IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Agrogenie: Bridging Agriculture & Technology Using Machine Learning And Deep Learning

[1] Poonam.Salunkhe, [2]Prerna.Gade, [3] Juhilee.Dhole, [4] Sanskruti.Itkar, [5] Dr.Kishor.Sakure,
[1,2,3,4] Final Year B.E. Student, [5] Professor,
[1,2,3,4,5] Department of Computer Engineering,
[1,2,3,4,5] *Terna Engineering College*
Navi Mumbai, India

Abstract: Agriculture is vital to the Indian economy, yet recent shifts have caused a crisis in the sector. To address this, the research aims to make agriculture profitable by using machine learning, bridging agriculture with technology, and deep learning. By predicting suitable crops based on climate and historical data, farmers can make informed decisions. An accompanying web application integrates features like crop recommendations, fertilizer guidance, disease detection, yield forecasting, and user feedback. This holistic approach revitalizes agriculture by leveraging modern technologies and data-driven insights. By bridging agriculture with technology and deep learning, the study aims to user in sustainable and profitable farming practices in India, ensuring the sector's resilience and growth.

Index Terms - Agriculture crisis, Profitable farming, Machine learning, Deep learning, Augmented learning. Sustainable farming practices, Web application for farmers.

1.INTRODUCTION

Agriculture, a vital pillar of the Indian economy, grapples with a crisis stemming from ongoing structural shifts, resulting in

diminished contributions to the GDP and a shift towards reliance on food imports. This crisis poses a threat to all sectors of the national economy. To address this, it is imperative to render agriculture profitable to sustain farmer engagement in crop production.

Traditionally, farmers relied on rudimentary methods and past experiences to navigate uncertainties in crop yield projections. However, with the advent of modern technologies like machine learning, the landscape is evolving. These advanced techniques offer a more refined and data-driven approach, promising greater accuracy and efficiency in predicting crop yields. Yet, despite these advancements, significant hurdles persist. A notable challenge is the lack of widespread awareness among farmers regarding optimal cultivation practices, exacerbated by the complexities of weather patterns and environmental variables. Furthermore, access to accurate historical crop yield data remains a bottleneck in effective decision-making, especially in agricultural risk management. In response to these challenges, this research advocates for integrating machine learning techniques to predict crop yields based on climatic conditions and historical data. This initiative seeks to equip farmers with indispensable insights into potential production outcomes, empowering them to make informed decisions before sowing crops. Moreover, recognizing the multifaceted nature of agricultural challenges, the proposed system goes beyond yield prediction, incorporating features such as crop and fertilizer recommendations, plant disease detection, and avenues for user feedback. By offering a comprehensive suite of tools and support mechanisms, this holistic approach endeavors to revolutionize Indian agriculture, steering it towards resilience and sustainability in the face of evolving challenges.

2. RELATED WORK

In[1] Smart Farming Prediction Using Machine Learning the paper is about using machine learning with various environmental factors like soil, pressure, weather, and crop type to predict the maximized profitable crop to grow. The paper primarily delves into algorithms for forecasting crop yield and predicting crop costs with precision.

The paper [2] presented at the ICCCI introduces a machine-learning methodology for crop yield forecasting based on climate parameters. It details the development of Crop Advisor, a user-friendly web tool, to predict the impact of climatic factors on crop yields. Utilizing the C4.5 algorithm, the study identifies key climatic parameters influencing crop productivity in select districts of Madhya Pradesh.

The paper [3] was published in the International Journal of Research in Engineering and Technology (IJRET), the paper explores the analysis of crop yield prediction through data mining methodologies. With a focus on rice production, the primary objective is to develop a user-friendly interface tailored for farmers, offering comprehensive analysis based on available data.

In [4], the research focuses on aiding beginner farmers through machine learning-guided crop selection. It employs a Naive Bayes algorithm, specifically a Gaussian classifier with boosting, to ensure accurate crop predictions. This supervised learning framework enhances crop selection precision, empowering novice farmers with informed decision-making capabilities.

In [5] Prakash M. Mainkar et al present a software solution for automated plant leaf disease detection and classification, aiming to enhance crop productivity. The methodology comprises image acquisition, preprocessing, segmentation, feature extraction, and classification steps.

In [6] Smart Farming Involves the Use of Machine Learning and Deep Learning Strategies: Senthil Kumar Swami Durai., And Mary Divya Shamili. One study discussed utilizes support vector machines and decision trees to predict crops based on micronutrients and meteorological characteristics, achieving an impressive 92% accuracy for rice, wheat, and sugarcane crops. Furthermore, the integration of deep learning techniques for weed detection is explored, where the proposed model incorporates pre-trained Resnet152V2 and Center Net algorithms, leading to improved accuracy and offering additional insights.

In [7] A Machine Learning-primarily Based Approach for Crop Yield Prediction and Fertilizer Recommendation., R. Harish., and Priya. The literature review explores crop recommendation systems, emphasizing the crucial role of accurate predictions in maximizing crop yield. It underscores the significance of precision agriculture and the necessity for reliable models to suggest crops based on specific parameters.

In a study [8], Amar Kumar Deya et al. employ an image processing algorithm to detect leaf rot disease in betel vine (Piper betel L.). They address limitations in manual detection and suggest utilizing color features of rotted leaf areas for disease identification. Subsequently, they segment the affected areas and extract the rotted leaf portion from various plant features. The research showcases the practical effectiveness of this automated vision-based system with straightforward validation.

3. PROPOSED METHODOLOGY

The objective is to underscore the urgent need for transforming agriculture into a profitable venture while highlighting the potential of modern technologies like machine learning to address challenges such as predicting crop yields accurately. By proposing the integration of machine learning techniques based on climatic conditions and historical data, the aim is to empower farmers with informed decision-making tools and provide holistic support through features like crop recommendations, fertilizer guidance, plant disease detection, and avenues for user feedback.

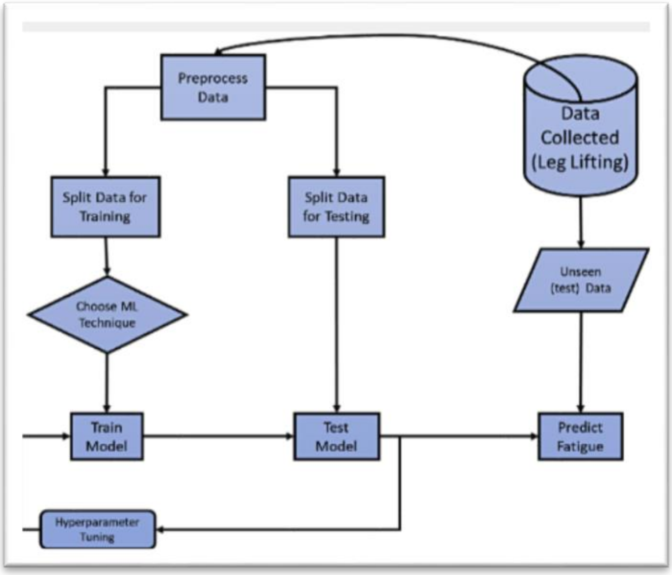


Fig 3.1: Prediction Framework

1) **Data Collection:** By proposing the integration of machine learning techniques based on climatic conditions and historical data, the aim is to empower farmers with informed decision-making tools. Data for these models can be sourced from various sources, including Kaggle, a platform known for hosting datasets relevant to agriculture. This data collection process will encompass variables crucial for fertilizer recommendation, plant disease detection, crop yield prediction, and user feedback, ensuring comprehensive and accurate insights for agricultural decision-making.

2) **Data Cleaning**: In preparing the data for machine learning in our applications, we first address missing values by either filling them in or removing them, ensuring our dataset is complete. Next, we detect and eliminate outliers that could skew our predictions.

Then, we normalize our numerical features to maintain consistency in scale across variables. Additionally, we convert categorical variables into numerical representations for easier integration into our models. If we encounter skewed distributions, we apply transformations to make our data more symmetric. Finally, we may balance our dataset to address any class imbalances, ensuring fair and accurate predictions. Through these steps, we ensure that our data is clean, standardized, and ready for training our machine learning algorithms effectively in the context of smart practices.

3) **Data Pre-processing**: During the pre-processing stage, the dataset is divided into training and testing sets to evaluate the machine learning models effectively. Finally, the data is formatted in a suitable structure for input into the machine learning algorithms, facilitating accurate predictions for fertilizer recommendation, plant disease detection, crop yield prediction, and user feedback.

4) Machine Learning Algorithms:

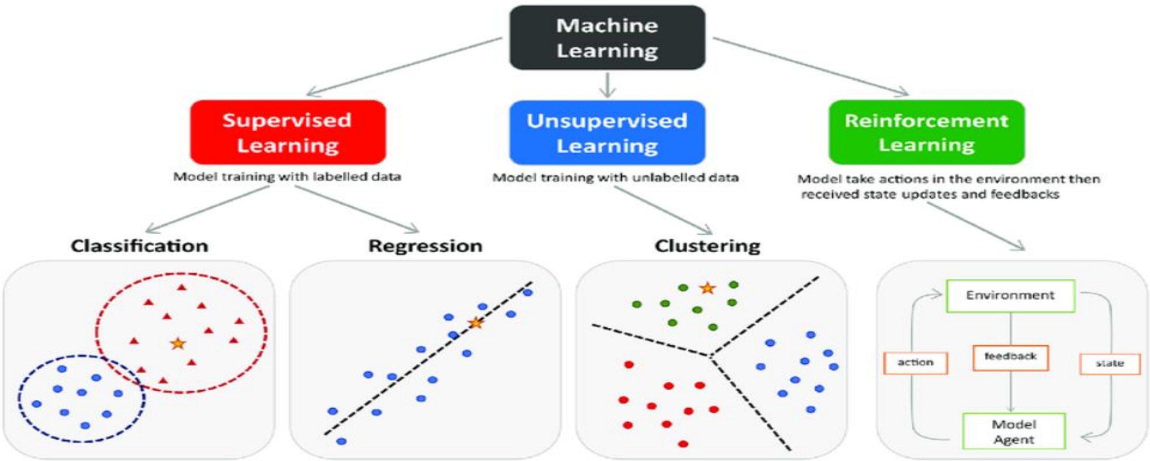
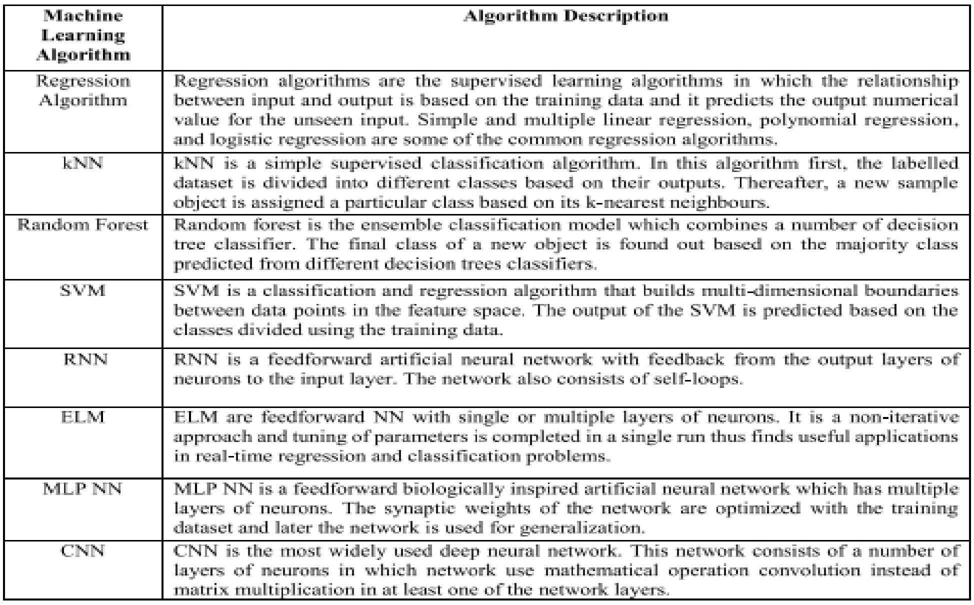


Fig 3.2: Classification of ML models

Machine learning algorithms fall into three primary categories: supervised learning, unsupervised learning, and reinforcement learning.. Supervised learning involves learning with guidance from a teacher, where algorithms use labeled data with known inputs and outputs. They learn from this labeled data to predict outputs for new inputs. Classification algorithms predict categorical values, while regression algorithms predict numerical values. Unsupervised learning algorithms work without labeled data and find similarities among objects, aiming to uncover hidden patterns. This approach is more challenging as it requires extracting knowledge without predefined labels. Reinforcement learning, on the other hand, learns from the environment through rewards and penalties. For instance, AlphaGo, a chess-playing game, uses reinforcement learning to outsmart other chess programs. In the agricultural context, various machine learning algorithms are used for tasks like crop yield prediction, fertilizer recommendation, and disease identification. These algorithms are analyzed and discussed, and Table 3.1 summarizes the types of algorithms employed for these purposes.

Table 3.1: Machine Learning models



Machine Learning Algorithm	Algorithm Description
Regression Algorithm	Regression algorithms are the supervised learning algorithms in which the relationship between input and output is based on the training data and it predicts the output numerical value for the unseen input. Simple and multiple linear regression, polynomial regression, and logistic regression are some of the common regression algorithms.
kNN	kNN is a simple supervised classification algorithm. In this algorithm first, the labelled dataset is divided into different classes based on their outputs. Thereafter, a new sample object is assigned a particular class based on its k-nearest neighbours.
Random Forest	Random forest is the ensemble classification model which combines a number of decision tree classifier. The final class of a new object is found out based on the majority class predicted from different decision trees classifiers.
SVM	SVM is a classification and regression algorithm that builds multi-dimensional boundaries between data points in the feature space. The output of the SVM is predicted based on the classes divided using the training data.
RNN	RNN is a feedforward artificial neural network with feedback from the output layers of neurons to the input layer. The network also consists of self-loops.
ELM	ELM are feedforward NN with single or multiple layers of neurons. It is a non-iterative approach and tuning of parameters is completed in a single run thus finds useful applications in real-time regression and classification problems.
MLP NN	MLP NN is a feedforward biologically inspired artificial neural network which has multiple layers of neurons. The synaptic weights of the network are optimized with the training dataset and later the network is used for generalization.
CNN	CNN is the most widely used deep neural network. This network consists of a number of layers of neurons in which network use mathematical operation convolution instead of matrix multiplication in at least one of the network layers.

5) **Prediction:** In the proposed system, leveraging machine learning techniques and datasets from sources like Kaggle, we aim to predict several key aspects crucial for agricultural decision-making.

1. **Crop Prediction:** By analyzing climatic conditions and historical data, our model predicts the most suitable crops for cultivation in specific regions. This prediction is based on factors such as temperature, precipitation, soil quality, and crop compatibility, empowering farmers to make informed choices about which crops to cultivate.
2. **Fertilizer Recommendation:** Utilizing data on soil composition, crop types, and nutrient requirements, our model generates personalized fertilizer recommendations. This prediction helps optimize crop growth and yield by ensuring that the soil receives the necessary nutrients for healthy plant development.
3. **Plant Disease Detection:** Through analysis of plant health indicators such as leaf color, texture, and overall appearance, our model detects signs of disease or pest infestation. Early detection enables prompt intervention, preventing widespread crop damage and minimizing yield losses.
4. **Crop Yield Prediction**: Leveraging historical yield data along with environmental factors, our model forecasts crop yields for upcoming seasons. This prediction assists farmers in estimating potential harvests, facilitating better resource allocation and planning.
5. **User Feedback:** Our system incorporates mechanisms for users to provide feedback on the accuracy of predictions and overall satisfaction with the platform. This feedback loop enables continuous improvement and refinement of the predictive models, ensuring their relevance and effectiveness in real-world agricultural settings.

6. Performance Measure :

Performance metrics such as Accuracy and Cross-Validation Score (CV Score) play crucial roles in evaluating the effectiveness of machine learning models. **Accuracy** measures the proportion of correctly predicted instances out of the total instances in the dataset, calculated as the ratio of the number of correct predictions to the total number of predictions made by the model. While Accuracy provides a straightforward assessment of the model's overall correctness, it may not always be the most reliable metric, particularly in scenarios with imbalanced classes.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

For, binary classification:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

On the other hand, the Cross-Validation Score (CV Score) is a technique used to assess the model's performance by splitting the dataset into multiple subsets, training the model on a subset, and testing it on the remaining subsets. The CV Score typically represents the average accuracy obtained across all folds in a cross-validation process. This approach helps mitigate issues such as overfitting and offers a more robust estimate of the model's performance on unseen data.

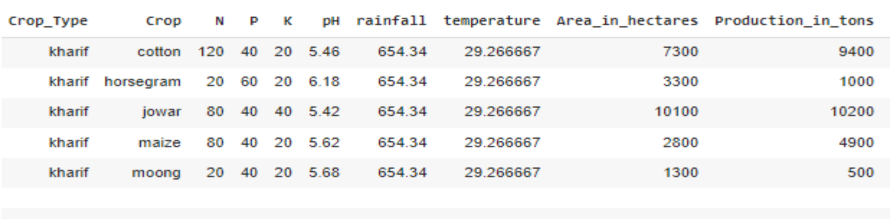
Common cross-validation techniques include k-fold cross-validation and stratified k-fold cross-validation, each providing unique advantages depending on the dataset characteristics. Both Accuracy and CV Scores are vital for evaluating machine learning models, with Accuracy offering a direct measure of correctness and CV Scores providing insights into the model's generalization capability. Additionally, other metrics such as precision, recall, and F1-score may be employed for a more comprehensive evaluation, especially in scenarios with imbalanced classes.

4 . IMPLEMENTATION

A. Data Description:

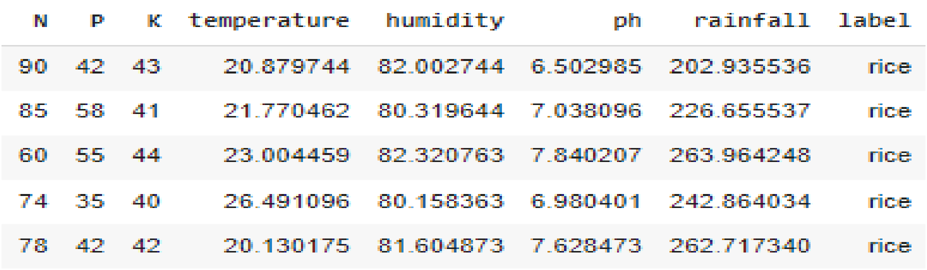
To implement the system, we decided to use a Kaggle dataset.

The following figures are the snapshots of the datasets that have been used for this project.



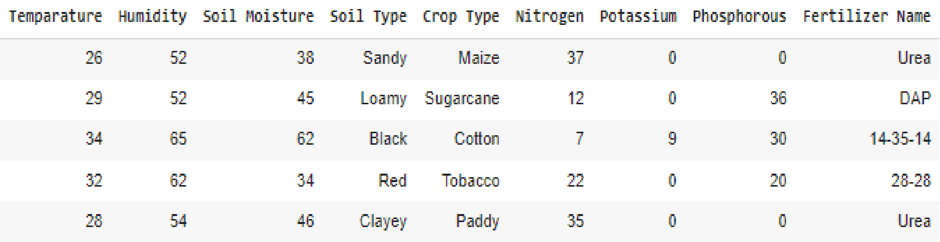
Crop_Type	Crop	N	P	K	pH	rainfall	temperature	Area_in_hectares	Production_in_tons
kharif	cotton	120	40	20	5.46	654.34	29.266667	7300	9400
kharif	horsegram	20	60	20	6.18	654.34	29.266667	3300	1000
kharif	jowar	80	40	40	5.42	654.34	29.266667	10100	10200
kharif	maize	80	40	20	5.62	654.34	29.266667	2800	4900
kharif	moong	20	40	20	5.68	654.34	29.266667	1300	500

Fig. 4.1. Dataset For Crop Yield Prediction Problem



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

Fig. 4.2. Dataset For Crop Prediction Problem



Temparature	Humidity	Soil Moisture	Soil Type	Crop Type	Nitrogen	Potassium	Phosphorous	Fertilizer Name
26	52	38	Sandy	Maize	37	0	0	Urea
29	52	45	Loamy	Sugarcane	12	0	36	DAP
34	65	62	Black	Cotton	7	9	30	14-35-14
32	62	34	Red	Tobacco	22	0	20	28-28
28	54	46	Clayey	Paddy	35	0	0	Urea

Fig 4.3. Dataset For Fertilizer Prediction Problem

```
dataset = tf.keras.preprocessing.image_dataset_from_directory(
    "PlantVillage",
    seed=123,
    shuffle=True,
    image_size=(IMAGE_SIZE,IMAGE_SIZE),
    batch_size=BATCH_SIZE
)
```

```
Found 4000 files belonging to 4 classes.
```

```
class_names = dataset.class_names
class_names
```

```
['Tomato___Bacterial_spot',
 'Tomato___Early_blight',
 'Tomato___Late_blight',
 'Tomato___healthy']
```

```
first image to predict
actual label: Tomato___Early_blight
1/1 ━━━━━━━━━━━━━━━━━━━━ 1s 747ms/step
predicted label: Tomato___Early_blight
```

Fig. 4.4. Dataset For Plant Disease Prediction

B.Implementation process:

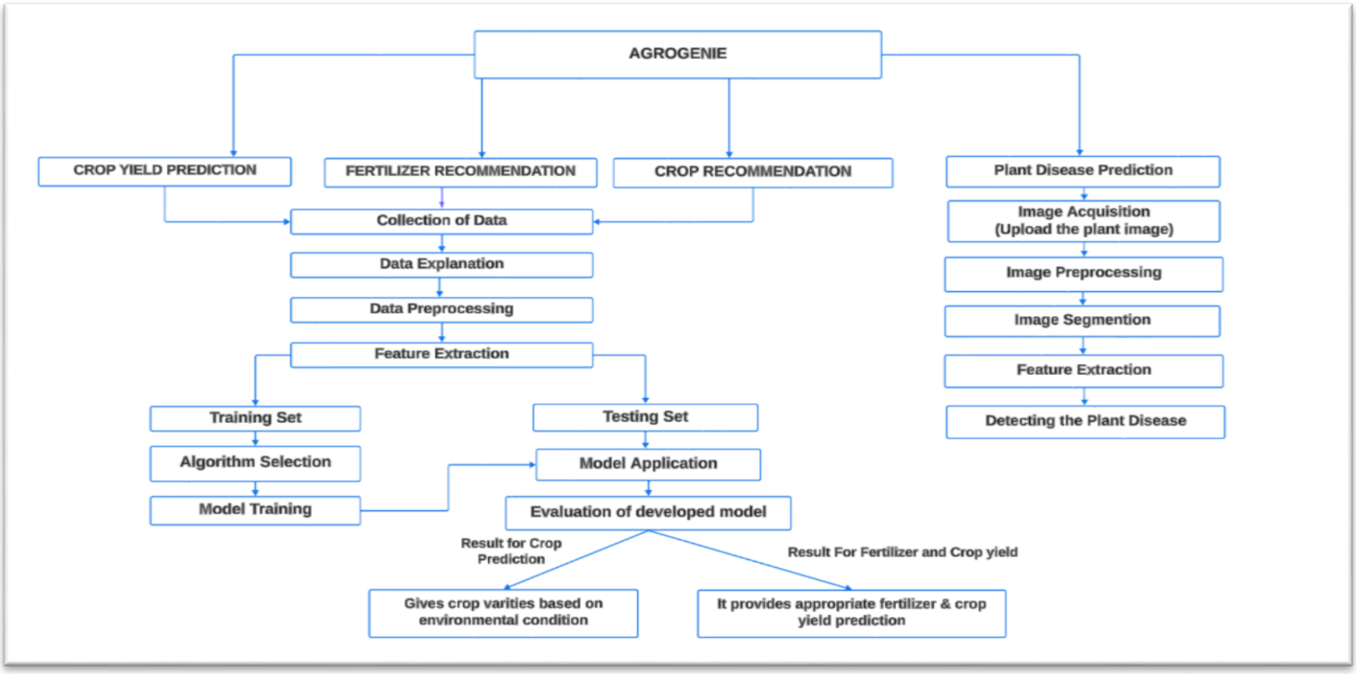


Fig 4.5.Implementation process

Crop Prediction:

Develop classification models (e.g., decision trees, random forests, support vector machines) to predict the most suitable crops for cultivation based on environmental factors such as climate, soil, and geographical location.

Fertilizer Recommendation:

Build regression models (e.g., linear regression, SVM) to provide personalized fertilizer recommendations based on soil nutrient levels, crop requirements, and environmental conditions.

Plant Disease Detection:

Implement computer vision algorithms (e.g., convolutional neural networks) to detect signs of disease or pest infestation in crops using images of plant leaves or other plant parts.

Crop Yield Prediction:

Develop time-series forecasting models (e.g., autoregressive integrated moving average, long short-term memory networks) to forecast crop yields for upcoming seasons based on historical yield data and environmental variables.

Integration of Features:

Integrate the developed models and features into a unified system or web application. Ensure seamless interaction between different modules to provide a cohesive user experience. Implement user authentication and authorization mechanisms to ensure data privacy and security.

User Interface Development:

Design an intuitive and user-friendly interface for the web application, featuring interactive visualizations, and dashboards.
Offer explicit guidance to users on implementing responsive design principles to ensure compatibility across diverse devices and screen dimensions.

Evaluation and Validation:

Evaluate the performance of the developed models using metrics such as accuracy, precision, recall, and F1-score.
Validate the models on separate test datasets or through cross-validation to assess their generalization ability.
Conduct sensitivity analysis to evaluate the robustness of the models to changes in input variables or model parameters.

Iterative Improvement:

Gather feedback from users and stakeholders to identify areas for improvement in the system. Continuously monitor model performance and update the system with new data or improved algorithms. Stay abreast of advancements in machine learning techniques and agricultural research to incorporate the latest innovations into the system.

By following this detailed process, the system aims to empower farmers with advanced tools and insights to optimize agricultural practices, increase productivity, and enhance sustainability.

5. RESULT AND ANALYSIS

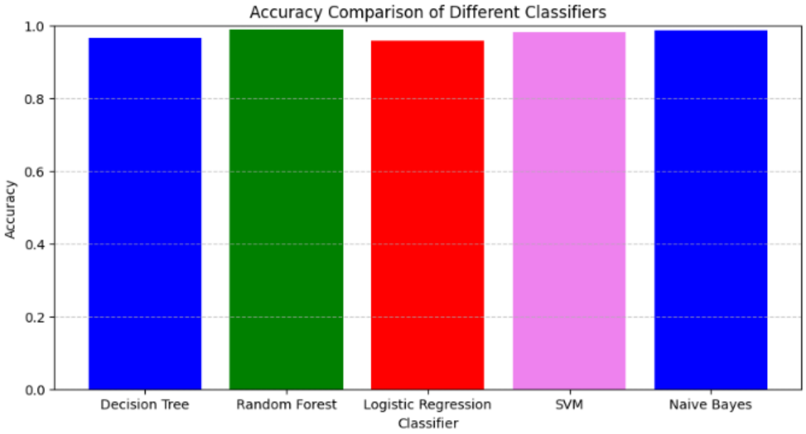
1. Crop prediction Model:

For the first module which is a classification problem, we built the following models and evaluated their performance.

Table 5.1:Comparison between various regressors

Classification Model	Accuracy
SVM	98.02
Naïve Bayes	98.587
Random Forest	98.870
Decision Tree	96.610
Logistic Regression	95.762

The accuracy of the Random Forest model surpassed that of alternative algorithms such as decision trees, support vector machines, and logistic regression, indicating its superior performance in crop prediction tasks.

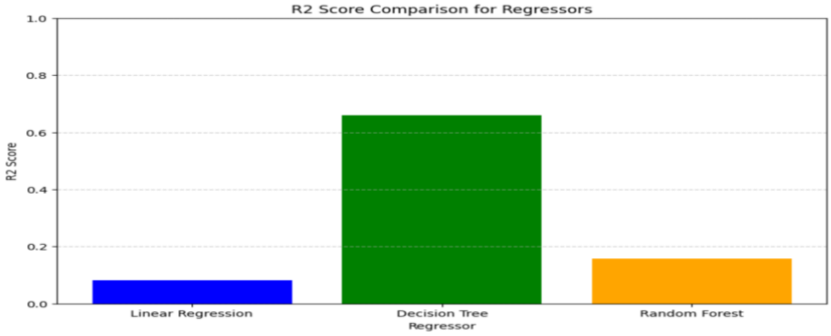


2. Crop yield prediction:

For the Second module which is a regression problem, we built the following models and evaluated their performance.

Table 5.2: Comparison between various regressors

Regression Model	R2 Score
Linear Regression	0.08187
Decision Tree Regressor	0.6587
Random Forest regressor	0.1566



Based on the graph, we can conclude that decision tree regressors have the highest R2 score, followed by random forest regressors and then linear regression. This suggests that decision tree regression performs better at fitting the data compared to the other two algorithms in this particular case.

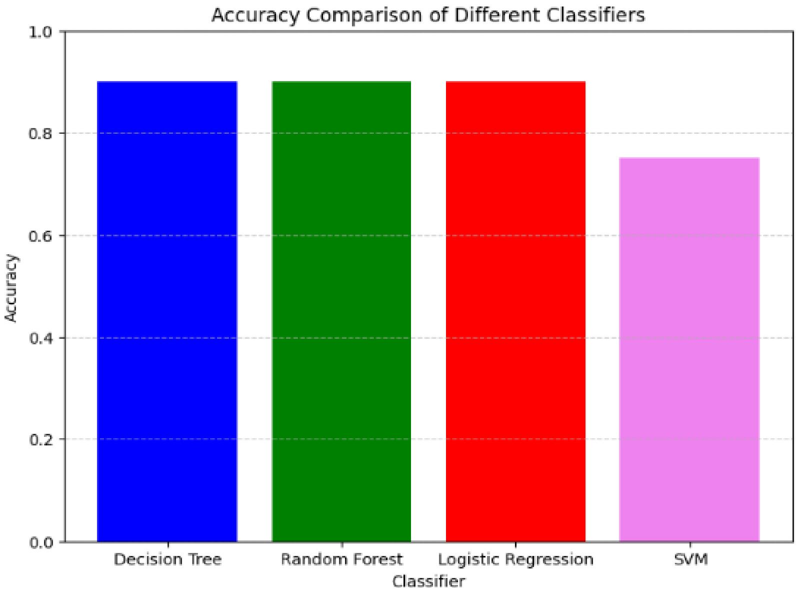
3. Fertilizer Recommendation Model:

For this module, focused on solving a classification problem, we constructed several models and assessed their effectiveness.

Table 5.3: Comparison between various regressors

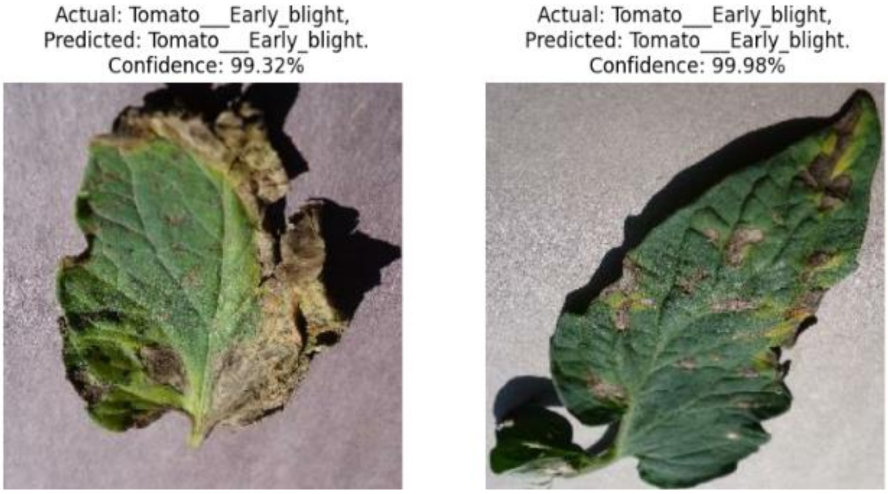
Classification Model	Accuracy
SVM	75.0
Random Forest	90.0
Decision Tree	90.0
Logistic Regression	90.0

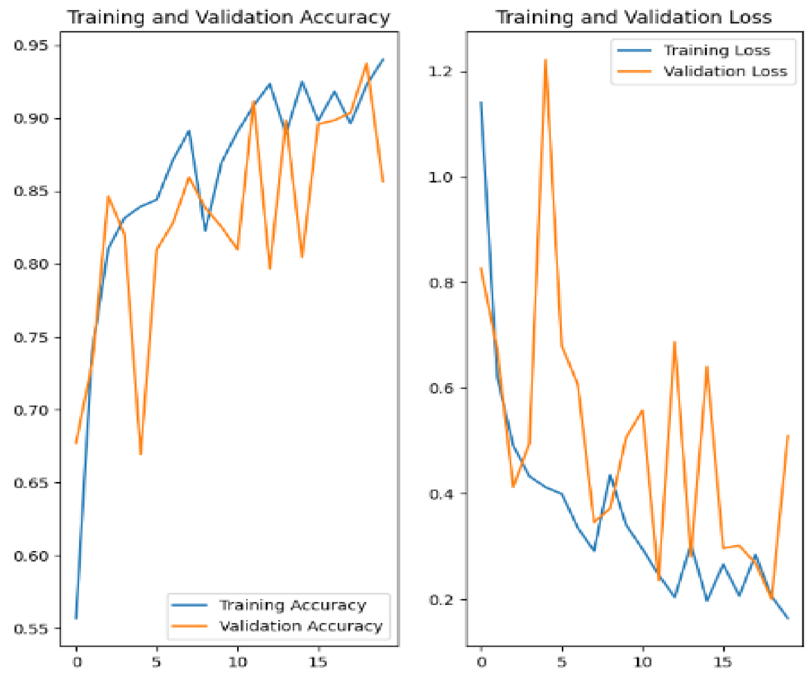
Linear Regression, Random Forest, and Decision Tree models demonstrated high accuracy in crop yield prediction, further investigation is necessary to understand SVM's performance and determine the most appropriate model for practical applications in agriculture. Linear Regression, Random Forest, and Decision Tree models all achieved an accuracy of around 90% in crop yield prediction, while SVM (Support Vector Machine) lagged, which suggests interesting findings. This is demonstrated with the help of the graph:



4. Plant Disease (CNN) Model:

We develop and deploy a plant disease prediction model using CNN with Keras and TensorFlow, helping to identify and mitigate crop diseases effectively.





An accuracy of 90% with a CNN model for plant disease prediction is impressive and indicates the effectiveness of the model in accurately identifying diseased plants.

6. CONCLUSION

The integration of advanced technologies like machine learning and data analytics holds immense promise for revolutionizing smart farming practices across various domains, including crop yield prediction, crop prediction, fertilizer recommendation, and plant disease detection. Through sophisticated predictive models and algorithms, farmers can harness actionable insights to optimize agricultural operations, mitigate risks, and enhance productivity and sustainability. Crop yield prediction models leverage historical data and environmental factors to forecast yields accurately, empowering farmers to make informed decisions regarding resource allocation and crop planning. Crop prediction algorithms utilize climatic conditions and soil characteristics to recommend suitable crops for cultivation, enabling farmers to diversify their crop portfolio and maximize returns. Fertilizer recommendation systems leverage soil nutrient analysis and crop requirements to provide personalized fertilization strategies, optimizing nutrient uptake and minimizing environmental impact. Plant disease detection models utilize computer vision and machine learning techniques to identify signs of diseases or pest infestations early, facilitating timely intervention and crop protection. By leveraging these advanced technologies, smart farming initiatives can drive transformative change in agriculture, fostering sustainable practices, improving resource efficiency, and ensuring food security for future generations.

REFERENCES

[1] P.Priya, U.Muthaiah. Balamurugan.Predicting the yield of the crop using a machine learning algorithm. International Journal of Engineering Science Research Technology.

[2]. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics (ICCCI)

[3]D Ramesh, B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology

[4]. Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E. Butler, Soo-Hyung Kim. Random Forests for Global and Regional Crop Yield Predictions. Institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America.

[5] Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks, Bin Liu 1,2,*,† Yun Zhang 1,†, DongJian He 2,3 and Yuxiang Li 4, Symmetry 2018, 10, 11; doi:10.3390/sym10010011.

[6].. Rashmi Priya, Dharavath Ramesh.2018." Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model".

[7]. Igor Oliveira, Renato L. F. Cunha, Bruno Silva, Marco A. S. Netto.2018." A Scalable Machine Learning System for PreSeason Agriculture Yield Forecast."

[8]Renato L. F. Cunha, Bruno Silva, Marco A. S. Nett. A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast. 2018 IEEE 14th International Conference on e-Science (e-Science)
"data.gov.in", [online] Available: https://data.gov.in/.

[10]Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal Jaypee Institute of Information Technology, India. Crop Yield Prediction Using Machine Learning Algorithms

[11] Ramesh Medar, Vijay S. Rajpurohit, Shweta Shweta KLS GIT, Belagavi. Crop Yield Prediction using Machine Learning Techniques

[12] 'District Wise rainfall data'Available: http://hydro.imd.gov.in/hydrometweb/(S(ctep2djbwexxp0qr1knt3a45))/DistrictRaifall.aspx

[13] Suvidha Jambekar, Shikha Nema, Zia Saquib. dept.of Electronics and Communication Usha Mittal Institute of Technology Mumbai, Maharashtra, India. Prediction of Crop Production in India Using Data Mining Techniques

[14]Predict Crop Production in India Using Machine Learning Technique: A Survey Bhavna Sharma, Pratap Singh Yadav, Sunita Yadav

[15]Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique. Rakesh Kumar, M.P. Singh, Prabhat Kumar, J.P. Singh