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

Machine Learning Approach for Plant Disease Detection And Suggestion

Yash Kamthe [1], Dnyaneshwari Garule [2],

Aishwarya Kudapane [3], Akash Khalekar [4]

Mr. Jitendra R. Chavan [5]

[1] Yash Kamthe Information Technology & MMCOE
[2]Dnyaneshwari Garule Information Technology & MMCOE
[3]Aishwarya Kudapane Information Technology &MMCOE
[4]Akash Khalekar Information Technology & MMCOE
[5] Mr. Jitendra R. Chavan, Assistant Professor, Information Technology & MMCOE

---***---

Abstract

Plant diseases are some of the most relevant topics affecting crop yield productivity because they lead to economic losses for farmers and impact food availability. To address such a concern, we developed a system that utilizes deep learning techniques, particularly Convolutional Neural Networks (CNNs), to identify diseases by analyzing images.

Detection is just one aspect of our project. In addition, we prepared responses that provide actionable steps for managing the illness in question. The suggested approach is to design the tool as user-friendly and intuitive for farmers. Additionally, users have the ability to inquire further about the illness and its management. Although this past research already demonstrated the efficacy of CNNs for such problems, our approach integrates an additional layer. It leverages a potent AI model to suggest relevant, real-time treatment options tailored to the situation. This model serves to issue practical recommendations comparable to those given by an expert. This system provides farmers with a sophisticated yet simple method for mitigating risks to their crops and enhancing their decision-making by merging straightforward and useful instructions with the automated identification of plant diseases.

Keywords: Streamlit, CNN, Deep Learning, AI API, Image Classification, Plant Disease Detection, Smart Agriculture.

1. Introduction

The world economy depends heavily on agriculture, and high agricultural productivity depends on maintaining plant health. One of the main causes of large economic losses is plant diseases. Conventional methods of identifying plant diseases depend on farmers or agricultural specialists visually inspecting the plants, which can be laborious, subjective, and unfeasible for extensive monitoring [3]. CNN models have shown better performance in identifying plant diseases from leaf photos thanks to developments in deep learning. When trained on sizable datasets, these models can accurately identify diseases while requiring less human involvement. However, the majority of current systems only classify diseases without offering additional treatment recommendations. We suggest an improved system that combines CNN-based disease detection with AI-driven recommendations in order to overcome this constraint..

2. Ease of Use

Plain and no technical skills required steps for uploading images of leaves have been implemented in our web interface so Everyone can use our sophisticated system. All of the necessary steps to upload the image are intuitive and easy. We undertook all of the required processes previously. This step does not require you to do anything. The image is then uploaded into the model for deep learning through convolutional neural networks, enabling it to in a fast and precise manner determine whether or not the plant is diseased.

Users have the ability to enhance their trust in our system as we allow for immediate feedback on the overall state of the plant post image analysis. The speed at which the feedback is provided during their workings aids build trust enabling users to achieve better results even when the provided information is time sensitive.

Users receive not only the diagnosis made by the system , but by attaching an AI-based module for suggestions we are enabling diagnosed diseases to be well manged as clear steps are provided enabling better control over it.

Interactive experiences were enhanced further through the integration of a chatbot.

3. Datasets in Plant Disease Detection

In creating and testing our own CNN, we accessed the PlantVillage dataset which is well-known in the field of plant pathology. It includes over 54000 labeled images of healthy and diseased plant leaves across 14 crop species and 26 diseases. This dataset comprises images taken under specific settings and provides grayscale, color, and segmented images. For our model, we used only color images. Additionally, we applied preprocessing steps, which included normalization and scaling the image dimensions to 256x256 pixels.

Implementing these data augmentation techniques helped enhance dataset variety and model robustness. Controlled image modifications included slight brightening, zooming up to 20%, random rotations not exceeding 40°, and horizontal flipping. In addition, the data was divided into training (80% of the data) and testing (20% of the data) sets while reserving 10% of the training set for validation.

4.Machine Learning and Deep Learning Approaches

Unlike deep learning models, classic ML techniques like SVM, KNN, and Random Forests rely on rigid features and are less scalable. The most used solutions are CNNs because they feature-extract on their own and are scalable. VGG16, ResNet50, and InceptionV3 are pre-trained CNNs with high accuracy but demanding computational resources. Hassan and Maji (2022) proposed a lightweight CNN consisting of inception layers and depthwise separable convolutions, achieving 99.39% on PlantVillage, 99.66% on a rice dataset, and 76.59% on a cassava dataset. Guo (2023) focused on diversity within the preprocessing and attained 92.23%. Lab images provided by Mohanty et al. (2016) achieved 99.35% accuracy, but real-world images only reached 31%. Singh et al. (2021) reached 98.9% on PlantVillage using Bayesian-enhanced CNN. In our implementation, a scratch-trained custom CNN is built with convolutional layers for feature extraction, max-pooling for dimensionality reduction, dense-layer classification, hidden-layer activation with ReLU, and Softmax for output.
For comparison, we conducted the exercise using custom CNNs, as well as transfer learning with ResNet50 and VGG16.

5. AI-Powered Cure Suggestion Systems

The majority of previous works focusing on applying machine learning techniques for plant disease detection have centered on image-based classification systems. While such systems are typically very accurate at detecting diseases, they frequently lack offering meaningful value or insights to the users, especially the farmers, regarding disease management actions after detection. Thus, there has been some attention to the design of intelligent decision support systems that not only provide diagnosis but also prescribe actions.

As an example, Chaudhary et al. (2021) implemented an expert system on the cloud which provides rule-based treatment recommendations for croplands after disease diagnosis via a CNN-based identification. Their solution, while workable in some limited, contrived situations, has no interface capable of accepting natural language inputs and is built upon an arbitrary framework of fixed logic rules 【6】.

In the same vein, Rathore et al. (2022) proposed a mobile application that integrates CNNs for image recognition and an expert-designed treatment database. While the application was able to provide treatment suggestions tailored to the diseases diagnosed, it lacked context-sensitive adaptability and did not offer conversational interfaces for user dialogue participation through voice or text【7】.

Many reasoning about automated deep learning model implementations for classification task, to the best of our knowledge, there are very few systems that integrate disease detection and responsive dynamic adaptive treatment guidance. Integrating these two concepts with Large Language Models (LLMs), a rapidly emerging

A.Emerging Directions in Smart Agriculture

Recent trends indicate a shift toward AI systems that incorporate:
1.**Recommendation engines**: for disease and fertilizer suggestions
2.**Conversational agents:** for natural language interaction
3.**Voice and multilingual support:** for broader accessibility

However, most of these approaches operate independently from image classification pipelines and lack real-time adaptability.

B. Contribution of Our Approach

To fill this gap, we developed a comprehensive solution that includes a *custom-trained CNN model* and integrates a Large Language Model (LLM) via the OpenAI API with LangChain. This model automates the entire pipeline from disease detection to treatment suggestions:

- During the classification stage of a leaf image, the predicted disease label is fed to the LLM.

- The LLM produces contextualized treatment advice for the specific disease in question and the crop.

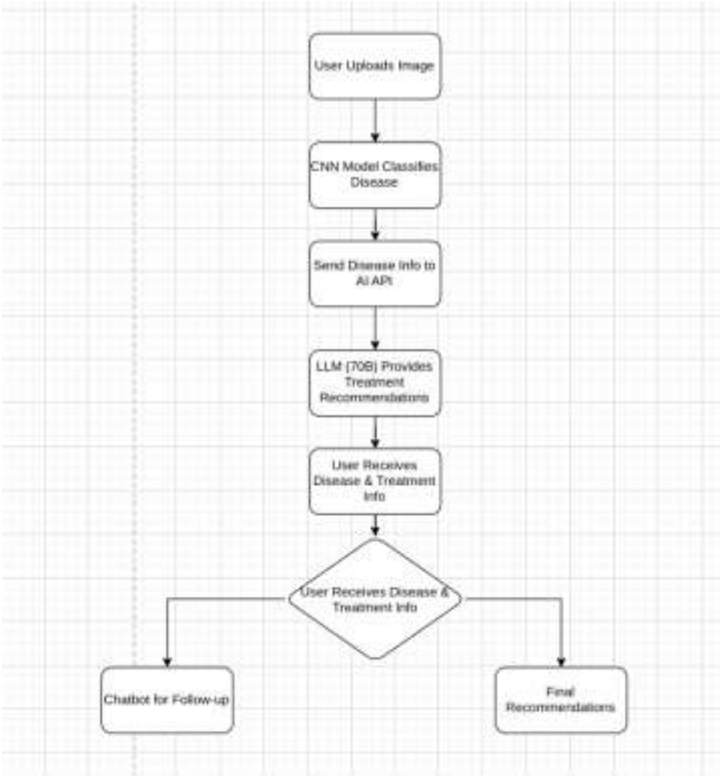
- A Streamlit-based bot allows users to enter conversational queries, and they can engage in dialogue with the system and ask follow-up questions related to the alternative controls, prevention, or remedial measures for the disease. This system facilitates dynamic dissemination of knowledge in real-time that empowers non-specialist users to receive tailored assistance for managing diseases.

6. Comparative Analysis

Study	Architecture	Dataset	Accuracy	Key Contributions
Hassan &Maji (2022)	Custm CNN (Inception + Depthwise)	PlantVillage, Rice,Cassava	99.39 / 99.66 / 76.59	Lightwei ght design for realtime use on mobile platform s
Guo (2023)	Custom CNN	Diverse crops (real-world)	92.23	Robust against realworld conditions via preprocessing
Mohanty et al. (2016)	AlexNet, GoogLeNet	PlantVillage	99.35 (lab), 31 (real)	Identifie d domain gap between lab and field conditio ns
Singh et al. (2021)	Bayesianenhanc ed CNN	PlantVillage	98.9	Feature depende ncy modelin g with Bayesian learning

7.System Design

The system consists of three primary modules:
1) Image Upload and Disease Detection Module: Users upload an image, which is classified by a CNN trained on the PlantVillage dataset.
2) AI-Powered Suggestion Module: The detected disease is sent to an AI API for tailored treatment recommendations.
3) Chatbot Module: Users can interact with a chatbot for additional disease management insights.



8.Training Configuration and Evaluation

Involving hyperparameters tuning, optimizers selection, and evaluation metrices, training plant disease classification models is very delicate work. Most researchers in the field stick to basic parameters such as learning rate between 0.001 to 0.0001, batch sizes from 32 to 128, and number of training epochs ranging from 20 to 100 considering dataset complexity and augmentation strategies used [1].

Inaccuracy, precision, recall, F1-scores, and confusion matrices remain the most popular and adopted metrics to evaluate model performance. In particular, relying on accuracy as the only metric can be problematic when dealing with imbalanced datasets; therefore, precision-recall curves alongside F1-scores shed more light [2].

Also reported is the improvement on generalization by reduction of overfitting using Cross-validation methodologies, in particular k-fold validation [3]. Using benchmark datasets like PlantVillage helps in providing a uniform point of reference for all the studies by comparison.

9.Real-Time Testing and Implementation Environment

Numerous academes emphasize the difficulty of adapting laboratory-based models to real life situations. In-field applications are greatly impacted by background noise, changing light conditions, leaf position, motion blur, and occlusion by nearby vegetation [4]. Domain shift—the difference between curated training datasets and actual data—presents an additional challenge in achieving model generalization.

As a solution, video capture at a smartphone level is often tested in open fields with UAVs and handheld cameras. According to surveys, models' accuracy was observed to decrease by 20–30% when using data from the field as compared to trained datasets such as PlantVillage [5].

In addition, other platforms such as cloud-based inference engines, mobile applications, and edge devices build [6]. These structures have to consider latency, energy consumption, and model size. In the context of real-time applications, these factors become crucial.

10.Future Scope

The field of plant disease detection using deep learning and AI is still evolving. Key opportunities for future research include:

- **Multimodal Learning**: Combining visual features with weather data, soil conditions, and geographical metadata to enhance prediction accuracy [7].
- **Few-Shot and Zero-Shot Learning**: Enabling models to detect new or rare diseases with minimal labeled data [8].
- **Explainable AI (XAI)**: Developing interpretable models that offer visual explanations (e.g., saliency maps) to build trust among farmers [9].
- **Federated Learning**: Supporting privacy-preserving model training across distributed farms and devices without centralizing data [10].
- **Disease Forecasting Models**: Moving from reactive to predictive analysis using time-series plant health monitoring and AI-driven alerts.
- **Generalization across crops and geographies**: A critical limitation of current systems is their specificity to particular crops or regions. Future research can aim at building **cross-crop** and **cross-geography models** that are resilient to seasonal, climatic, and morphological variations in plants.
- **Integration with IoT and weather data**: Merging ML models with **real-time sensor data** from IoT devices (like soil sensors, weather stations, or drones) can facilitate more **context-aware predictions** and support **preventive strategies** rather than reactive ones.
- **Few-shot and zero-shot learning**: These learning paradigms allow models to detect **novel diseases or variants** with very few or no examples. Their adoption could dramatically reduce the time and cost associated with dataset collection and labeling.
- **Explainable AI (XAI)**: The deployment of AI in agriculture, particularly for smallholder farmers, requires that predictions be **interpretable**. Future models should provide **visual and textual justifications** to increase user trust, regulatory acceptance, and actionable outcomes.

11.Conclusion

Numerous academes emphasize the difficulty of adapting laboratory-based models to real life situations. In-field applications are greatly impacted by background noise, changing light conditions, leaf position, motion blur, and occlusion by nearby vegetation [4]. Domain shift—the difference between curated training datasets and actual data—presents an additional challenge in achieving model generalization.

As a solution, video capture at a smartphone level is often tested in open fields with UAVs and handheld cameras. According to surveys, models' accuracy was observed to decrease by 20–30% when using data from the field as compared to trained datasets such as PlantVillage [5].

In addition, other platforms such as cloud-based inference engines, mobile applications, and edge devices build [6]. These structures have to consider latency, energy consumption, and model size. In the context of real-time applications, these factors become crucial.

12.Refrences

[1] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture, 145*, 311–318.

[2] Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence, 31*(4), 299–315.

[3] Too, E. C., et al. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture, 161*, 272–279.

[4] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419.

[5] Picon, A., et al. (2019). Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Computers and Electronics in Agriculture, 161*, 280–290.

[6] Sladojevic, S., et al. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience, 2016*.

[7] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90.
[8] Wang, Y., et al. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys, 53*(3), 1–34.

[9] Chouhan, S. S., et al. (2020). A survey on applications of deep learning in agriculture. *Computers and Electronics in Agriculture, 173*, 105393.

[10] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST), 10*(2), 1–19.

[11] Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture, 153*, 46–53.

[12] Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors, 17*(9), 2022.

[13] Zhang, S., et al. (2020). Attention-based neural networks for plant disease detection and classification. *IEEE Access, 8*, 36439–36451.