International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Driven Crop Disease Prediction and Management System

Hima S, Varsha S, Poojitha S
CSE, Presidency University, Bengaluru, Karnataka, India
Guided By: Mr.T Ramesh, Assistant Professor, SOE, Presidency University
hima.20211cse0567@presidencyuniversity.in varsha.20211cse0576@presidencyuniversity.in
poojitha.20211cse0632@presidencyuniversity.in

ABSTRACT

There are severe agricultural challenges faced by most countries such as India, because of the outbreaks of plant diseases and climate change, which jeopardize the livelihoods of millions dependent on agriculture. A web application has been designed to ameliorate this situation by giving real-time crop recommendation suggestions based on environmental parameters such as temperature, humidity, soil nutrients, pH, and rainfall. The use of AI and ML for plant disease management and crop yield prediction follows a path of transformation toward helping in accurate, data-driven decision-making for agriculture. Seven machine learning models, namely Decision Tree, Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, XGBoost, and k-Nearest Neighbors (KNN), were used to test and validate the system.
Besides, the project also incorporates a plant disease identification system based on a Convolutional Neural Network (CNN); it scans leaf images for the high-accuracy detection and classification of plant diseases.
This means farmers will be able to act early, preventing crop loss and enhancing agricultural productivity.

KEYWORDS: CNN, Crop Recommendation, Machine Learning, Plant Disease Identification, Random Forest

I. INTRODUCTION

To a great extent,this could be attributed to artificial and machine intelligence which have integrated agriculture into almost all its activities, thus making it an industry considered by many as significantly benefitting from AI and ML. Examples of recent breakthrough developments are intelligent crop recommendation systems and plant disease detection. Agriculture has already been transformed by machine learning and data science with the inception of
image-based classification systems apt to recognise plant diseases accurately and offer tailor-made recommendations to farmers relative to crops they should plant.

The advanced Crop Recommendation Solution is well integrated into a proposed system done in Python with some machine learning algorithms that can perform classification works to assist farmers to make informed decisions that can bring a rational act of growing a specified plant while considering parameters like soil nutrient status, season, and geography. The farmer uses data input such as rainfall, location, and soil concentration of potassium, phosphorus, and nitrogen. It predicts crop suitability for different regions using huge datasets with several features and numerous classification models. This expert knowledge will quickly assist the farmer in better crop selection and finally minimize crop failure risk along with increased productivity in agriculture.

Also, new feature of plant disease detection in the System using image classification algorithms. Formerly, these enable the Convolutional Neural Networks, one of the deep learning techniques employed for the analysis and processing of images. This prevents loss in crops by uploading photos of diseased symptoms in crop leaves. After uploading the photos, the plant can identify itself and detect possible diseases from the millions of labeled samples to which the image is compared.

This paper employs an easy web interface developed using Streamlit to allow for easy access and interactivity of the Crop Recommendation System to farmers and agricultural professionals. It also allows them, through the interface, to input the required data and get personalized crop suggestions as well as give insight on crop health and probable diseases that plants may have.
This paper is organized as follows: Section 2 gives a review of the relevant literature, while Section 3 discusses the proposed system. Experimental results are shown in Section 4, while Section 5 concludes the study with a summary of key findings, a presentation of the methods behind the efficient crop recommendation and disease prediction system, and suggestions for future development.

II. RELATED WORKS

Somehow to be very smart farming techniques are one of the most important parameters in transforming agriculture characterized by higher productivity, better resources management systems, and controlling crop disease. Intelligent crop recommendation systems along with plant disease identification methods have been elaborated in this paper. This would allow for timely diagnosis of plant diseases and empower the farmer with conscious decisions to check the spread of the disease and minimize crop loss, which ultimately gives a big relief to the farmer from large-scale losses because of stochastic factors [11].

By S. M. Pande et al. (2021), ML is the very backbone for predicting and managing crop diseases. Some of the algorithms for yield prediction and suggestion of suitable crops are based on certain environmental and geographical conditions, such as Decision Trees and K-Nearest Neighbor (K-NN). All for the maximization of productivity and profitability [10].

Decision trees classify and predict crops according to environmental factors like temperature, moisture, and soil pH. The image representation makes decision trees more fit to assess suitability for certain crops and in decision-making [3][10].
With its ensemble learning approach, random forests successfully resolve the overfitting concern of decision trees while enhancing predictive ability. It thus empowers decision- making in highly heterogeneous agricultural systems with respect to inputs for crop yield prediction and resource management [6].

Convolutional neural networks functions just like the human visual system and do extremely well in identifying plant diseases based on images. They help in identifying visible symptoms of diseases, such as discoloration and injury as a result of pests, by either in-situ images or by acquired drone images, and enable almost real time diagnosis [2][4][7].

AI and deep learning have further enhanced the efficiency of plant disease detection.
Deep Learning Models such as CNNs form a true game-changer for disease detection due to the analysis of huge datasets of labeled crop images. Apps like DeepCrop allow farmers to upload infected plants' pictures and get immediate diagnostic feedback [2][5].
Bouguettaya et al. (2023) state that UAV-Based Monitoring concerns drones provided with high-resolution cameras for aerial observations. On the contrary, images are processed to check for signs of disease symptomatology through deep learning models like YOLO and Faster R-CNN, although issues such as image resolution and operational costs still pose challenges [7].

The web-based crop recommendation platforms essentially fortify the farmers with practical insights.

Responsive Design ensures availability across multiple devices, including PCs, tablets, providing great usability for the end-user. For instance, DeepCrop is easy to navigate, with intuitive UX for any end-user that doesn't possess any kind of technical training [2].

Real-time data integration is a true asset for live weather data, as it gives higher recommendation

III. PROPOSED SYSTEM

The actual implementation phase is among the most important stages where development takes place for the framework of plant disease identification and smart crop recommendation. This is where the theoretical concepts and methodologies discussed in previous sections get translated into practical applications [10]. The implementation phase hypothetically exhibits a systematic development of the system supplemented with rigorous testing regimes and analytical methods in order to verify conformity of the system at earlier-defined standards. In order to guarantee reliability as well as effectiveness of the system, this section outlines methodology, testing and validation approaches, results analysis, and quality assurance mechanisms employed [8]

A. Methodology 1.Datasets

Crop recommendations data set contains 2,200 records, having eight attributes: temperature, humidity, pH, rainfall, nitrogen, phosphorus, potassium, and a target label. While temperature, humidity, and rainfall are indicators of environmental conditions, the NPK values indicate soil fertility. The label identifies the best crop to be cultivated based on those inputs, especially soil pH values.

For plant disease detection, an auxiliary dataset contains 70,295 high-resolution images of diseased plants, approximately 5 GB in size. The images were standardized to 128x128 pixels and represent 38 classes, including 26 of disease types and 14 of plant species. A model was established according to the training and assessment processes; other images were, therefore, utilized for validation of detection performance.

2. Data Preprocessing

Data preprocessing: schemes involved in the transformation of raw data into something understandable for analysis and machine learning. In crop recommendation, preprocessing involves handling the issue of missing values, normalizing features, and other tasks that are beneficial for building accurate models.

Image preprocessing for disease detection involves resizing all incoming images to a size of 128x128 pixels, augmentation methods such as flipping, rotation, and changing brightness, and some de-noising techniques that help improve the image consistency while acting against model overfitting [4][7].

3. Train-Test Split

The scikit-learn function train_test_split() is used to assess model performance. In the case of crop recommendation datasets, the split is done with 80% (i.e., 1,760 records) allocated for training and 20% (i.e., 440 records) for testing. In this way, 80% training and 20% validation splits are done also for the plant disease dataset [2][7].

4. Model Development and Prediction Prediction is where the trained model is deployed on new data and makes its forecast. The output is generated on the test set using the predict() function.

The Random Forest Classifier is used for crop recommendation because of its proven robustness and accuracy in agricultural forecasting, with the ability to manage very complex interactions between input features yielding reliable results [6].

Identification of plant diseases implements a Convolutional Neural Network (CNN). The architecture consists of an input layer to take the images, Convolution layers that extract spatial features, Pooling layers that reduce the dimensionality of features to prevent over-fitting, and Fully connected layers which classify the extracted features into the different categories of diseases [2][4][5].

5. Classification Report and Confusion Matrix

The classification report and confusion matrix from the Scikit-learn metrics module are used to assess the performance of the model.

Confusion matrix indicating counts of true positives, false positives, false negatives, and true negatives.

Precision:

Precision = TP / (TP + FP)

Recall measures the ability of the model to identify all relevant positive cases as follows:
Recall = TP / (TP + FN)

The F1 score is given by:
F1 = 2 * (Precision * Recall) / (Precision + Recall)
An F1 score close to 1 is considered excellent performance.

6. Accuracy
Accuracy is one of the basic criteria for measuring the efficiency of a model and is defined as a ratio of the number of correct predictions versus total predictions, which can be expressed mathematically as:
Accuracy=(TP+TN)/(TP+TN+FP+FN)
The implications of this measure are implemented in crop recommendation and disease detection models for authenticating prediction reliability [1][5][6].

B. Model Evaluation
Performance evaluation of the plant disease detection model is by holding out 20% of the image dataset for validation, which is used for assessing the model's precision so that it accurately recognizes disease characteristics. The evaluation demonstrates that the model satisfies accuracy standards and works aptly in real-world situations.

C. Integration and Deployment of Systems
Some of the components proposed for the functioning of the suggested system could be integrated into a web-based platform:
1. Crop Recommendation-an interactive and responsive interface via which users are allowed to input soil and environmental parameters to obtain personalized crop recommendations [2][10].
2. Disease Detection-This feature allows users to upload images of the crops, which are then analyzed instantly to detect plant diseases and suggest viable solutions.
3. Real-Time Data Integration-the integration of data from external sources such as live weather forecasting-to give farmers current and thorough insight into agriculture [2][8].

IV. Results And discussions

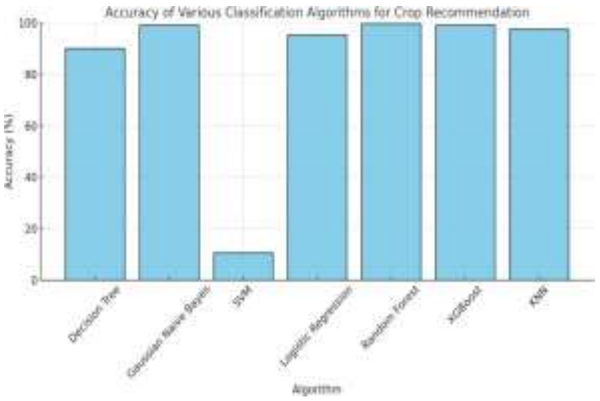
The study evaluated the performance of seven classification algorithms in the crop recommendation task. The results shown in Table 1 reveal with a whopping 99.55% accuracy how Random Forest surpassed all other models. This indeed makes it highly dependable and robust for accurately forecasting relevant crops under varying conditions.

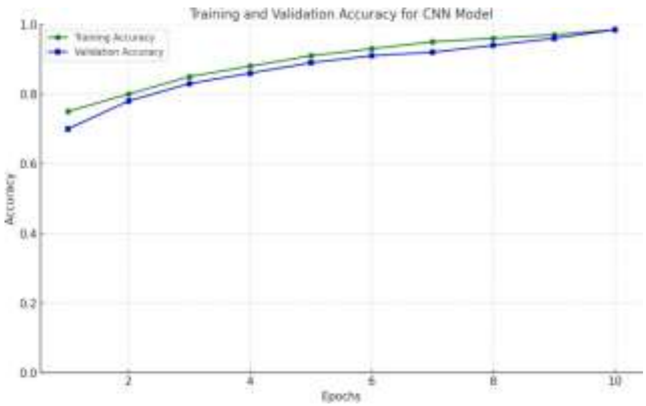
Table 1. Accuracy Across Classification Algorithms

Algorithm	**Accuracy (%)**
Decision Tree	90.00
Gaussian Naive Bayes	99.09
Support Vector Machine (SVM)	10.68
Logistic Regression	95.23
Random Forest	99.55
XGBoost	99.09
K-Nearest Neighbor (KNN)	97.50

As stated in the above table, the Random Forest model always outperforms the other algorithms when different types of crops are involved. The performance of certain models like XGBoost and Gaussian Naive Bayes were similar but varied with the crop. Support Vector Machine Out of this condition, in which SVM performed better on the other side by showing an accuracy of 10.68%, is most probably due to its sensitiveness to the large and complex dataset. Overall study validates the robustness and versatility of Random Forest for crop recommendation. With these properties, it can have a place in agricultural applications for its comparative stability and versatility denoting strong and consistent results.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930





B. Plant Disease Detection Using CNN

In a research study, CNN was applied to diagnose diseases in plants using a total of 17,572 images featuring pictures of 38 classes of diseases as well as healthy plant cases. These images were reshaped to 128 by 128 pixels and then split into training and validation sets using the `tf.keras.utils.image_dataset_from_directory` method.The CNN model to predict was a pre-trained one loaded from a .keras file[5].

During training, a dropout layer is used as a regularization mechanism to prevent the model from learning very specific patterns by randomly turning off input neurons. To test and visualize the images, an early blight infected potato plant was taken. The image was resized to the required dimension, passed from BGR to RGB grayscale using OpenCV, and then given to the CNN model. The model provided a probability distribution over all classes, and the maximum likelihood class was selected using the `np.argmax` function.The predicted class label, along with the original image, was shown for confirming the accuracy of the disease under consideration.

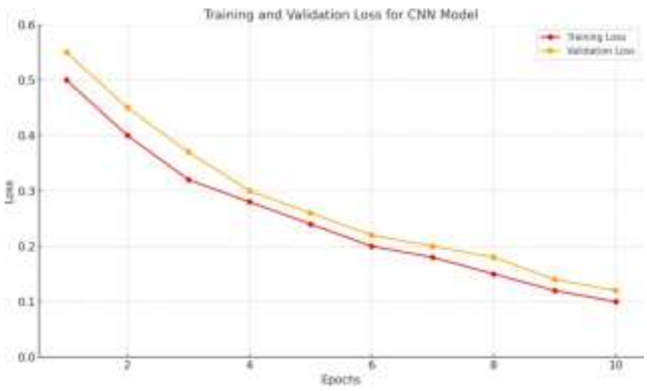
This procedure illustrates how efficient CNN technology can be in automating plant diseases. These subtle features like tiny spots, discoloration, and many other visible indicators within the plant can help make diagnosis more rapid and accurate to timely intervene and improve crop health and management practices.

C. CNN Model Performance Analysis

The CNN based plant disease detection model was evaluated for its performance with the help of measures like training accuracy and validation accuracy as well as training losses and validation losses as displayed in the graph here above.

High accuracy: The model has demonstrated a fairly good discriminative ability of healthy plants from diseased ones, with further evidence of increasing training and validation accuracies. At one point, the validation accuracy soared to an all- time high of 98.5%, clearly indicating that the model is highly robust.

Low Loss: The training and validation loss curves moved very much in parallel downward with only a small divergence indicating the success of training as well as good generalization on unseen data.



It conceals a strong ability to fit even with minor variances in results on the validation set. In addition to covering the capacity for accurate disease diagnosis, it can extract complex features like such as discolorations, leaf spot damage, and wilting patterns. These observations allow us to conclude that CNNs indeed work quite well for plant disease detection and, especially, smart agriculture with increased crop yields**.**

V. Conclusion and Future Work

In abstract, the proposed scheme for this study provides the state of the art established technologies and methodologies in order to establish a hi-tech solution regarding smart crop recommendation and plant disease detection. It generated real-time data processing and machine learning techniques that derive the actionable improvement in managing vegetables for minimizing the dependence of plant diseases. Making the system modular and scalable, as well as following industry standards, ensures the reliability, maintainability, and adaptability of the system.

So the well-structured database and user-centered interface will be very user-friendly and they'd be able to interact with the platform without much technical knowledge. Ease of use encourages seamless interaction between farmers and broader use.

The development in precision agriculture has made strides in using strategic planning, cutting-edge technology, and best practice in this bold move. When accurate diagnostics and intelligent crop recommendations become available to the farmer, that is the sustenance of the data-driven system in farming.

For future developments:
The following enhancements may be undertaken to enrich the system:

- **Update Dataset Regularly:** New region-specific data samples will continuously be integrated to keep the models up to speed and accurate at all times.
- Use of Better Techniques: Reinforcement learning may be added to improve existing models and strategies applied.
- Multispectral and Hyperspectral Imaging: In these lines, superior and early plant health and disease manifestations can be examined.
- Weather and Other Environmental Data Use: In addition to other crop recommendations, it would include weather conditions, soil moisture conditions, and rainfall statistics to improve their precision.
- Profit Estimation Tools: Farmers can then make wise recommendable decisions regarding crops that can maximize their returns and the most profitable crops in relation to study market prices, trends, input costs, and expected returns.
- Guidance to New Farmers: Helping the new farmers by providing tools for crop scheduling, irrigation, fertilization, and pest control for timely and efficient farming. Very limited multilingual input capability.
- Edge Computing Integration: Affordable portable devices integrated with cameras and real-time sensors will be deployed on-site for agricultural monitoring.
- Collaborative Knowledge Platform: Enabling farmers, researchers, and agronomists to establish case studies, data, and insights for continuous improvement of the system's performance.
- Crowdsourcing Data Collection: Bringing all kinds of users to the model to greatly increase its accuracy and relevance.
- AI-Based DSS: Build systems with intelligent decision-making through reinforcement learning that optimize resource use and maximize yield.

This system could make it a very comprehensive assistant at the end, changing the paradigm in which farming will be held in the age of smart agriculture.

VI. References

[1] M. Shripathi Rao, A. Singh, N. V. Subba Reddy, and D. U. Acharya, “Crop prediction using machine learning,” Journal of Physics: Conference Series, vol. 2161, no. 1, p. 012033, Jan. 2022.

[2] Md. Manowarul Islam et al., “DeepCrop: Deep learning-based crop disease prediction with web application,” Journal of Agriculture and Food Research, vol. 14, pp. 100764–100764, Dec. 2023.

[3] A. Kar, N. Nath, U. Kemprai, and Aman, “Performance Analysis of Support Vector Machine (SVM) on Challenging Datasets for Forest Fire Detection,” International Journal of Communications, Network and System Sciences, vol. 17, no. 2, pp. 11–29, Feb. 2024.

[4] G. Rani, E. T. Venkatesh, K. Balaji, Balasaraswathi Yugandher, Adiki Nithin Kumar, and M Sakthimohan, “An automated prediction of crop and fertilizer disease using Convolutional Neural Networks (CNN),” 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Apr. 2022.

[5] D. Tirkey, K. K. Singh, and S. Tripathi, “Performance analysis of AI- based solutions for crop disease identification, detection, and classification,” Smart Agricultural Technology, vol. 5, p. 100238, Oct. 2023.

[6] K. Ramu and K. Priyadarsini, “A Review on Crop Yield prediction Using Machine Learning Methods,” 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Oct. 2021.

[7] A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, “A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images,” Cluster Computing, Aug. 2022.

[8] K. Jhajharia, P. Mathur, S. Jain, and S. Nijhawan, “Crop Yield Prediction using Machine Learning and Deep Learning Techniques,” Procedia Computer Science, vol. 218, pp. 406–417, 2023.

[9] M. Shripathi Rao, A. Singh, N. V. Subba Reddy, and D. U. Acharya, “Crop prediction using machine learning,” Journal of Physics:

[10] S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA, and K. SHAURYA, “Crop Recommender System Using Machine Learning Approach,” 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Apr. 2021.

[11] C. Raju, A. D, and A. P. B, “CropCast: Harvesting the future with interfused machine learning and advanced stacking ensemble for precise crop prediction,” Kuwait Journal of Science, p. 100160, Dec. 2023.

[12] P. Chauhan, Hardwari Lal Mandoria, A. Negi, and R. S. Rajput, “Plant Diseases Concept in Smart Agriculture Using Deep Learning,” Advances in environmental engineering and green technologies book series, pp. 139–153, Oct. 2020.