INTEGRATED CROP PROTECTION SYSTEM

[1]VAISHNAVI M, [2] KUSUMA B, [3] YAMUNA A K

Guide: Serin V Simpson, Assistant Professor CSE Department

Computer Science and Engineering, Presidency University, Bengaluru

Abstract

Farming is the backbone of human survival as it provides food for the increasing population. However, the sector faces problems such as plant diseases, bad selection of crops, and improper use of fertilizers that lead to lower yields and economic loss. We will introduce an AI-driven web app that integrates plant disease detection, crop suggestions, and fertilizer recommendations. The system applies advanced machine learning to analyse data while providing actionable insights to the farmers and researchers. In addition, it has an easy interface for them to navigate and implement plans of solution. The module for plant disease detection finds infections before they get severe via image processing, thus reducing losses through infections. The crop recommendation system suggests suitable crops that can be grown in consideration of environmental and soil data, resulting in better yields and sustainability for farmers. The fertilizer module provides data-driven guidance on optimal usage, minimizing waste and damage. It improves agricultural decision-making and boosts sustainability because of the proper predictions that help boost productivity, reduce risks, and ensure success for farmers.

I. INTRODUCTION

AgroDoc is an AI-driven platform designed to help farmers and researchers make informed agricultural decisions. It uses advanced image processing and machine learning to offer key functionalities:

1. Plant Disease Detection: Identifies early-stage plant diseases through image analysis, enabling timely intervention.

2. Crop Recommendation: Suggests ideal crops based on soil type, history, and environmental conditions to maximize yield.

3. Fertilizer Suggestions: Recommends fertilizers tailored to soil nutrients and With its user-friendly interface, AgroDoc simplifies farming decisions, enhancing productivity, cutting costs, and promoting sustainability.

II. RESEARCH ELABORATION

The study resolves farming problems by incorporating the ML models and datasets into a single decision support system. Conventional techniques are mainly based on human labor and experience, and the contemporary systems are very fragmented; for example, there are some specific systems for disease detection or crop recommendations.

Data Sources

- Public Data: Crop-leaf images labeled with diseases and healthy samples.
- Historical Agriculture Data: Soil nutrients, weather, and rainfall to suggest crop recommendations.
- Fertilizer Dataset: The optimum NPK levels of commonly grown crops.

ML Models:

1. CNN Model based on ResNet9

For plant disease classification with 95% validation accuracy

Trained on PlantVillage dataset with data augmentations

2. Random Forest

For crop recommendation with 90% accuracy

Features include soil nutrients, pH, temperature, humidity, and rainfall

Other Integration

- OpenWeatherMap API to give real- time weather data, which can be used for recommendations

This system provides a comprehensive platform for farmers to detect diseases, give crop recommendations, and fertilizer suggestions all in one go with high accuracy and time saving.

III. SYSTEM ANALYSIS

The research has recognized the greatest problems that the farmers are facing in agriculture and, hence, devised an integrated, internet-based system to tackle them.

1. Plant Disease Detection

This uses ResNet9 CNN to detect diseases through scanning leaf images.

Gives prescriptions on treatments; it saves time and ensures no more spreading of disease.

2. Crop Recommendation

Scan for nutrients in soil (N, P, K), pH, rainfall, temperature, and humidity.

o Gives correct crops via a Random Forest model to ensure maximum yields and healthy soils.

3. Fertilizer Tips:

Balances soil nutrient levels with the best crop values. No Suggests a balanced fertilizer for growth and a preservation-friendly environment.

Characteristics of the Application Has features of disease detection, crop recommendation, and fertilizer advice.

Has features that make it user-friendly and easily scalable to various technical knowledge

Has real-time information from weather reports for accuracy in application recommendations.

This system automates significant agriculture functions, supports sustainable processes, and helps farmers apply scientific decision-making to better the productivity and minimize the loss in the farms.

IV. REQUIREMENT ANALYSIS

The system requirements are divided into:

Functional Requirements:

1. **Plant Disease Detection:**

○ Users upload plant leaf images.

○ ResNet9 identifies diseases and suggests treatments.

2. **Crop Recommendation:**

○ Users provide soil and environmental data (N, P, K, pH, rainfall, etc.).

○ Random Forest recommends the best-suited crops.

3. **Fertilizer Suggestions:**

○ Users input soil nutrient values and crop type.

○ The system advises on fertilizer type and quantity for optimal growth.

Non-Functional Requirements:

1. **Scalability:**

○ Handles large datasets and supports many users.

2. **Accuracy and Speed:**

○ Models ensure high accuracy and fast predictions with minimal delay.

3. **User-Friendly Interface:**

○ Intuitive design for all users, with clear inputs and outputs.

4. **Reliability:**

○ Robust performance with error handling for diverse scenarios.

This system integrates advanced ML models with a user-centric design for efficient and reliable farming support.

V. SYSTEM DESIGN

Disease Detection: An Overview

Motivation:

Agriculture is an essential component of a country's development, especially in India. The sector fuels the economy and provides employment. Indian farmers have issues like selecting the right crop for the soil and proper time for disease protection, which causes crop losses.

Objective:

Early and accurate detection of plant diseases using machine learning prevents disease spread, minimizes crop loss, and ensures reliable results.

Plant Leaf Disease:

Plant diseases interfere with normal plant functions, either by means of infectious agents (such as fungi, bacteria, or viruses) or by way of non-infectious elements (like temperature, humidity, or soil pH).

Determinants of Pathogenesis 1.Temperature
Optimum temperatures favor development of spores of most pathogens. Severe conditions can obscure symptoms, making diagnosis difficult.

2. Humidity

3. High humidities facilitate spore germination and fungal development. Diseases in greenhouses can often be controlled through humidity management.

4. Soil pH :

Soil acidity or alkalinity influences disease occurrence, such as potato scab being controlled at pH < 5.2.

5. Disease Triangle:

Infectious diseases need:

- A susceptible host
- A virulent pathogen
- Favorable environmental conditions

Breaking this triangle is the essence of prevention.

Types of Diseases

1. Infectious Diseases:

Caused by pathogens like fungi, bacteria, nematodes, viruses, and viroids.

- Fungi cause most plant diseases, affecting all major crops.
- Nematodes damage roots, reducing plant resistance and vigor.

2. Non-Infectious Diseases:

Results from environmental conditions such as poor pH, extreme temperatures, or pollution that tend to weaken the plants and cause secondary infections.

Importance of Identification

Disease identification helps in effective management and optimal utilization of resources. Incorrect identification leads to ineffective treatment and additional crop loss.

Machine learning-based detection systems ensure reliable and efficient management of plant diseases and are beneficial to farmers and the agricultural ecosystem.

VI. IMPLEMENTATION AND RESULTS

This project involves developing an AgroDoc system that detects diseases, recommends crops, and provides fertilizer suggestions using machine learning and deep learning techniques. The purpose is to make agricultural practices efficient and accessible for farmers.

1. Disease Detection

• Objective: Plant disease detection using leaf images using CNN- based models such as AlexNet and LeNet-5.

Method:

PlantVillage Dataset (augmented) contains 87,000 images of crops with 38 disease classes.

Train pre-trained models (e.g., VGG- 16, ResNet-50) for better performance.

Use LIME for interpretability by highlighting key segments responsible for predictions.

2. Crop Recommendation

• Objective: Recommend optimal crops based on soil and environmental conditions.

• Dataset: Kaggle Crop Recommendation Dataset with features like Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, and rainfall.

Models Used:

Decision Tree, Naive Bayes, SVM, Logistic Regression, Random Forest, XGBoost.

Use 5-fold cross-validation to ensure reliability.

•Output: Suggests the best crop based on soil and weather conditions.

3. Fertilizer Suggestion

• Objective: Recommend fertilizers to optimize soil nutrients and reduce environmental harm.Dataset: Custom dataset with features like crop type, N, P, K, pH, and soil moisture.

Method:

Rule-based classification based on NPK levels.

Suggests one of six fertilizers depending on nutrient deficiency or excess.

Key Features:

1. User-Friendly Interface: Accessible for non- experts via a web app.

2. The Actual Predictions: gives prompt feedback using Flask API over disease, crop and fertilizers

3. DFDs: Showing a diagrammatic representation to comprehend input, processing and outcomes

Tools and Frameworks:

1. Deep Models: CNNs (AlexNet, LeNet-5 and VGG- 16)
2. ML Models: Dec-Tree, SVM and Rand-Forrest, XGBO with
3. Frameworks:Tensor-flow, Keras Flask -API, LIME
4. Platforms: kaggle, Google Coloab, DigitalOcean For Production.

This system uses the power of advanced AI techniques for improving yield, diagnosing diseases, and promoting sustainable farming practices for farmers.

VII. SYSTEM STUDY AND TESTING

Crop Recommendation

The results for our crop recommendation experiments are shown in the Table I. The Fig. 5 also depicts these scores on a bar-chart for easy comparison. We can see that the RandomForest and Naïve Bayes models perform the best, followed by
the XGBoost model. It is expected that boosting (RandomForest) and bagging (XGBoost) models will usually perform and generalize better than non- ensemble methods.

TABLE I: Accuracy Comparison of Crop Recommendation models

Model Type	5-Fold Cross-Val Accuracy
Decision Tree	0.914
Naive Bayes	0.995
SVM	0.983
Logistic Regression	0.955
Random Forest	0.995
XGBoost	0.992

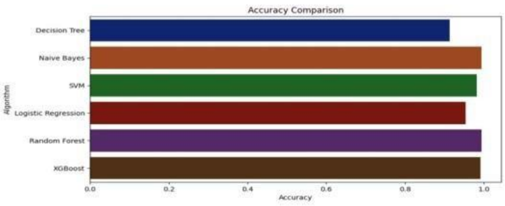


Figure 5: Accuracy Comparison of Crop Recommendation models

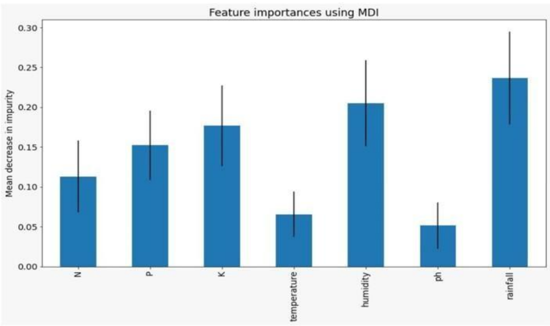


Figure 6: Feature importance for Crop Recommendation RandomForest model

We choose the RandomForest model, which has a cross-validation accuracy of 0.995 for our application because we are able to easily understand the feature importances of the feature used, which tells us how important the features are for our classification . The Fig. 6 depicts the feature importance we calculate from the RandomForest model. We observe that rainfall plays the most significant role in determining the crop type. The second highest importance is that of humidity, followed by K, P and N. This means that overall water content is of the highest importance, followed by soil quality. Thus, using this model, we are able to also understand which features are overall important to our model for crop recommendation.

Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Crop Disease Detection

After getting the optimal validation scores for VGG, ResNet and EfficientNet, we conclude that the EfficientNet model performs the best of all. Fig. 7 shows the accuracy comparison for the three models, where this is apparent. We also plot accuracy and loss curves for the three models in Fig. LABEL:val_loss_curve_pdd. From the accuracy curves, we observe that EfficientNet reaches the highest score very soon, compared to the other two models. ResNet reaches a score close to EfficientNet, but much later in its training. VGG is a smaller model, which could be why it fails to learn the data well and does not perform as well as the other two models. The EfficientNet is a class of several state-of-the-art convolutional neural network based models which are created using a structured approach for scaling neural networks. They have been pretrained on the ImageNet dataset, and achieve the highest score of all the previously designed CNN-based models on ImageNet. For these reasons, we choose the EfficientNet model and deploy it on our application to perform leaf image classification.

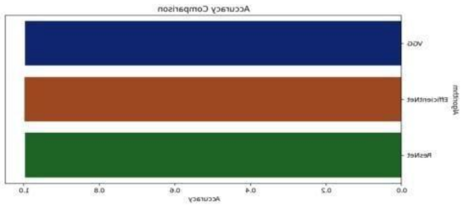
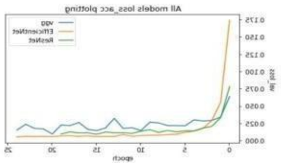


Figure 7: Accuracy Comparison of Plant Disease Detection models



VIII. CONCLUSION

This paper introduces the 'Farmer's Assistant', which is a user-friendly web application system that uses machine learning and web scraping to provide some of its features, like crop recommendation with the help of the Random Forest algorithm, fertilizer recommendation with a rule-based classification, and crop disease detection through images of leaves by using the EfficientNet model. From our user interface, the user can input their information through forms, and fast results will be fetched from there. We also employ the use of LIME to explain our predictions of disease detection images. This would in return help to explain the predictions to users on why our model makes certain predictions and using this, to possibly enhance datasets and models.

Our application is running fine, but there is still room for improvement. In order to make recommendations on crops and fertilizers, we should let them know where they can obtain it from popular shopping websites and perhaps enable the user to purchase crops and fertilizers right from our application.

We also gather all the information based on various types of brands and products associated with N, P, and K values. Till the present date, we have only six types of recommendation types available; however, in the future scenarios, we plan to enhance it further with high-performance machine learning systems which can give excellent and improved suggestions.

The dataset we have used in our paper for the classification of diseases has to be incomplete.

This means that our model would only work well on images that belong to classes it's aware of. It fails to tell us what correct class would be for data outside those classes. To be fixed in future releases, there are two methods through which this can happen. One possible solution is finding datasets with similar sizes and containing different crop types or diseases. The other alternative is making or modifying these datasets using generative modeling and then incorporating them into our training set. This will therefore enable our model to work well with different scenarios.

The second is to allow the users to upload their images by creating a portal in our web application where they can do the labeling themselves. This study shows that LIME explanations can be misleading, as these carry only the local information about an instance and not what a model emphasizes on its global level. We can attempt approaches like Grad- CAM, or Integrated Gradients, etc or we will take approaches like training inside of sparse linear layers augmented with LIME just to explain the model even better at certain prediction spots.

Finally, it's also interesting to present detailed segmentations of the sick part of the dataset.

Currently, this is not something to be done as data has not been available like that so far.

In our application, we can add a tool for the segmentation of annotations wherein users can help us in completing the missing parts. We can also employ some unsupervised algorithms in order to find diseased areas in the image. We are planning on incorporating these features and correcting such errors in our next work.

IX. FUTURE ENHANCEMENT

Some possible additions in the future versions are that enhance functionality, performance, and the user experience of the system. These are:

E-Commerce Connection: Connect the platform to online shopping websites so that users can purchase the suggested crops, fertilizers, and other farming products directly on the platform.

Extended Fertilizer Recommendations: Add fertilizer recommendations based on NPK values for specific brands and offer farmers information about products available commercially.

Improved Somatic Testing: Make the disease classification model better by inputting more datasets with the types of crops and kinds of diseases.

Use generative modeling approaches to augment training data and improve generalization of the model.

User annotation portal: Make a website where users can upload and mark images of sick crops. The dataset will be improved this way.

Advanced Interpretability Techniques: GradCAM and Integrated Gradients can be used to generate better explanations for disease predictions. It will make the system clearer and more reliable.

Segmented Disease Detection: Using detailed segmentation, it can locate and tag the diseased parts of the plant image to be treated for specific applications.

Making Web/Mobile Apps: It can make the system into a mobile version so that farmers from rural and remote areas can easily access it.

Real-Time Monitoring: Use IoT sensors in getting actual soil and environmental data so the crop advice and fertilizer advice will become more authentic.

X. REFERENCES

1. D.Jayanarayana Reddy, Dr. M. Rudra Kumar, “Crop Yield Prediction using Machine Learning Algorithm”, 2021.
2. R, Harish D, Priya B, “A Machine Learning- based Approach for Crop Yield Prediction and Fertilizer Recommendation”, 26th August 2022.
3. Suruliandi, G. Mariammal & S.P. Raha, “Crop prediction based on soil and environmental characteristics using feature selection techniques”, 12 Mar 2021.
4. Nishu Bali & Anshu Singal, “Deep Learning Based Wheat Crop Yield Prediction Model in Punjab Region of

North India", 16 Sep 2021.

5. Reyana, Sandeep Kautish, P. M. Sharan Karthik, Ibrahim Ahmed Al-Baltah, Muhammed Basheer Jasser, (Member, IEEE) And Ali Wagdy Mohamed," Accelerating Crop Yield: Multisensor Data Fusion and Machine Learning for Agriculture Text Classification", 27 February 2023.

6. Alexandros Oikonomidis, Cagatay Catal & Ayalew Kassahun, "Deep learning for crop yield prediction: a systematic literature review", 06 Feb 2022.

7. Sujatha M., Jaidhar C.D, "Machine learning- based approaches to enhance the soil fertility", 14 November 2023.

8. P K Devan, Swetha B, Uma Sruthi P, Varshini, "Crop Yield Prediction and Fertilizer Recommendation System Using Hybrid Machine Learning Algorithms", 14th March 2023.

9. Mamatha, J.C. Kavitha c, "Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques",

10. 10 January 2023. 10. Chetan R, D.V. Ashoka, Ajay Prakash B, "IMLAPC: Interfaced Machine Learning Approach for Prediction of Crops", 11 January 2022.

11. Simraneet Kaur, Getanajali Babar, Navneet Sandhu, Dr. Gagan Jindal, „Various Plant disease detection using Image Processing Methods", June 2019.

12. Er. Varinderjit Kaur,Dr. Ashish Oberoi ,

„Wheat disease detection using svm classifier" , Aug 2018.Barbedo, „Plant Disease Identification from individual lesions spots using deep learning", Apr 2019.

13. K. Naga Subramanian, A.K. Singh, A. Singh, S. Sarkar and Ganpath Subramanian, "Usefulness of interpretability methods to explain deep learning-based plant stress phenotyping", July 2020.

14. W. Yang, C. Yang, Z. Hao, C. Xie, and M. Li,

„„„Diagnosis of plant cold damage based on hyperspectral imaging and convolutional neural network,""" IEEE Access, vol. 7, pp. 118239–118248, 2019.

15. K. Nagasubramanian, S. Jones, A. K. Singh, S. Sarkar, A. Singh, and B. Ganapathy subramanian, „„„Plant disease identification using explainable 3D deep learning on hyperspectral images".

16. N. Zhang, G. Yang, Y. Pan, X. Yang, L. Chen,and C. Zhao, „„„A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades,""" Remote Sens., vol. 12, no. 19, Sep. 2020, Art. no. 3188.

17. J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, „„„Using deep transfer learning for image-based plant disease identification,""" Comput. Electron. Agricult., vol. 173, Jun. 2020, Art. no. 105393.

18. M. Agarwal, A. Singh, S. Ajaria, A. Sinha, and

S. Gupta, „„„ToLeD: Tomato leaf disease detection using convolution neural network,""" Procedia Comput. Sci., vol. 167,
pp. 293–301, Jan. 2020.

19. G. Hu, H. Wu, Y. Zhang, and M. Wan, „„„A low shot learning method for tea leaf's disease identification,""" Comput. Electron. Agricult., vol. 163, Aug. 2019, Art. no. 104852.