International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02745
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Agri Ai: Intelligent Plant Disease Surveillance and Predictive Forecasting

Dr. Jagini Naga Padmaja
Department of Artificial Intelligence and Machine Learning
Vardhaman College of Engineering
Hyderabad, India
nagapadmaja_14@vardhaman.org

Sai Asritha Thati
Department of Artificial Intelligence and Machine Learning
Vardhaman College of Engineering Hyderabad, India
Saiasrithathati21aiml@vardhaman.org

Sai Sreeja Bandi
Department of Artificial Intelligence and Machine Learning
Vardhaman College of Engineering Hyderabad, India
bsaisreeja21aiml@student.vardhaman.org

Shalini Tejavath
Department of Artificial Intelligence and Machine Learning
Vardhaman College of Engineering Hyderabad, India
Tejavathshalini21aiml@student.vardhaman.org

Uma Devi R
Department of Artificial Intelligence and Machine Learning
Vardhaman College of Engineering Hyderabad, India
rumadevi21aiml@student.vardhaman.org

Dr. Gagandeep Arora
Department of Artificial Intelligence and Machine Learning
Vardhaman College of Engineering Hyderabad, India
Gagandeeparora250379@gmail.com

***Abstract*—Agriculture is the backbone of many economies, especially in India, as it is a considerable contributor to GDP and employment. Farmers struggle to predict crop yield due to the unpredictable nature of weather conditions, soil quality and other environmental factors. This paper implements machine learning (ML) technologies for yield prediction along with the remedies based on the temperature, rainfall, and soil composition. Different ML algorithms, such as Random Forest, Decision Tree, Neural Networks, and Regression models, are compared in terms of their accuracy. The proposed system mainly aims to help farmers with data-driven insights to optimize crop yield and enhance agricultural productivity.**

Keywords: Machine learning; Random Forest; Yield prediction; Agricultural productivity.

I. INTRODUCTION

For thousands of years, agriculture has been the backbone of human civilization providing food, raw materials, and economic stability for communities globally. Agriculture holds great significance in India, accounting for about 15.4% of GDP and employing a large segment of the population. It is number 1 in the world in net cropped area, followed by China and the United States. Indian agriculture, despite its political importance and economic contribution, has been governed by the rule of capture that has made reforms hard by facing challenges ranging from climate change to soil degradation, water scarcity, and changing market dynamics.

Environmental uncertainties, including unpredictable changes in rainfall, temperature, and fertility, pose one of the biggest challenges to agricultural productivity. Crop yield prediction, by virtue, has heavily relied on the experience of farmers combined with statistical models, is another area that has failed to keep pace with rapidly changing climatic conditions. Machine learning techniques are emerging as a transformative approach for tackling them. With vast datasets to draw upon, ML algorithms are capable of detecting patterns, forecasting future yields, and making recommendations for best farming practices.

The use of machine learning algorithms like Random Forests, Neural Networks, and Regression methods also have been explored in agricultural research to increase crop yield predictions. They use historical data-like weather patterns, soil properties, and previous crop yields to create predictive frameworks that aid farmers in making data-driven decisions. These ML models could also leverage data in real -time, which enabled on the fly reactive measures based on current environmental conditions.

In crop yield prediction, proper selection of input features greatly improves the accuracy of prediction. These models can lower computational and the few parameters have high predictive power such as soil moisture, nitrogen 1 and temperature fluctuations by identifying the most relevant parameters.

This study compares different feature selection techniques and ML algorithms to find the best technique for predicting crop yield and suggesting remedies.

This study aims to be a bridge between technology and agriculture using an intelligent decision support system. This research aims to help farmers to optimize crop yield, minimize loss and promote sustainable farming practices through actionable recommendations by merging several ML algorithms together with feature extracting methods. The proposed system will be evaluated on metrics like predictive accuracy, computational efficiency, and real-world applicability.

II. LITERATURE SURVEY

With the projected demands of food security, advanced technological solutions have become essential in order to maximize crop yield. Machine Learning(ML) techniques have been increasingly used in agriculture and especially in crop yield prediction (CYP) and disease management (Ghosal et al.2023; Rinaldi et al.2021; Zhao et al.2023). Several studies confirmed the ability of ML algorithms to predict agricultural yield based on environmental and soil conditions

1. Crop Yield Prediction using Machine Learning

Predicting the crop yield is critical in resource allocation, risk management, and strategic planning. Various ML-based yield prediction methods have been reported in some studies.

The Random Forest (RF) algorithm was first proposed by Brieman [1], which uses the ensemble method to improve the accuracy of prediction by combining multiple decision trees. Balamurugan [2] developed RF for CYP using data from historical rainfall, temperature, and soil quality. However, this approach did not run up against a deep learning model, so it was not well-suited for dynamic execution environments.

Mishra et al. [8] performed a separate important study For example, [3] investigated several ML models including Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN). While a good deal of theoretical analysis was performed, the study did not apply these models to real-world datasets, so practical validation is lacking.

Another more recent study by Shruti Kulkarni et al. [4] described the use of Neural Networks for predictive analysis of crop yield. This long Short-Term Memory

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02745
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

(LSTM) model was able to pick up long-term weather and soil properties dependencies and it also yielded superior accuracy results compared to traditional ML models. But CNN-based techniques take up a lot of compute power, which might not be available to small-scale farmers.

2. Agricultural ML Models: Feature Selection

By minimizing redundant variables, feature selection allows ML models to save resourcing, allowing computation to only occur for relevant features. Multiple studies have used PCA, RFE, and GA to identify features that are most impactful to CYP.

PCA-based feature selection, for example, was shown to reduce correlated predictors while maintaining relevant prediction accuracy and improving overall prediction performance. Analogously, RFE was also used on regression-based CYP models and found to improve generalization performance.

An additional research study used ensemble learning methods (XGBoost, GBM, and Bagging). Ensemble models combined several weak learners to provide improved robustness, thereby outperforming standalone such as Decision Trees.

3.Prediction and Management of Crop Disease

Aside from CYP, cinnamon ML has also won a role in disease monitoring and preempted diagnosis of vegetation diseases. Traditional disease identification methods are based on expert consultation, which is usually not available in rural areas. Deep learning has been successfully used to help automate the detection and identification of diseases and pests affecting plants.

It reached an accuracy of over 90% in identifying different types of crop diseases using AlexNet and Inception-ResNet-v2 for the classification of different diseases. To perform such a holistic assessment, researchers are integrating disease prediction models into CYP systems and vice versa with the aim of creating decision-support tools for farmers.

Stated differently, even though these models are advanced, there are still limitations including but not limited to low generalizability across diverse climate conditions, sparseness of data, and absence of real-time deployment and standards. To mitigate these limitations, the proposed research combines feature selection, ensemble learning, and deep learning techniques to create a scalable, high-accuracy system for predicting crop yield and recommending remedies.

III.METHODOLOGY

In this study, we followed a structured ML pipeline such as data collection, preprocessing, feature selection, model selection, training, evaluation, and recommendation generation. This project aims to construct an accurate and scalable machine learning model for crop yield prediction and give remedial recommendations on the basis of factors which are environmental and agricultural.

1. DATA COLLECTION

Data is gathered from various sources to maintain high accuracy and robustness:

- Data on historical crop yields from government agricultural departments and research institutions.
- Weather agencies providing meter
- Weather agencies providing meteorological information like temperature, rain fall, humidity and wind speed.
- Soil health reports on pH, nitrogen(N), phosphorus(P), potassium(K), organic carbon and moisture.
- Cloud-based data is the information gathered by remote sensing satellites and IoT-based sensors for immediate environmental supervision.
- Farmer documentation of traditional practice and knowledge.

2. DATA PREPROCESSING

Raw data can be inconsistent, contain missing values, noise , and use different measurement units. These steps will help you clean and prepare your input data:

- HANDLING MISSING DATA

Mean/Median Imputation: This could be used for numerical variables (like temperature, and pH levels).

KNN Imputation: For categorical features like soil type.

- NORMALIZATION AND SCALING

Min-Max scaling: Apply on temperature, rainfall, soil nutrient values to maintain a uniform range (0-1).

Standardization: used where the features that were normally distributed.

- FEATURE ENCODING

One-Hot Encoding(OHE): Effectively used to represent categorical variables effectively such as crop type and soil texture in numerical format.

Label Encoding: Applied on binary categorical data i.e. presence of pest(y/n).

- OUTLIER DETECTION AND REMOVAL

Interquartile Range(IOR) Method: Outliers for rainfall and temperature datasets are identified.

Z-Score of Validation: Validates that the distribution of data is within acceptable limits.

3. FEATURE SELECTION

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02745
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

This is the case where a selection of features and methods is necessary to optimize the accuracy and computational effort of the model. Different feature selection methods are employed:

- FILTER-BASED METHODS

Pearson Correlation Coefficient: Features that are very correlated can lead to duplication; these features can be dropped.

Mutual Information score: It calculates the dependence or closeness of the crop yield with each feature.

- WRAPPER-BASED METHODS

Recursive Feature Elimination (RFE): It recursively removes the least significant features based on the performance of the model.

- EMBEDDED METHODS

LASSO Regression: It is a Regularization technique that penalizes fewer important features.

Random Forest Feature Importance: Ranks the list of features based on their contribution, using decision tree approach.

4. DEVELOPING A MACHINE LEARNING MODEL

- MODEL SELECTION

In order to have a complete analysis, many machine learning models are trained and compared:

- Traditional Models:

Multiple Linear Regression (MLR): Settle a trend for yield prediction

Decision Tree (DT): Captures non-linear dependencies in data.

- Ensemble Learning Models:

Random Forest (RF): Combines several decision trees to minimize variance.

Gradient Boosting Machines (GBM): Improving the weak learner sequentially.

XGBoost(Extreme Gradient Boosting): An optimized gradient boosting package designed to be highly efficient, flexible, and portable.

- Deep Learning Models:

A family of machine learning which imitates human decision-making using layers of multiple individual neural networks.

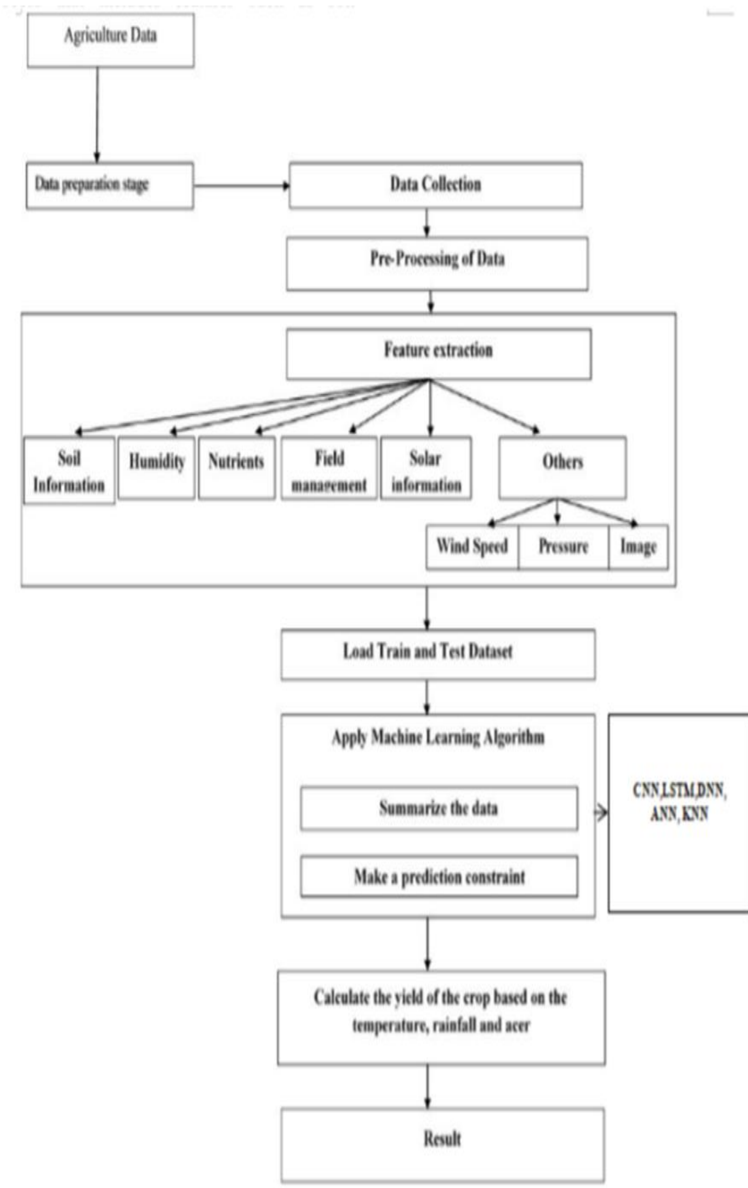
Long Short-Term Memory (LSTM) Networks: Best for time-series agriculture yield prediction

- HYPERPARAMETER TUNING

Each of these models is fine-tuned for optimal performance:

- Grid Search and Random Search for hyperparameter tuning.
- Cross-validation (k-fold CV) for improving model robustness.

MODEL DIAGRAM:



International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02745
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

IV. RESULTS

- CROP PREDICTION

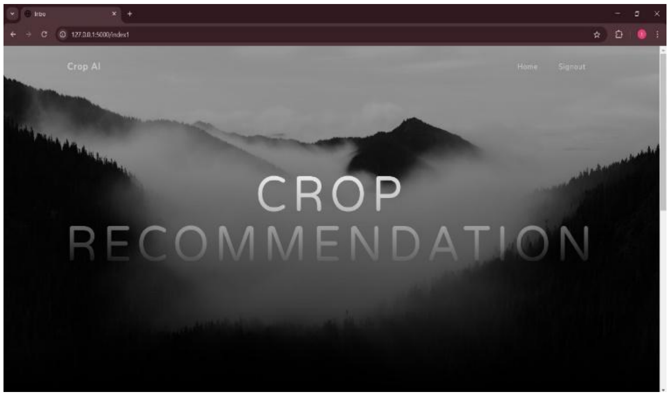


Figure 2. Crop Recommendation Page

Crop Prediction is the process of using AI and machine learning to forecast crop yields based on environmental factors like soil quality, temperature, rainfall, and past farming data. It helps farmers plan better by predicting which crops will grow best under given conditions. By analyzing historical and real-time data, crop prediction minimizes risks, improves productivity, and supports sustainable farming practices.

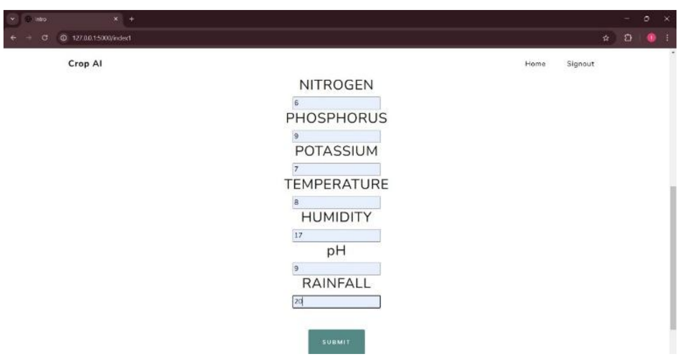


Figure 3. Testing Soil Moisture

Testing soil moisture is the process of measuring the water content in the soil to ensure optimal conditions for crop growth. It helps farmers manage irrigation efficiently, preventing both overwatering and drought stress. Common methods include moisture sensors, the gravimetric method, tensiometers, and digital meters. Proper soil moisture testing improves water conservation, enhances crop yield, and supports sustainable farming practices.

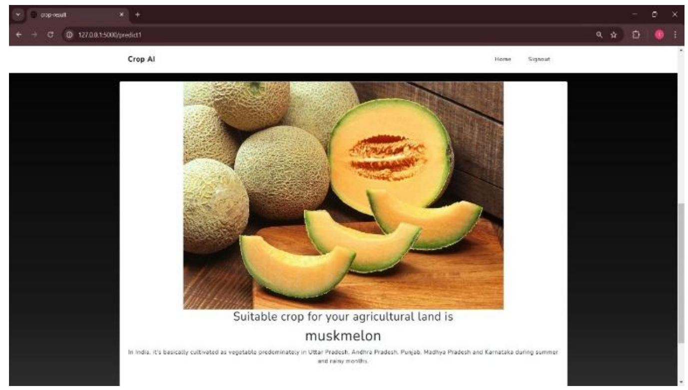


Figure 4. Crop Recommendation

A **Crop Recommendation System** uses AI and data analysis to suggest the best crops for cultivation based on factors like soil type, weather, temperature, rainfall, and nutrient levels. It helps farmers make informed decisions to improve yield and sustainability. Machine learning models analyze historical data and environmental conditions to provide accurate recommendations, reducing risks and optimizing agricultural prediction.

- YEILD PREDICTION

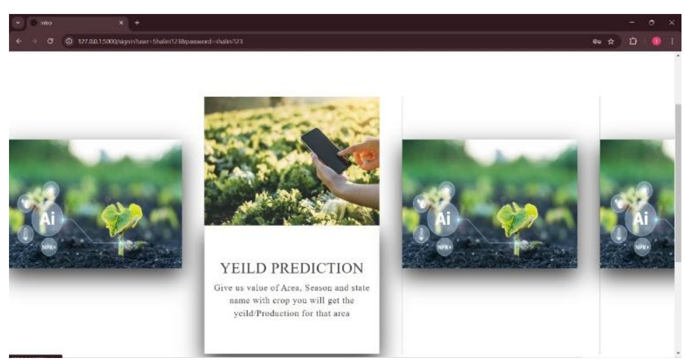


Figure 6. Crop Yeild Prediction

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02745
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Crop yield prediction is the process of estimating the expected agricultural output using AI, machine learning, and historical data. It considers factors like soil quality, weather conditions, rainfall, and farming practices to forecast yields accurately. This helps farmers optimize resources, plan better, and improve productivity while reducing risks.

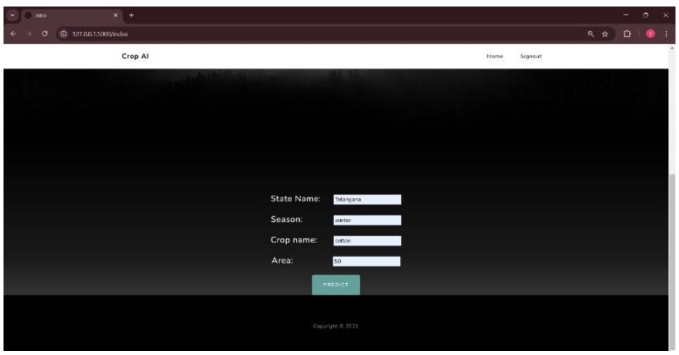


Figure 7.Yeild Prediction Testing

Yield prediction testing evaluates the accuracy of models used to estimate crop production. It involves testing machine learning algorithms with real-world data on soil quality, weather conditions, and farming practices. This process helps improve prediction accuracy, enabling farmers to make informed decisions for better resource management and higher productivity.

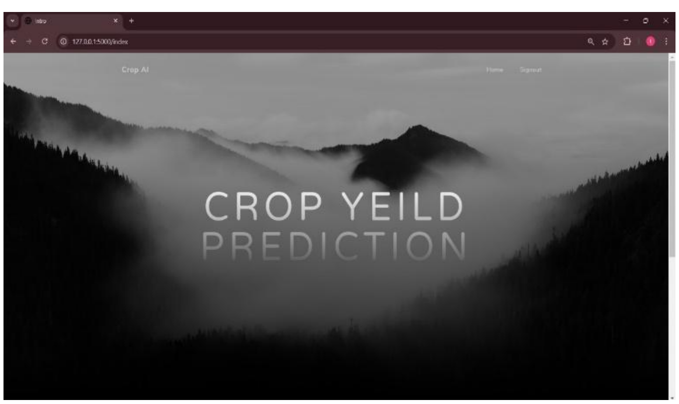


Figure 8: Crop yield prediction

Crop yield prediction is the process of estimating the amount of crop that will be produced in a given area during a specific growing season. It involves analyzing factors like weather, soil conditions, irrigation, and farming practices. Advanced techniques such as machine learning, satellite imagery, and historical data are often used to make more accurate predictions. These predictions help farmers optimize resources, improve crop management, and ensure food security by forecasting supply

Prediction Rate refers to the accuracy or efficiency of a model in forecasting outcomes based on input data. It is commonly used in machine learning, weather forecasting, agriculture, finance, and other fields. The prediction rate is influenced by factors such as data quality, model complexity, and environmental variables. A high prediction rate indicates reliable and precise forecasting, while a low rate suggests the need for model improvement.

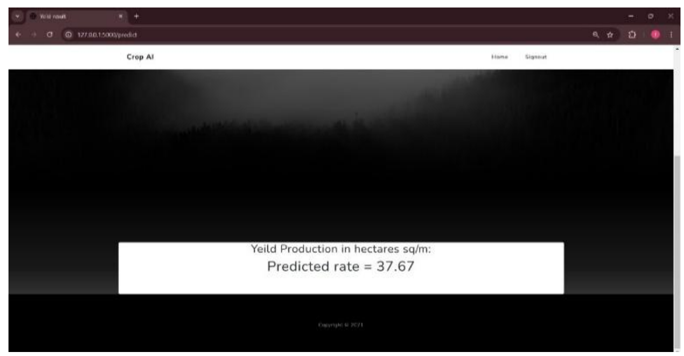


Figure 9: prediction rate

- PLANT DISEASE DETECTION

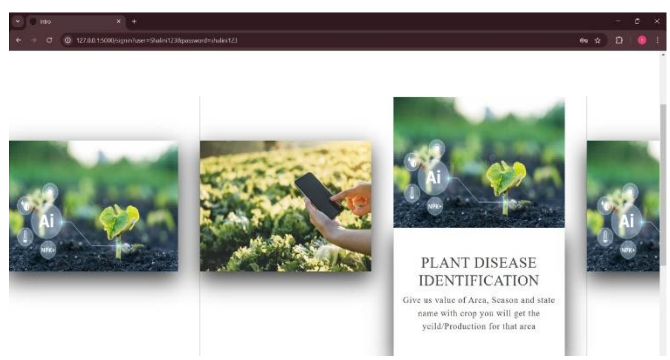


Figure 9. Plant Disease Identification

Plant Disease Identification is the process of detecting and diagnosing diseases in plants caused by pathogens like fungi, bacteria, viruses, or environmental factors. It involves visual

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02745
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

inspection, laboratory testing, and advanced technologies such as image processing, machine learning, and remote sensing. Early and accurate identification helps in effective disease management, reducing crop losses and ensuring better agricultural productivity.

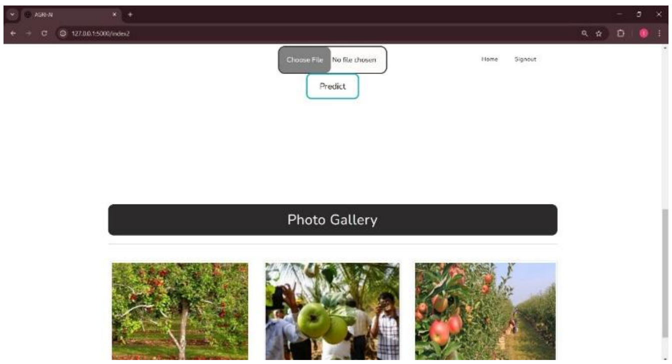
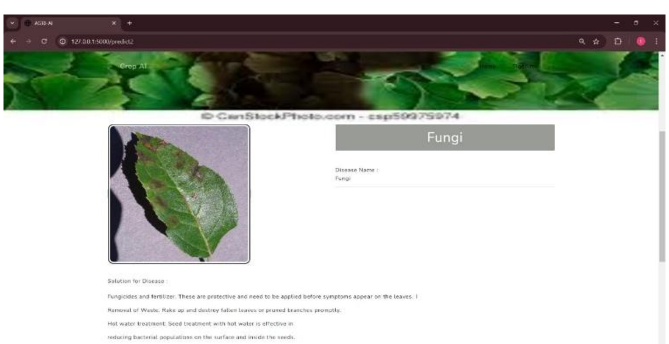


Figure 10: upload photo

Figure 11: soil moisture monitoring

Soil Moisture Monitoring is the process of measuring the water content in soil to ensure optimal conditions for plant growth. It helps in efficient irrigation management, preventing overwatering or drought stress. Methods include **sensor-based monitoring**, **remote sensing**, and **gravimetric analysis**. Accurate soil moisture data improves crop yield, conserves water, and supports sustainable agriculture.



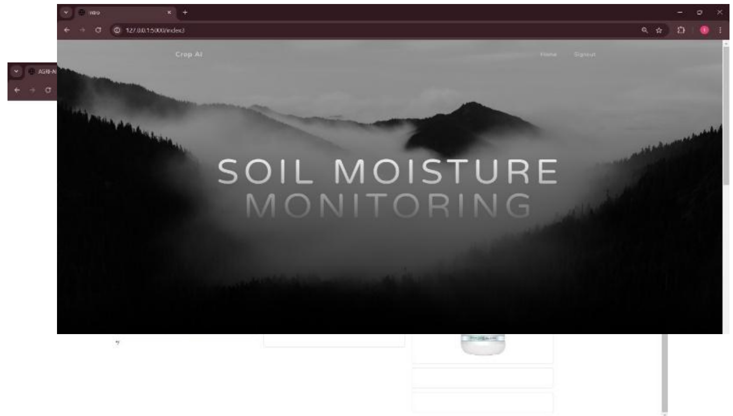


Figure 12. Identification Of Plant Disease

Figure 12. Product Recommendation

Product recommendation suggests items to users based on their interests, browsing history, and purchase behavior. It improves customer experience and increases sales using techniques like machine learning and filtering methods.

- SOIL MOISTURE MONITORING

Figure 13: testing

Soil Testing Monitoring is the process of analyzing soil properties to assess its nutrient levels, pH, moisture, and contamination. It helps in optimizing fertilizer use, improving crop productivity, and maintaining soil health. Regular monitoring ensures sustainable farming and better land management.

Figure 14: result

V. CONCLUSION

The present research discussed the variety of features mainly dependent on the data availability. Each research will investigate CYP using ML algorithms that differ from the features. The features were chosen based on the geological position, scale, and crop features and these choices were mainly dependent upon the data-set availability, but the more features usage was not always giving better results. Therefore, finding the fewer best-performing features were tested that also have been utilized for the studies. Most of the existing models utilized Neural networks, random forests, KNN regression techniques for CYP, and a variety of ML techniques were also used for best prediction. From the studies, most of the common algorithms used were CNN, LSTM, and DNN algorithms but still improvement was required further in CYP. The present research shows several existing models that consider elements such as temperature, and weather conditions, performing models for effective crop yield prediction.

REFERENCES

[1] Crop Yield Prediction using ML Algorithms, 2019 Fifth International Conference on Image Information Processing (ICIIP),2019|[1] Aruvansh Nigam, Saksham Garg, Archit Agrawal

[2] {plantum: googable} Leo Brieman, Random Forests h Australia, 8/13, b.0140-4873. 8 {cite: im} Journal of Biomedical Science and Engineering - MAKING THE MOST OF LIFE AS WE KNOW IT; AND HERE!! In: Machine Learning 45 (2001) 1, 5-32, tps://www.scirp.org/(S(czeh2tfqw2orz553k1w0r45))/journal/home.aspx?journal id=30" \t "_blank" Journal of Biomedical Science and Engineering, blank" Vol. 9 No.5, 2001

[3] Priya, P., Muthaiah, U., Balamurugan, M.Predicting Yield of the Crop Using Machine Learning Algorithm, Volume 9 Issue 4, April 2020 http://www.ijsr.net2015 www. ijsr. net2015

[4] Applications of machine learning techniques in agricultural crop production[4] Mishra, S., Mishra, D., Santra, G. H.Indian Journal of Science and Technology 10.17485/ijst/2016/v9i38/95032 volume 9,2016
Mandal, G Srivatsa Sharma, “Predictive Analysis to Improve Crop Yield using a Neural Network Model” IEEE-2018

[5] Shruti Kulkarni, Shahnawaz Mandal, G Srivatsa Sharma, “Predictive Analysis to Improve Crop Yield using a Neural Network Model” IEEE-2018

[6] Xu G., Zhang F., and Shah S.G. along with others looked at how leaf color images can help spot tomatoes that lack nitrogen and potassium. This was published in Pattern Recognition Lette rs in 2011

[7] n another study, Vibhute A and Bodhe SK. explored how image processing can be used in agriculture. Their survey appears in the International Journal of Computer Applications from 2012.

[8] Andy Robinson, Assistant Professor and Potato Extension Agronomist, Adjunct NDSU/University of Minnesota Gary Secor, Professor, Plant Pathology Department, NDSU Neil Gudmestad, University Distinguished Professor, Plant Pathology Department, NDSU

[9] Punajari, J.D., Yakkundimath, R., Byadgi, A.S., ‘Image Processing Based Detection of Fungal Diseases In Plants’, International Conference on Information and Communication Technologies, Volume 46, pp. 1802-1808, 2015

[10] Huang K.Y.: ‘Application of artificial neural network for detecting phalaenopsis seedling diseases using color and texture features’, Comput. Electron. Agric., 2007, 57, (1), pp. 3–11