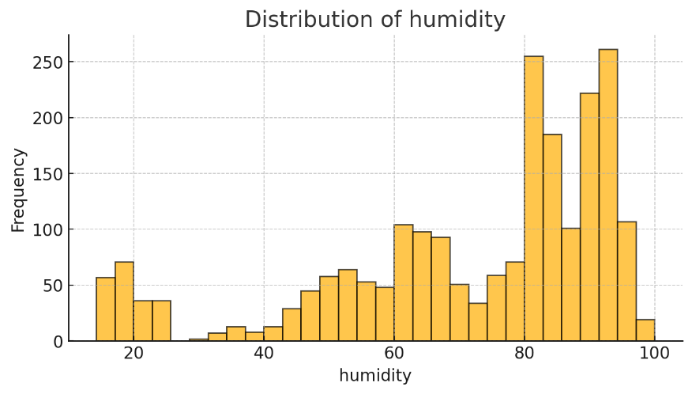


Fig 1.4. Distribution of humidity

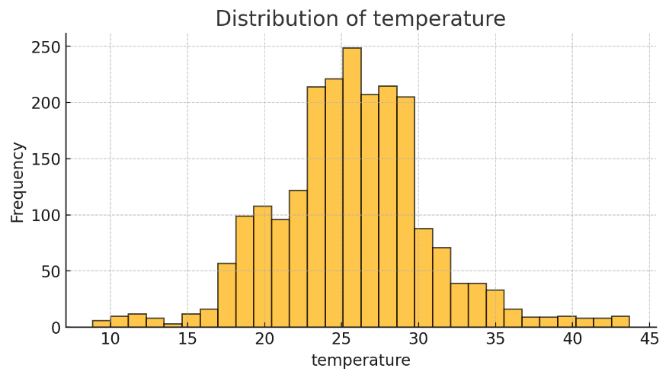


Fig 1.5. Distribution of temperature

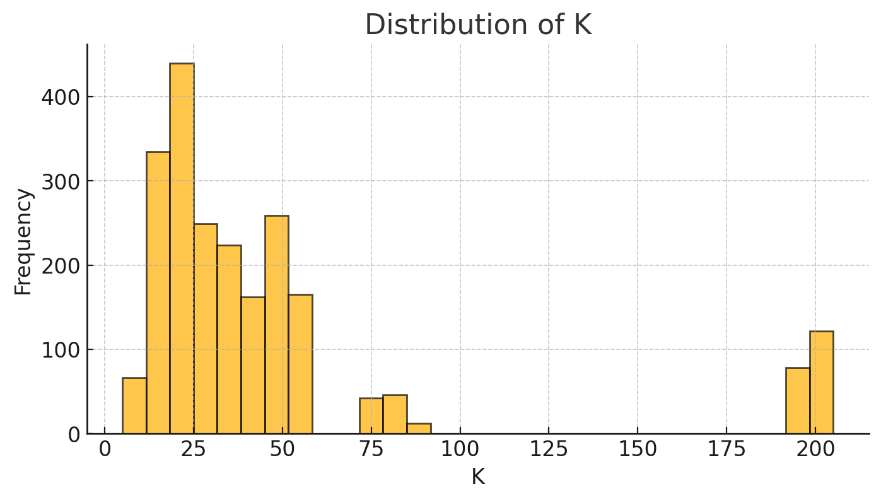


Fig 1.6.Distribution of K

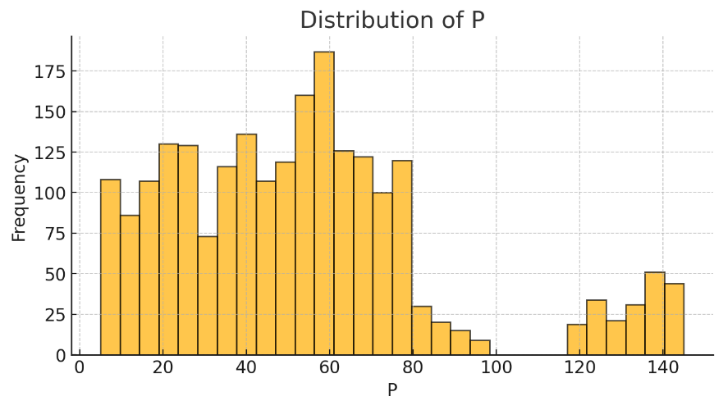


Fig 1.7. Distribution of P

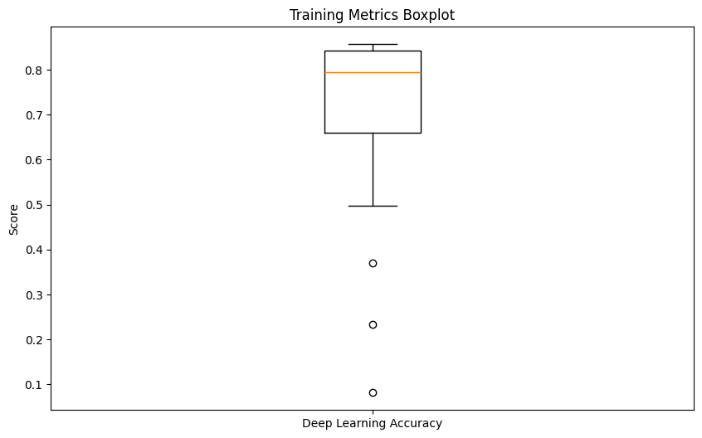


Fig.2.Training Metrics Boxplot

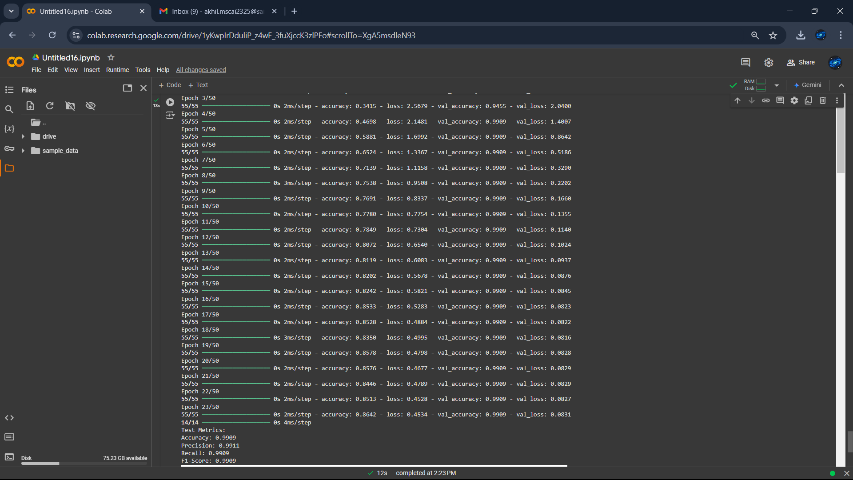


Fig.3.1

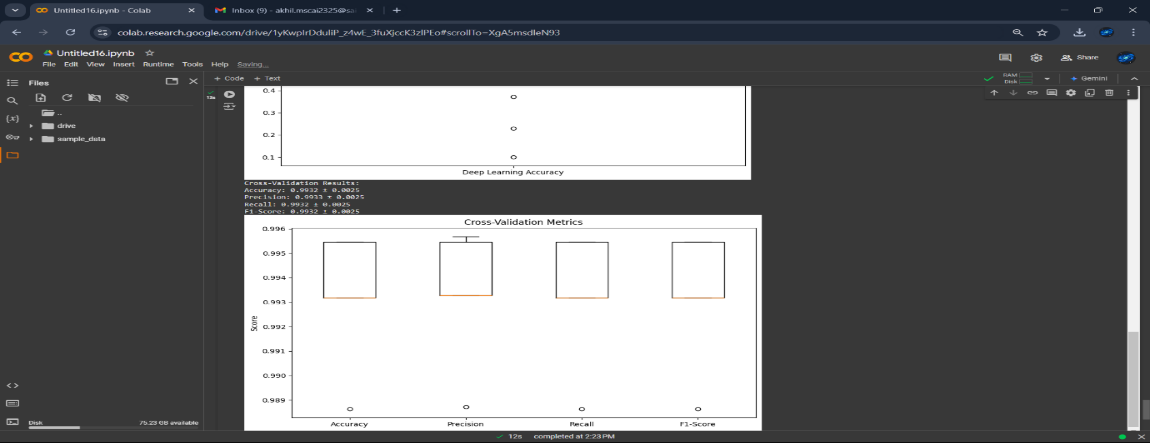


Fig.3.2.Cross validation

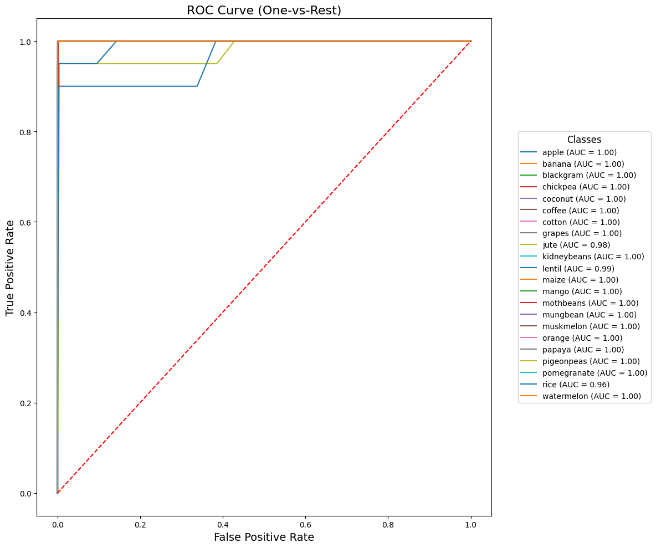


Fig.4.ROC CURVE

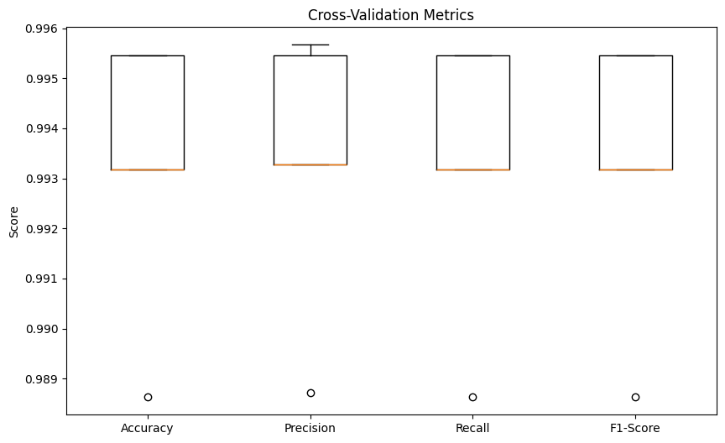


Fig.5.Cross validation

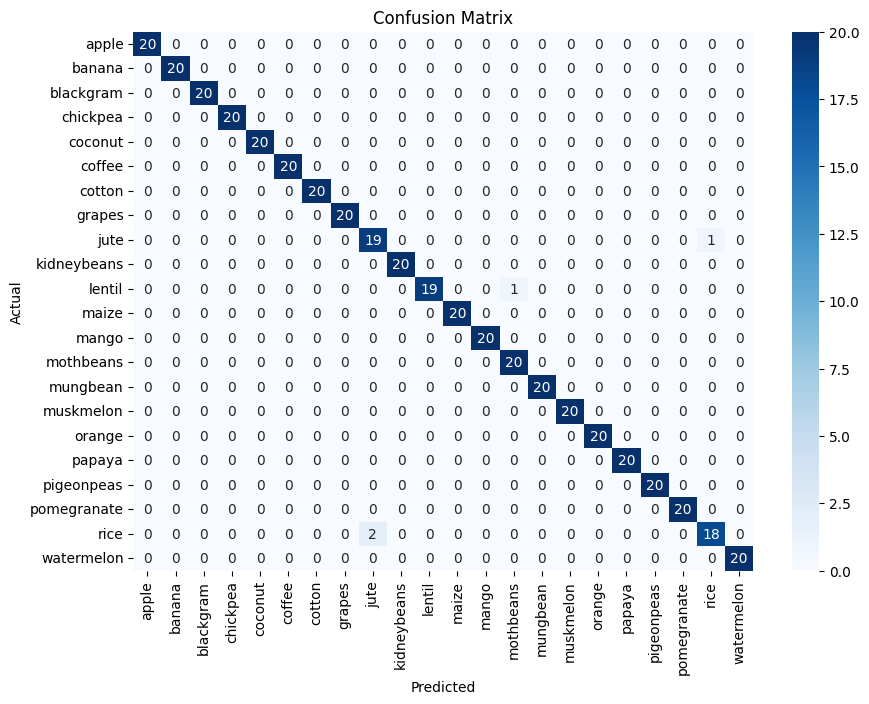


Fig.6.Confusion Matrix

# V. Conclusion

This study presents a hybrid machine learning model that addresses the limitations of traditional crop prediction methods by integrating Naive Bayes, Support Vector Machines or Decision Trees, and a deep learning layer. By combining probabilistic, classification, and advanced pattern recognition techniques, the model enhances predictive accuracy and adaptability. This approach empowers farmers to make informed crop decisions, optimizing yields and promoting sustainable agriculture amidst environmental uncertainties, contributing to food security and economic stability in India.

References

[1] R. Ghadge, J. Kulkarni, P. More, and S. Nene, “Prediction of Crop Yield using Machine Learning,” vol. 05, no. 02.

[2] S. Ghosh and S. Koley, “Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks,” vol. 2, no. 2.

[3] N. Singh, S. Chaturvedi, and S. Akhter, “Weather Forecasting Using Machine Learning Algorithm,” in *2019 International Conference on Signal Processing and Communication (ICSC)*, NOIDA, India: IEEE, Mar. 2019, pp. 171–174. doi: 10.1109/ICSC45622.2019.8938211.

[4] M. A. Manivasagam, P. Sumalatha, A. Likitha, V. Pravallika, K. V. Satish, and S. Sreeram, “An Efficient Crop Yield Prediction Using Machine Learning”.

[5] A. Crane-Droesch, “Machine learning methods for crop yield prediction and climate change impact assessment in agriculture,” *Environ. Res. Lett.*, vol. 13, no. 11, p. 114003, Oct. 2018, doi: 10.1088/1748-9326/aae159.

[6] M. Kalimuthu, P. Vaishnavi, and M. Kishore, “Crop Prediction using Machine Learning,” in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India: IEEE, Aug. 2020, pp. 926–932. doi: 10.1109/ICSSIT48917.2020.9214190.

[7] Z. Hong, Z. Kalbarczyk, and R. K. Iyer, “A Data-Driven Approach to Soil Moisture Collection and Prediction,” in *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, St Louis, MO, USA: IEEE, May 2016, pp. 1–6. doi: 10.1109/SMARTCOMP.2016.7501673.

[8] “Analysis of soil prediction of crop yield using machine learning approach,” *Int. J. Adv. Res. Comput. Sci.*.

[9] Xu Qiao, Feng Yang, and Xianlei Xu, “The prediction method of soil moisture content based on multiple regression and RBF neural network,” in *Proceedings of the 15th International Conference on Ground Penetrating Radar*, Brussels: IEEE, Jun. 2014, pp. 140–143. doi: 10.1109/ICGPR.2014.6970402.

[10] N. Suresh *et al.*, “Crop Yield Prediction Using Random Forest Algorithm,” in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India: IEEE, Mar. 2021, pp. 279–282. doi: 10.1109/ICACCS51430.2021.9441871.