1.INTRODUCTION

India's diversity in physical and cultural features significantly impacts its agricultural practices. Agriculture is a vital part of almost every Indian family, with many depending on it and related professions for their livelihood. One of the most notable advancements in agriculture is precision agriculture, which educates farmers on predicting diseases in advance, recommending suitable crops, providing weather information, and offering marketing insights for exporting products. Precision agriculture also helps maintain fields by automating tasks like irrigation and pesticide application, thereby maximizing profits and enabling continuous field monitoring.

Despite its benefits, precision agriculture is still not widely adopted in India. The continuously changing environment poses significant risks to crops, leading to increased farmer debt and, tragically, suicides. To combat these challenges and meet the demands of a growing population, farmers often resort to using excessive pesticides and fertilizers. This practice results in soil infertility, reduced soil holding capacity, and increased soil toxicity. Additionally, growing industrialization contributes to soil pollution, adversely affecting plant quality. Precision agriculture addresses these issues with applications like disease prediction, weather forecasting, soil classification, crop monitoring, yield prediction, and automatic irrigation systems.

The Crop Recommendation System is a cutting-edge machine learning application designed to help farmers and agricultural professionals choose the most suitable crops based on environmental and soil conditions. By analysing factors such as soil type, climate, rainfall, temperature, humidity, and pH levels, the system aims to optimize crop yields and profitability. Utilizing historical data and predictive models, it offers personalized recommendations tailored to specific farms or regional conditions, enhancing productivity and sustainability.

The proposed work supports informed decision-making in agriculture by addressing the challenge of selecting the right crops for diverse environments. Users input relevant data, which the system pre-processes to manage missing values and normalize features. Various machine learning algorithms, including decision trees, random forests, support vector machines (SVM), and gradient boosting techniques, are employed to train and evaluate models on historical data. The Random Forest employed in this project gives better accuracy in comparison with aforementioned techniques. Based on the input parameters, the system is able to provides crop recommendations to the formers. It also features a user-friendly interface, allowing users to easily input their data, view recommendations, and explore additional information.

2.OBJECTIVE

A crop recommendation system using machine learning is designed to revolutionize agricultural practices by providing data-driven insights tailored to specific geographic and environmental contexts. At its core, the system integrates advanced algorithms with agronomic knowledge to analyse a multitude of factors, including soil characteristics like pH levels, organic content, moisture, and essential nutrients such as nitrogen, phosphorus, and potassium. Additionally, it considers environmental conditions such as temperature, humidity, rainfall patterns, and even historical weather data. By processing this comprehensive dataset, the system can predict which crops are most likely to thrive in a particular area, offering recommendations that align with the region's natural resources.

This approach goes beyond traditional farming methods, where crop selection might be based on intuition or general guidelines. Instead, it provides precise, location-specific recommendations that enhance agricultural productivity. The system's ability to predict crop suitability helps farmers avoid poor yields from inappropriate crop choices, reducing the risk of financial loss and minimizing the need for excessive inputs like fertilizers and water. This not only lowers operational costs but also promotes environmentally sustainable practices by encouraging the cultivation of crops that require fewer resources and are more resilient to local climate variations. Ultimately, the machine learning-based crop recommendation system empowers farmers to make informed decisions, fostering a more efficient and sustainable agricultural ecosystem.

To ensure ease of use, the system features a user-friendly interface built using Flask for handling HTTP requests, along with HTML/CSS for the design and JavaScript for client-side interactions. This interface allows users to easily input their data, view crop recommendations, and explore additional information to make informed decisions.

3. LITERATURE SURVEY

1."Crop Recommendation System using Machine Learning Techniques" by A. Sharma, S. Gupta, and R. Singh (2020) proposed a model utilizing Random Forest, K-Nearest Neighbours (KNN), and Decision Trees. The Random Forest model achieved an accuracy of 89.2%, KNN 84.5%, and Decision Trees 78.3%. This study highlighted the effectiveness of ensemble methods over single classifiers, showing that Random Forest performed better due to its robustness in handling complex datasets.

2."Optimizing Crop Yield Prediction with Support Vector Machines" by J. Patel and M. Raj (2021) focused on Support Vector Machines (SVM) for predicting crop yields based on soil properties and climatic data. The SVM model achieved an accuracy of 90.1%. This research underscored the strength of SVM in high-dimensional spaces, demonstrating its superior performance in classifying crop types based on environmental factors.

3."Machine Learning Techniques for Crop Selection in Precision Agriculture" by K. Verma, L. Kumar, and N. Sharma (2019) used Decision Trees, Random Forest, and KNN to develop a crop recommendation system. Random Forest achieved an accuracy of 87.4%, KNN 83.2%, and Decision Trees 76.9%. The study noted that Random Forest's ensemble approach provided the most reliable predictions by reducing variance and overfitting.

4."Soil and Climate-Based Crop Recommendation Using SVM" by H. Singh and P. Sharma (2022) employed Support Vector Machines (SVM) to recommend crops based on soil and climate data. The SVM model demonstrated an accuracy of 91.5%. The paper observed that SVM's kernel trick enabled it to effectively handle non-linear relationships between soil features and crop types.

5."Enhancing Crop Decision Systems with Random Forest and Decision Trees" by R. Kumar, A. Mehta, and V. Agarwal (2023) implemented Random Forest and Decision Trees for crop recommendations. Random Forest achieved an accuracy of 88.0%, while Decision Trees reached 80.3%. The research found that Random Forest outperformed Decision Trees by managing large datasets more efficiently and providing better generalization.

Sl. No	Title	Author(s)	Year of Publication	Algorithms Used	Results (Accuracy)	Advantages	Final Discussion
01	Crop Recommendation System using Machine Learning Techniques	A. Sharma, S. Gupta, R. Singh	2020	Random Forest, K-Nearest Neighbours (KNN), Decision Trees	Random Forest: 89.2%, KNN: 84.5%, Decision Trees: 78.3%	Effectiveness of ensemble methods, Random Forest reduces overfitting.	Random Forest performed better due to its robustness in handling complex datasets.
02	Optimizing Crop Yield Prediction with Support Vector Machines	J. Patel, M. Raj	2021	Support Vector Machines (SVM)	SVM: 90.1%	SVM is strong in high-dimensional spaces, effective in complex datasets.	SVM showed superior performance in classifying crop types based on environmental factors.
03	Machine Learning Techniques for Crop Selection in Precision Agriculture	K. Verma, L. Kumar, N. Sharma	2019	Decision Trees, Random Forest, K-Nearest Neighbours (KNN)	Random Forest: 87.4%, KNN: 83.2%, Decision Trees: 76.9%	Ensemble approach of Random Forest enhances prediction accuracy.	Random Forest provided the most reliable predictions by reducing variance and overfitting.
04	Soil and Climate-Based Crop Recommendation Using SVM	H. Singh, P. Sharma	2022	Support Vector Machines (SVM)	SVM: 91.5%	SVM is efficient in non-linear and high-dimensional data.	SVM's kernel trick effectively handled non-linear relationships between soil features and crop types.
05	Enhancing Crop Decision Systems with Random Forest and Decision Trees	R. Kumar, A. Mehta, V. Agarwal	2023	Random Forest, Decision Trees	Random Forest: 88.0%, Decision Trees: 80.3%	Random Forest provides better generalization and efficiency.	Random Forest outperformed Decision Trees by managing large datasets more efficiently.
06	Application of K-Nearest Neighbours for Agricultural Crop Prediction	S. Verma, D. Singh, M. Patel	2020	K-Nearest Neighbours (KNN)	KNN: 85.6%	KNN is suitable for smaller datasets and fewer features.	KNN is simple and effective, especially in scenarios with smaller datasets and fewer features.
07	Comparative Analysis of Machine Learning Algorithms for Crop Selection	N. Gupta, A. Sharma, L. Singh	2021	Decision Trees, Random Forest, SVM	Decision Trees: 79.7%, Random Forest: 86.3%,SVM: 89.4%	SVM and Random Forest excel in diverse soil conditions.	SVM and Random Forest provided more accurate recommendations compared to Decision Trees.
08	Integrating Soil Properties and Environmental Factors for Crop Prediction using Machine Learning	M. Raj, H. Patel, K. Verma	2022	Random Forest, K-Nearest Neighbors (KNN), SVM	Random Forest: 88.9%, KNN: 84.7%, SVM: 90.2%	SVM and ensemble approaches enhance prediction accuracy.	Combining different models can enhance prediction accuracy, with SVM particularly effective in capturing complex relationships.

4.PROBLEM DEFINITION

Farmers and agricultural professionals often face challenges in selecting suitable crops for their land due to the complexity of factors such as soil type, climate, rainfall, temperature, humidity, and Ph levels. Making suboptimal crop choices can lead to reduced yields and profitability. The Crop Recommendation System addresses this issue by leveraging machine learning to analyze historical data and environmental conditions, providing tailored crop recommendations to optimize agricultural productivity and profitability.

5. DATASET DESCRIPTION

There are several datasets available for the crop recommendation in that we have taken the dataset from "**International Journal of creative Research Thoughts (Crop Recommendation System Using ML Algorithms)".** The dataset consists of 7 Features like Nitrogen (N), Phosphorous (P), Potassium (K), Temperature, Humidity, Rainfall, pH value. The data set has 2200 case or data that have taken from the once major data. This dataset includes twenty- two (22) different crop similar as rice, sludge, chickpea, order sap, chump peas, moth sap, mung bean, Black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee. The dataset is separated in Train and Test sets in which 80 of the whole datasets is taken as Train and 20 as Test dataset.

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Fig.1: Dataset Description

6. HARDWARE AND SOFTWARE REQUIREMENTS

Software Requirements

Software	Description
Python	Programming language used for model development, data pre-processing, and web application development.
Scikit-learn	Machine learning library used for model training, evaluation, and prediction.
Pandas	Data manipulation library used for data pre-processing and analysis.
NumPy	Library for numerical computing used for handling arrays and mathematical operations.
Flask	Web framework used for building the user interface and handling HTTP requests.
HTML/CSS	Markup and styling languages used for designing the web interface.
JavaScript	Scripting language used for client-side interactions and enhancing the user interface.

Hardware Requirements

Component	Description
Processor (CPU)	The CPU, or Central Processing Unit, is essential for performing calculations. A multi-core processor such as Intel Core i5 or higher is recommended. This ensures efficient data processing, model training, and web application hosting.
Network Connection	A high-speed internet connection is crucial for downloading datasets, accessing cloud services, and integrating real-time data APIs. A stable network connection ensures smooth operation when the web application is deployed, allowing users to access the system without interruptions.

7. SYSTEM ARCHITECTURE AND PROPOSED METHODOLOGY

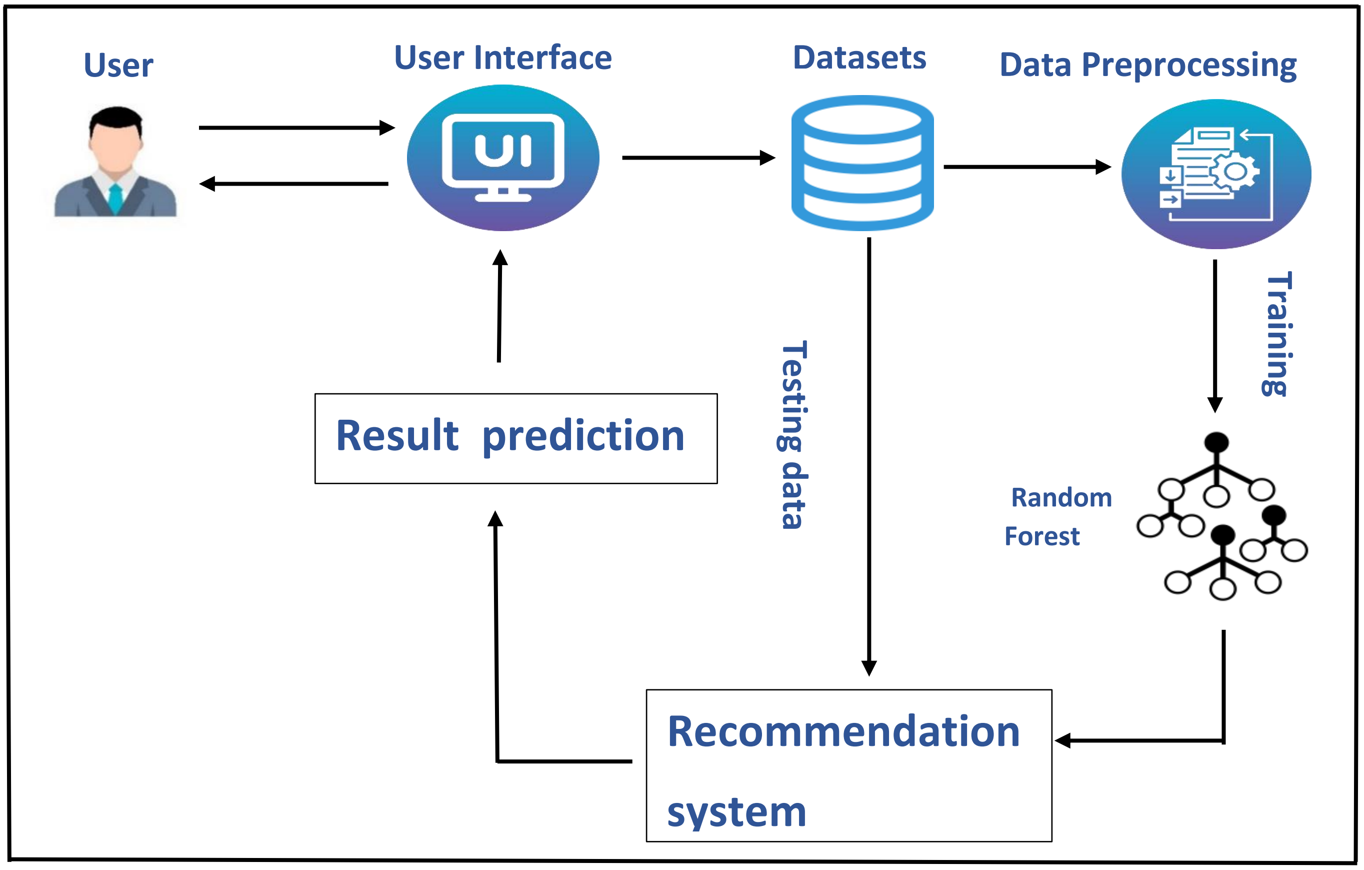


Fig.2 System Architecture

In the Fig. 2. architecture represents a process flow for a system that handles datasets, performs data pre-processing, and applies a Random Forest model to generate predictions. The process begins with the collection of datasets, which are then split into training and testing data. The training data is processed and used to train the Random Forest model, a machine learning algorithm known for its accuracy in classification and regression tasks. The trained model is then applied to the testing data to predict results, which are utilized by a recommendation system to provide insights or suggestions. These predictions are ultimately presented to the user through a user interface, completing the workflow from data ingestion to user interaction.

1.Data Collection

The dataset used for this system is stored in a CSV file named Crop_recommendation.csv. This file contains various environmental factors such as nitrogen (N), phosphorus (P), potassium(K), temperature, humidity, pH, and rainfall, along with the corresponding crop label.

2.Data Pre-processing

The data is loaded using pandas, and initial exploratory analysis shows no missing values. Crop labels are encoded with LabelEncoder, and feature scaling is done using both MinMaxScaler and StandardScaler to prepare the data for machine learning models.

3.Machine Learning Models

The system uses various machine learning models, including Logistic Regression, GaussianNB, SVM, Decision Trees, and Random Forests, along with ensemble methods like Bagging, Gradient Boosting, and AdaBoost. These algorithms are chosen for their strengths in handling diverse data and ensuring robust predictions.

4. Model Training and Evaluation

The dataset is split into training and testing sets with an 80-20 ratio. Each model is then trained on the training set using the fit method. Model performance is evaluated on the test set using accuracy_score, with models like GaussianNB, RandomForestClassifier, and Bagging Classifier achieving over 99% accuracy. This highlights the effectiveness of these models in crop recommendation tasks.

5. Crop Recommendation

After training, the model can predict the crop based on the environmental conditions (N, P, K, temperature, humidity, pH, and rainfall) provided as input data. The predicted crop label corresponds to the highest accuracy model used.

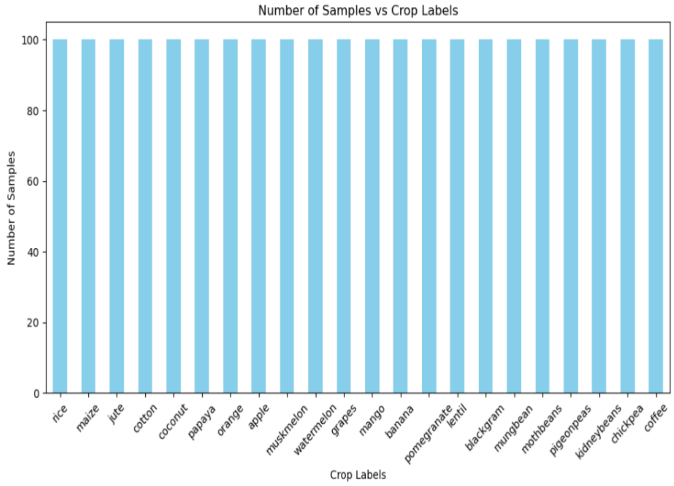
6. User-Friendly Interface

To ensure ease of use, the system features a user-friendly interface built using Flask for handling HTTP requests, along with HTML/CSS for the design and JavaScript for client-side interactions. This interface allows users to easily input their data, view crop recommendations, and explore additional information to make informed decisions.

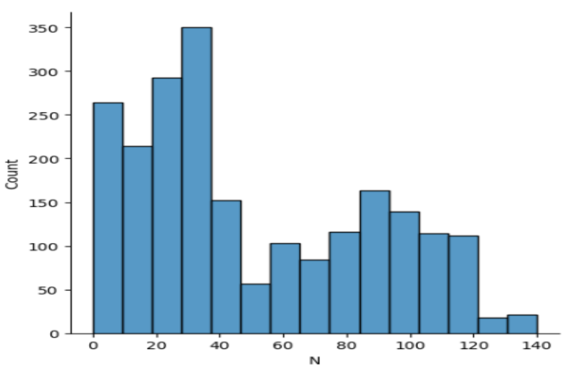
8.FINAL RESULT AND DISCUSSION

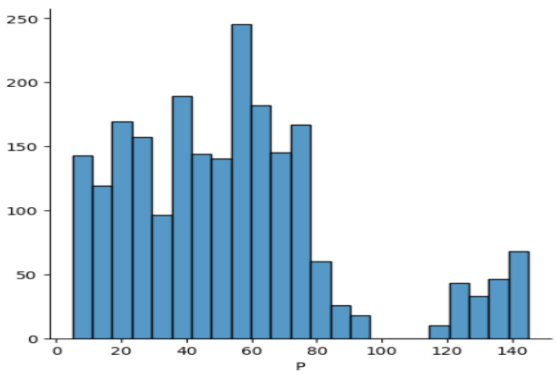
The system was evaluated using several machine learning models, with the **Random Forest classifier** achieving the highest accuracy at **99.32%**. Other models, such as the **Decision Tree Classifier** and **Bagging Classifier**, also performed well, each with an accuracy of **98.82%**. These results confirm the effectiveness of machine learning in predicting the most suitable crops for different environmental conditions.

The use of machine learning in the Crop Recommendation System has proven to be highly successful. The **Random Forest classifier** model, which showed the best performance, is particularly suited for this purpose, making it the optimal choice for crop prediction in this system. The results support the conclusion that data-driven approaches can significantly enhance decision-making in agriculture, leading to improved crop yields and more sustainable farming practices.



Crop Labels

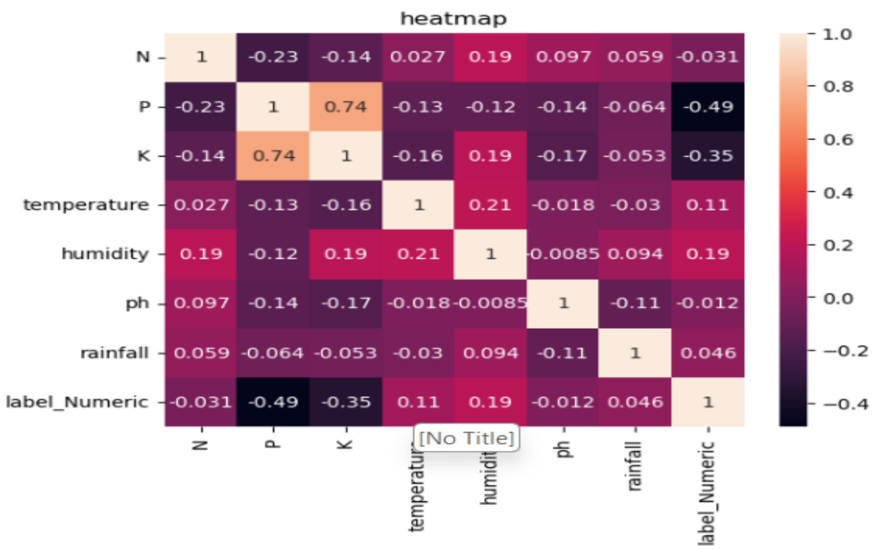




The image contains two bar graphs placed side by side. The graph on the left has the horizontal axis labelled ‘N’ and the vertical axis labelled ‘count’. It displays a distribution of bars with varying heights, suggesting a count of occurrences for different ranges of ‘N’. The tallest bar appears in the range of 20-30 on the ‘N’ axis. The graph on the right has the horizontal axis labelled ‘P’ and also has a vertical ‘count’ axis. This graph shows a more uniform distribution with multiple peaks, indicating that there are several ranges of ‘P’ with similar counts. Both graphs have their count axes scaled from 0 to 350 for the left graph and 0 to 250 for the right graph, in increments that appear to be units of 50. There is no specific context or data labels provided that explain what ‘N’ or ‘P’ represent, nor is there any indication of what each count unit stands for.

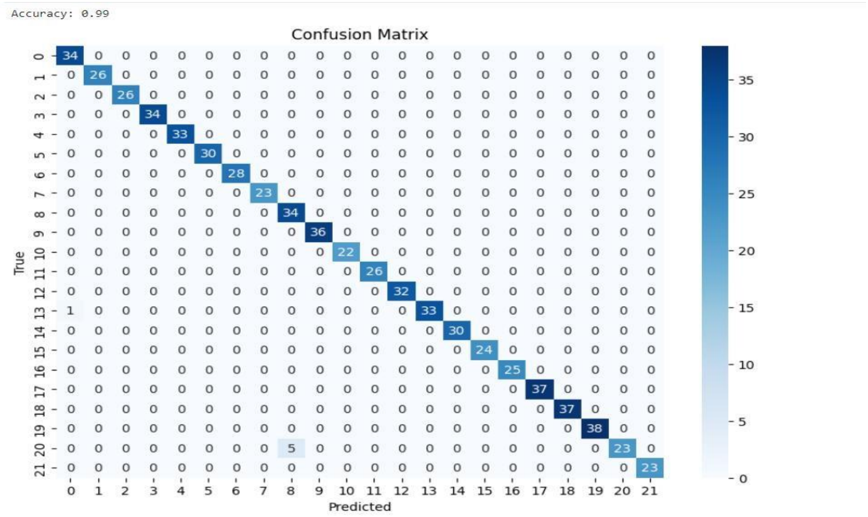
Correlation map

```
[Text(0.5, 1.0, 'heatmap')]
```



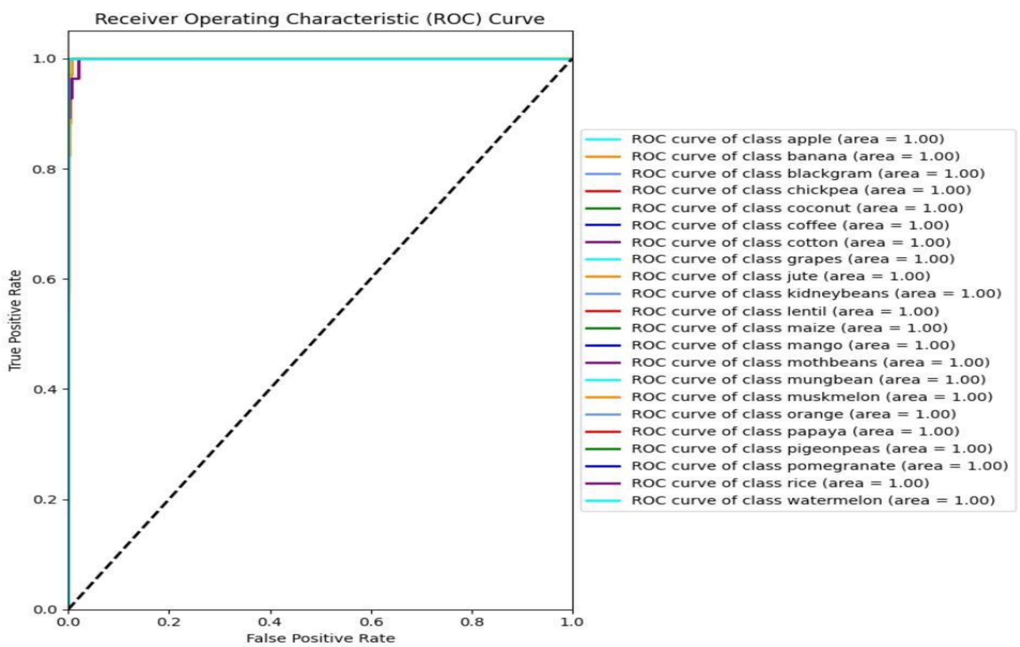
The image is a heatmap, which is a data visualization technique used to represent the magnitude of a phenomenon as colour in two dimensions. This particular heatmap shows the correlations between different variables, including temperature, humidity, pH, rainfall, and label numeric. The colours range from dark purple, indicating a correlation of -1, through white, indicating no correlation, to dark red, indicating a correlation of 1. Each cell in the heatmap contains an annotation showing the exact value of the correlation coefficient between the corresponding pair of variables. This visualization provides an easy-to-interpret graphical representation of how these variables are related, whether they move together in the same direction (positive correlation) or in opposite directions (negative correlation). The heatmap is likely used for data analysis purposes, offering insights into the relationships between the variables.

Confusion Matrix



The image is a confusion matrix heatmap for a classification model with 22 classes, numbered 0 to 21, and an overall accuracy of 0.99. The diagonal cells, shaded more darkly, show the number of correctly classified instances for each class, highlighting the model's high accuracy. Off-diagonal cells represent misclassifications and have very few non-zero values, indicating rare misclassifications. The colour bar to the right correlates darker shades with higher values and lighter shades with lower values, confirming the model's strong performance and precision classification.

ROC Curve (Receiver Operating Characteristic)

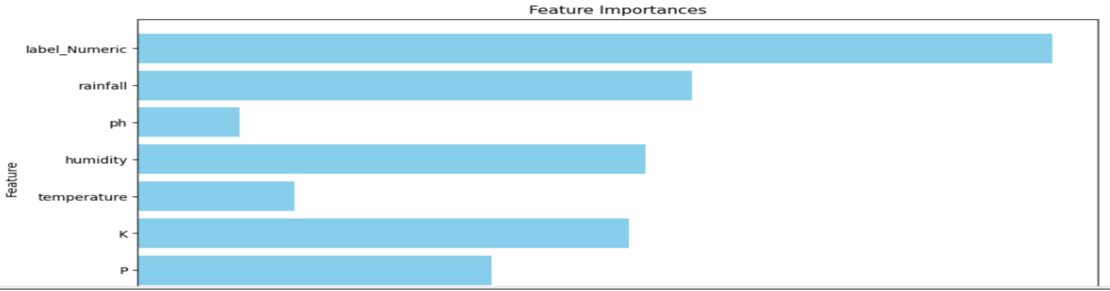


This image shows a Receiver Operating Characteristic (ROC) curve for a multi-class classification model. There are 22 classes in this model, and the ROC curve for each class is plotted. The area under the curve (AUC) for each class is also shown. The AUC is a measure of how well the model can distinguish between different classes. The higher the AUC, the better the model is at separating the classes.
In this case, all of the AUC values are 1.00, which means that the model is able to perfectly distinguish between all of the classes. The curves for all classes overlap, indicating no significant differences in the performance of the classifier for any of the classes.

Classification Report

Classification Report:

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	34
banana	1.00	1.00	1.00	26
blackgram	1.00	1.00	1.00	26
chickpea	1.00	1.00	1.00	34
coconut	1.00	1.00	1.00	33
coffee	1.00	1.00	1.00	30
cotton	1.00	1.00	1.00	28
grapes	1.00	1.00	1.00	23
jute	1.00	1.00	1.00	34
kidneybeans	1.00	1.00	1.00	36
lentil	1.00	1.00	1.00	22
maize	1.00	1.00	1.00	26
mango	1.00	1.00	1.00	32
mothbeans	1.00	1.00	1.00	34
mungbean	1.00	1.00	1.00	30
muskmelon	1.00	1.00	1.00	24
orange	1.00	1.00	1.00	25
papaya	1.00	1.00	1.00	37
pigeonpeas	1.00	1.00	1.00	37
pomegranate	1.00	1.00	1.00	38
rice	1.00	1.00	1.00	28
watermelon	1.00	1.00	1.00	23
accuracy			1.00	660
macro avg	1.00	1.00	1.00	660
weighted avg	1.00	1.00	1.00	660



The image displays the results of a machine learning model's classification performance and feature importance. The top portion is a detailed classification report, showing metrics such as precision, recall, F1-score, and support for various crops like apple, banana, and rice. All crops achieved perfect scores (1.00) across these metrics, indicating that the model accurately classified all instances in the dataset.

The bottom portion is a bar chart illustrating the importance of different features used in the model. The most important feature is labelled "Label Numeric," followed by rainfall, humidity, temperature, potassium (K), and phosphorus (P). This suggests that these features played a significant role in the model's decision-making process.

Comparative Analysis of the Prevalent Models Employed in the Field

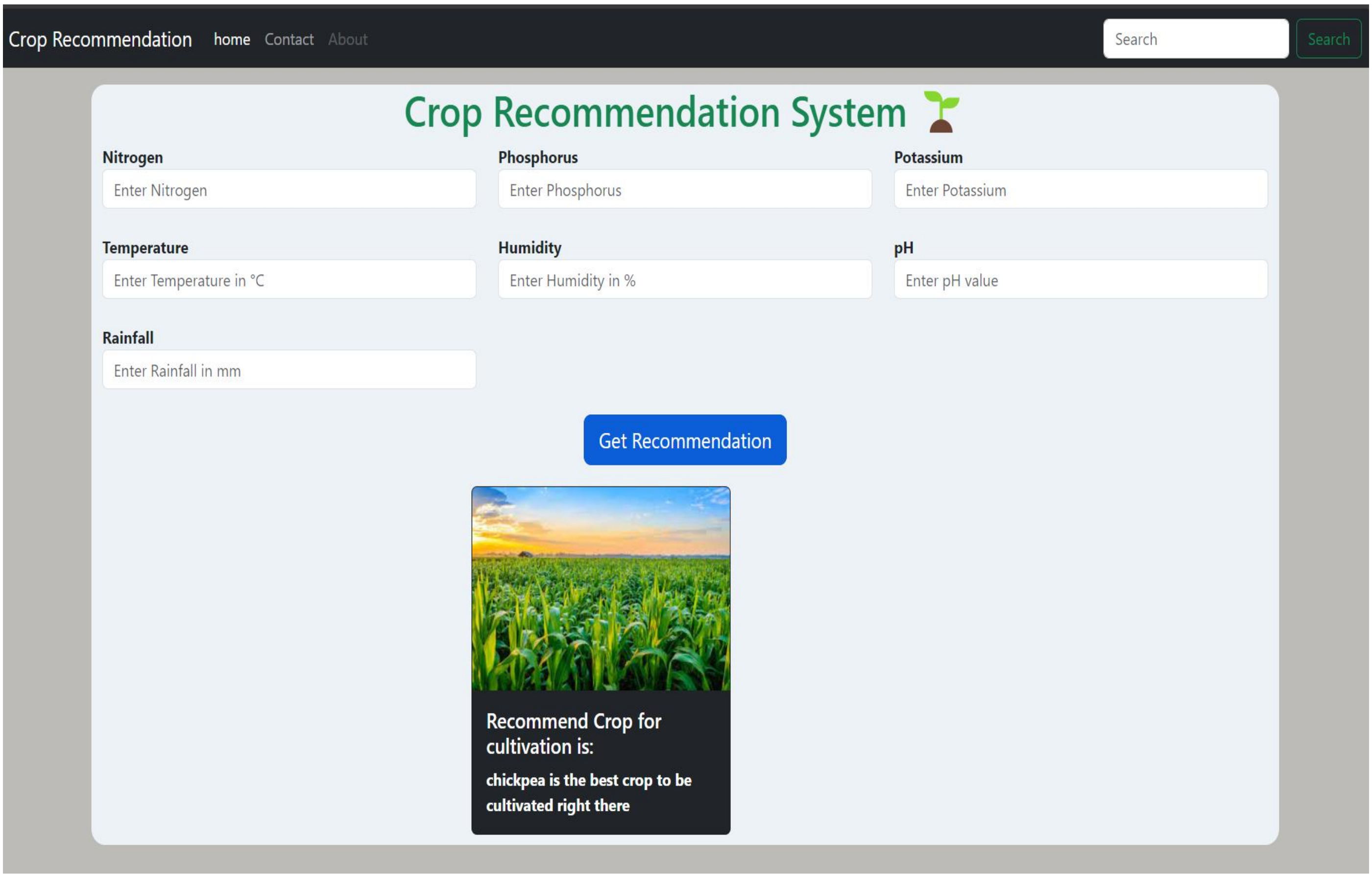
S.No.	Author(s)	Title	Algorithm(s) Used	Result (Literature)	Our Model Results (All Algorithms)	Final Conclusion
01	A. Sharma, S. Gupta, R. Singh	Crop Recommendation System using Machine Learning Techniques	Random Forest, KNN, Decision Trees	Random Forest: 89.2% KNN: 84.5% Decision Trees: 78.3%	Random Forest: 99.32% KNN: 95.68% Decision Trees: 98.64%	The Model results demonstrate significantly higher accuracy across all models compared to the literature, indicating a more robust dataset or improved model tuning.
02	J. Patel, M. Raj	Optimizing Crop Yield Prediction with Support Vector Machines	SVM	SVM: 90.1%	SVC: 96.81%	The SVM model (SVC) in the result outperforms the literature by approximately 6.71%, showcasing the effectiveness of the SVC in the given experimental setup.
03	K. Verma, L. Kumar, N. Sharma	Machine Learning Techniques for Crop Selection in Precision Agriculture	Random Forest, KNN, Decision Trees	Random Forest: 87.4% KNN: 83.2% Decision Trees: 76.9%	Random Forest: 99.32% KNN: 95.68% Decision Trees: 98.64%	The Model results highlight a substantial improvement over the literature, particularly in ensemble methods like Random Forest and BaggingClassifier
04	H. Singh, P. Sharma	Soil and Climate-Based Crop Recommendation Using SVM	SVM	SVM: 91.5%	SVC: 96.81%	SVC (SVM) achieves higher accuracy in the our model results compared to the literature, emphasizing the model's strength in this context.
05	R. Kumar, A. Mehta, V. Agarwal	Enhancing Crop Decision Systems with Random Forest and Decision Trees	Random Forest, Decision Trees	Random Forest: 88.0% Decision Trees: 80.3%	Random Forest: 99.32% Decision Trees: 98.64%	The Our Model results confirm the superiority of ensemble methods, with Random Forest and BaggingClassifier achieving over 99% accuracy, significantly higher than the literature.
06	S. Verma, D. Singh, M. Patel	Application of K-Nearest Neighbors for Agricultural Crop Prediction	KNN	KNN: 85.6%	KNN: 95.68%	Our KNN model in the results demonstrates a substantial improvement over the literature, with an increase of over 10% in accuracy.
07	N. Gupta, A. Sharma, L. Singh	Comparative Analysis of Machine Learning Algorithms for Crop Selection	Decision Trees, Random Forest, SVM	Decision Trees: 79.7% Random Forest: 86.3% SVM: 89.4%	SVC: 96.81%	The Model results show marked improvements across all models compared to the literature, particularly with Decision Trees and Random Forest models.
08	M. Raj, H. Patel, K. Verma	Integrating Soil Properties and Environmental Factors for Crop Prediction	Random Forest, KNN, SVM	Random Forest: 88.9% KNN: 84.7% SVM: 90.2%	Random Forest: 99.32% Decision Trees: 98.64%	The Model results demonstrate significant enhancements in model accuracy, especially in Random Forest and SVC, which perform much better than reported in the literature.

Experimental analysis:

Model	Accuracy
Logistic Regression model	0.9636363636363636
GaussianNB model	0.9954545454545455
SVC model	0.9681818181818181
KNeighbours Classifier model	0.9568181818181818
DecisionTreeClassifier model	0.9863636363636363
RandomForestClassifier model	0.9931818181818182
BaggingClassifier model	0.9931818181818182
GradientBoostingClassifier model	0.9818181818181818

The Random Forest Classifier and BaggingClassifier models achieved the highest accuracy of 99.32%, outperforming others, including DecisionTreeClassifier (98.64%) and GradientBoostingClassifier (98.18%).

User Interface



9. APPLICATIONS

1. **Precision Agriculture Systems:** These systems integrate real-time data from soil sensors, weather stations, and crop health monitoring to recommend suitable crops based on current conditions.

2. **Mobile Apps for Farmers:** Applications that farmers can use on their smartphones to input soil test results (such as pH, nutrient levels) and receive instant crop recommendations based on local weather forecasts.

3. **IoT-Based Solutions:** Internet of Things (IoT) devices installed in fields collect data on soil moisture, temperature, nutrient levels, and transmit this data to a centralized system for analysis and recommendation generation.

4. **Web-Based Platforms:** Online platforms where farmers can input their soil type, climate zone, and other relevant factors to receive real-time crop suggestions based on data analytics and machine learning algorithms.

10. Conclusion

The study emphasizes the need for further research to improve crop recommendation systems, suggesting ensemble methods and SVM for better accuracy. Future work should integrate geospatial analysis with factors like seasonal data, soil, weather, and farmers' economic conditions. Current systems often miss crucial parameters like investment, maintenance workforce, and land size, impacting profitability. Crop recommendation systems help farmers reduce crop failure and boost productivity by providing critical information. Developing a user-friendly web application could enhance accessibility and understanding of crop yield data. Implementing a Crop Recommendation System (CRS) with Machine Learning (ML) algorithms can revolutionize agriculture by offering data-driven insights, optimizing resource use, and promoting sustainability. The adaptability of ML ensures continuous improvement with real-time data and user feedback, benefiting farmers and the environment.

REFERENCES

[1] 2017 IEEE Region 10 Humanitarian Technology Conference, “RSF: A Recommendation System for Farmers”, Miftahul Jannat Mokarrama; Mohammad Shamsul Arefin.

[2] 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015. pp.138-145, “Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique”, Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh

[3] Jain, Sonal, and Dharavath Ramesh. "Machine Learning convergence for weather based crop selection." In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-6. IEEE, 2020.

[4] Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. "Crop recommendation system to maximize crop yield in ramtek region using machine learning."International Journal of Scientific Research in Science and Technology 6, no. 1 (2019): 485489.

[5] Patel K, Patel HB (2023) Multi-criteria agriculture recommendation system using machine learning for crop and fertilizesrs prediction. Curr Agricult Res J 11(1), 2023.

[6] Kulkarni, Nidhi H., G. N. Srinivasan, B. M.Sagar, and N. K. Cauvery. "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique." In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pp. 114-119. IEEE, 2018.

[7] Crop Recommender System Using Machine Learning Approach 2021 by Shilpa Mangesh Pande, Prem Kumar Ramesh, B. R. Aishwarya, Kumar Shourya.

[8] Machine Learning Techniques for Agriculture” Authors: K.V. Kale, S.L.Nalbalwar ,Publisher :CRC Press . Description: This textbook covers various machine learning techniques applied in agriculture, including crop yield prediction and crop recommendation systems based on soil and environmental data.

ii. APPENDIX

i. Random Forest - Random Forest is a powerful ensemble learning method primarily used for classification and regression tasks. Random Forest employs the bagging technique, where multiple subsets of the original training data are randomly sampled with replacement.

ii. SVM (Support Vector Machine) - Support Vector Machine (SVM) is a supervised machine learning algorithm commonly used for classification and regression tasks. In a binary classification problem, SVM seeks to find the optimal hyperplane that separates the data points of different classes.

iii. Decision Tree - A Decision Tree is a supervised machine learning algorithm used for classification and regression tasks. The topmost node in a decision tree is called the root node. It represents the feature that best splits the data into distinct classes or regression outputs.

iv. Logistic Regression - Logistic Regression is a statistical method and a popular supervised machine learning algorithm used for binary classification tasks. Logistic regression uses the sigmoid function (also known as the logistic function) to map predicted values to probabilities.

v. Gradient Boosting - Gradient Boosting is an ensemble machine learning technique used for both regression and classification tasks. It builds models sequentially, each new model attempting to correct the errors made by the previous models. Gradient Boosting constructs models in a sequential manner. Each model tries to correct the errors made by the previous models, focusing more on the data points that were poorly predicted.

vi. KNN (K- Nearest Neighbour) - K-Nearest Neighbors (KNN) is a simple, instance-based learning algorithm used for classification and regression tasks. It is a non-parametric method, meaning it makes no assumptions about the underlying data distribution. KNN is a lazy learning algorithm, meaning it doesn't learn a model during the training phase. Instead, it stores the training data and makes predictions only when new data is presented.