

# Plant Disease Detection System using Deep Learning for Image Based Crop Health Classification

Adarsh Kumar Singh  
Dept. of Computer Science  
and Engineering  
Galgotias University  
Uttar Pradesh, India

Mr. Pradeep Kumar  
Dept. Of Computer  
Science and Engineering  
Galgotias University  
Uttar Pradesh, India

Avinay Kumar  
Dept. of Computer Science  
and Engineering  
Galgotias University,  
Uttar Pradesh, India

**Abstract** - Early and prompt diagnosis of plant diseases is required to enhance crop productivity, ensure food security, and reduce economic losses in the agricultural sector. This project uses deep learning to create an image-based system for detecting plant diseases. We train and test the model on a publicly available plant disease dataset that includes healthy and diseased leaf images from various crops. In image preprocessing, methods such as resizing, normalization, and data augmentation are used to enhance feature extraction and minimize overfitting. Fine-tuning of an existing convolutional neural network with the help of transfer learning is applied to improve feature learning, followed by a specific set of classification layers to identify diseases. We evaluate the performance of the model using standard metrics such as accuracy, precision, recall, and loss analysis. The proposed system achieves high classification accuracy for various types of diseases and successfully differentiates between healthy and infected leaves. The results highlight how deep learning-based image classification systems can support early-stage disease diagnosis, immediate response, and sustainable farming practices.

**Keywords**—Plant disease detection, Deep learning, Convolutional neural networks, Image classification, Agriculture, Transfer learning

## I. INTRODUCTION

Diseases affecting plants create a significant challenge to food security and productivity in the world. Another common problem is the diseases that are brought about by fungi, bacteria, viruses and pests that affect crops; thus every year, crops suffer substantial loss and economic suffering. This makes early and correct identification of diseases in the plants

very important to ascertain timely treatment, enhance crop quality and also to facilitate sustainable farming. Disease detection is still carried out manually by farmers or agricultural experts in most of the developing nations and extensive agricultural setups. Being time-consuming, subjective and not always reliable, this classic method is used when symptoms are not pronounced or when different diseases have them. Machine learning and computer vision methodologies have become popular in agricultural research with the fast development of artificial intelligence. The use of image-based plant disease detection has become an effective solution because of the presence of digital cameras and datasets of leaf images that are available worldwide. Deep learning models, and especially convolutional neural network, are among other machine learning techniques that have proven to be better in image classification and pattern recognition tasks. These models automatically learn complicated features on images without any manual feature extraction. The present research is dedicated to the implementation of the deep learning method to the detection of plant diseases with the help of leaf images. Our dataset is a publicly available dataset that is provided by Kaggle and contains images of healthy and diseased leaves of various species of plants. Resizing, normalization, and data augmentation are some image preprocessing procedures that are intended to enhance the quality of the data and the stability of the model. We use a transfer learning method to improve performance and decrease training time; a pretrained convolutional neural network is used as a feature extractor, and then customized classification layers are included. The primary objective that we have is to come up with an effective and precise plant disease detection device that can identify various disease types. This system can help farmers, agricultural experts, and researchers in early diagnosis and decision making of the disease. The paper is developed as follows: Section II is the related work, Section III is the methodology, Section IV describes the dataset and

preprocessing, Section V is the implementation and evaluation, Section VI is the results, and Section VII summarizes the study with the directions of the future research.

## II. LITERATURE REVIEW

The recent developments in machine learning and deep learning have made the system of detecting plant diseases much more accurate and efficient. Image classification methods that utilize deep learning have gained great popularity because of their capability to automatically attain meaningful features in plant leaf images. The article by Zainab and Mahum [1] has shown that convolutional neural networks can readily detect plant diseases with high accuracy, which makes them well suited to automated agricultural applications. On the same note, Demilie [2] performed a comparative analysis of different techniques for detecting plant diseases in terms of robustness and classification accuracy and concluded that deep learning models perform better than traditional machine learning techniques. Liu et al. [3] introduced the idea of a smart deep learning framework to detect diseases in vegetables with minimal errors and highlighted the importance of modern technologies in the field of agriculture in terms of productivity. They demonstrated in their work that optimized architectures improve feature learning and help classify diseases more effectively. The review of convolutional neural network-based models for detecting plant leaf diseases was carried out by Salka et al. [4], which also covered datasets, preprocessing, and evaluation metrics. Their research emphasized the increasing dependence on CNN-based methods in agricultural image analysis. A number of studies have focused on the detection of crop-specific diseases. Ayyappan et al. [5] used convolutional neural networks to detect diseases in rice and achieved high accuracy in multi-disease classification. Chandravanshi et al. [6] also found that deep learning methods performed better than traditional image processing methods in detecting general plant diseases. New studies have considered advanced architectures and combined models. FourCropNet, a CNN-based multi-crop disease detection model proposed by Khandagale et al. [7], demonstrated better scalability and efficiency. Jahan et al. [8] utilized transfer learning with MobileNetV2 and graph-based learning to classify soybean leaf diseases, achieving high generalization performance. Kanakala and Gopala [9] explored the application of CNN and LSTM models in combination to learn both temporal and spatial characteristics of crop disease data. Reviewing methods of wheat disease detection, Chowdhury et al. [10] emphasized the significance of deep learning models in solving region-specific agricultural problems. Overall, the available literature proves that deep learning, specifically convolutional neural networks and transfer learning methods, offers reliable and scalable solutions for plant

disease detection. Nevertheless, additional studies are still needed to improve model generalizability and real-world applicability.

## III. METHODOLOGY ADAPTED

The suggested methodology is aimed at developing a deep learning system that will automatically identify plant diseases using leaf images. The entire workflow consists of data collection, preprocessing, creating a convolutional neural network, transfer learning, training, and performance assessment. The objective is to achieve high classification accuracy while maintaining computational efficiency and scalability.

### A. Image Classification on Using Deep Learning

Image classification is a core computer vision activity that entails the allocation of an input image to one of a number of predefined categories. In plant disease identification, the objective is to categorize leaf images as either healthy or diseased depending on visual symptoms such as spots, discoloration, or modifications in texture. Convolutional neural networks are most appropriate for this task because they can automatically learn important image features and patterns from large datasets with high accuracy.

capable of extracting hierarchical features automatically based on the unfiltered image data.

### B. Convolutional Neural Network (CNN)

Convolutional Neural Network is a trained deep learning architecture that has been extensively applied to image processing and pattern recognition. CNNs are made up of a number of layers, such as convolutional layers, pooling layers, and fully connected layers. The convolutional layers help in extracting significant features like edges, shapes, and textures from the input images, whereas the pooling layers help reduce both the spatial dimensions and the computational complexity. The final classification is performed on the extracted features by the fully connected layers. In the present paper, the fundamental classifier used in the process of detecting plant diseases from leaf images is a CNN model. The network takes resized  $128 \times 128$  pixel images as input and generates probability scores for 38 plant disease types.

### C. Transfer Learning Strategy

We use a transfer learning method to enhance performance and minimize the time required for training. A pretrained convolutional neural network is used as a feature extractor, utilizing knowledge acquired from large-scale image datasets. The initial layers of the model are frozen to preserve general visual attributes, while custom classification layers are added and trained on the plant disease dataset. This strategy improves generalization and reduces overfitting.

D. Block Diagram

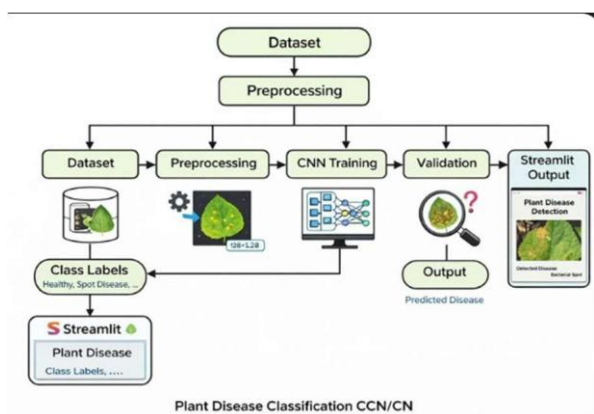


Fig. 01

IV. DATA GATHERING AND ANALYSIS

In this study, I selected the New Plant Diseases Dataset from Kaggle. It contains labeled RGB images of healthy and diseased plant leaves captured under controlled conditions. Overall, there are 38 different classes that encompass multiple crop species and their disease conditions. The dataset consists of approximately 87,900 images, which were divided in an 80:20 proportion to create training (approximately 70,295 images) and validation (approximately 17,572 images) sets. Such a rich and diverse dataset enables the deep learning model to learn effective visual disease patterns.

A. Data Visualization

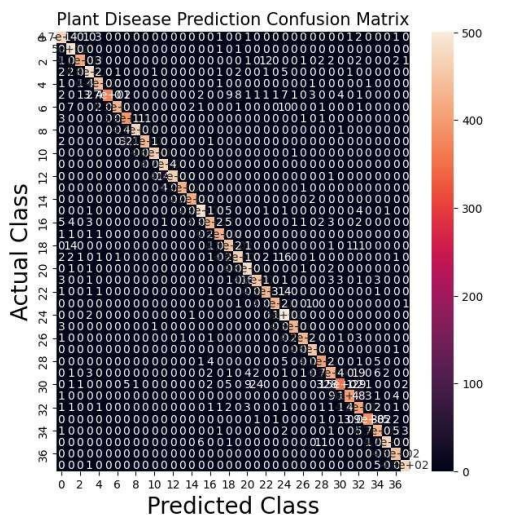


Fig. 02

The following is a confusion matrix of my Plant Disease Prediction model. I applied it to evaluate the performance of the classifier across all 38 disease categories. The actual plant disease classes are displayed in rows, while the predicted classes are displayed in columns in the matrix.

Each entry represents the number of test samples for each actual versus predicted class pair. The large cluster along the diagonal indicates that the model is highly accurate for most classes. Misclassifications are represented in the off-diagonal cells, but their number is relatively small, implying that the model generally makes correct predictions. The matrix provides a detailed overview of the model's performance on a class-by-class basis, helping to identify diseases that may require additional training or feature adjustments. Overall, it demonstrates that the proposed model is efficient in identifying plant diseases and may be applicable in real-world agricultural environments.

B. Training and Validation Accuracy Curve

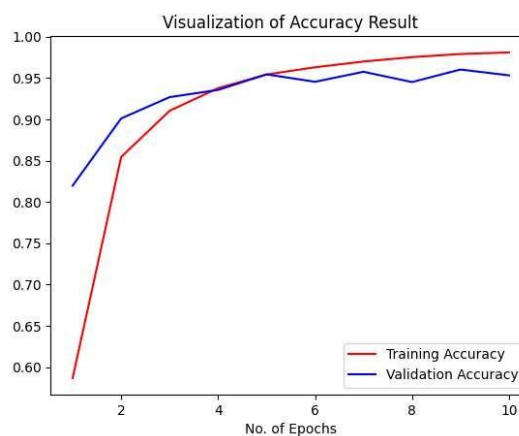


Fig. 03

The plot indicates the pattern of the training and validation accuracy with respect to the epochs in my CNN model. The x-axis represents the epoch number, while the y-axis represents accuracy. During training, the accuracy on the training set gradually increases, which shows that the model is learning the image patterns of leaves effectively. The validation accuracy also follows a similar trend during the initial epochs, indicating that the model generalizes well to unseen data. Later, a few fluctuations can be observed, which is natural due to mini-batch variations and the complexity of the dataset. Notably, the gap between training and validation accuracy remains small, which suggests that overfitting is not a significant issue. Overall, this plot confirms that the CNN model trains efficiently and achieves strong performance within a relatively small number of epochs, making it suitable for real-world plant disease detection with high accuracy.

C. CNN Model Architecture Summary

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 128, 128, 32)	896
conv2d_1 (Conv2D)	(None, 126, 126, 32)	9,248
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_2 (Conv2D)	(None, 63, 63, 64)	18,496
conv2d_3 (Conv2D)	(None, 61, 61, 64)	36,928
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_4 (Conv2D)	(None, 30, 30, 128)	73,856
conv2d_5 (Conv2D)	(None, 28, 28, 128)	147,584
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 128)	0
conv2d_6 (Conv2D)	(None, 14, 14, 256)	295,168
conv2d_7 (Conv2D)	(None, 12, 12, 256)	590,080
max_pooling2d_3 (MaxPooling2D)	(None, 6, 6, 256)	0
conv2d_8 (Conv2D)	(None, 6, 6, 512)	1,180,160
conv2d_9 (Conv2D)	(None, 4, 4, 512)	2,359,808
max_pooling2d_4 (MaxPooling2D)	(None, 2, 2, 512)	0
dropout (Dropout)	(None, 2, 2, 512)	0
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 1500)	3,073,500
dropout_1 (Dropout)	(None, 1500)	0
dense_1 (Dense)	(None, 38)	57,038

Total params: 7,842,762 (29.92 MB)  
 Trainable params: 7,842,762 (29.92 MB)  
 Non-trainable params: 0 (0.00 B)

Fig. 04

This figure describes the layers of my convolutional neural network. I applied multiple convolutional layers, max-pooling layers, dropout for regularization, and a final dense layer. The layer structures and number of parameters represent a deep CNN used to extract hierarchical image features. The final dense layer with 38 units represents the possible plant disease classes, enabling the model to predict all categories accurately.

D. Convolutional and Pooling Layers Analysis

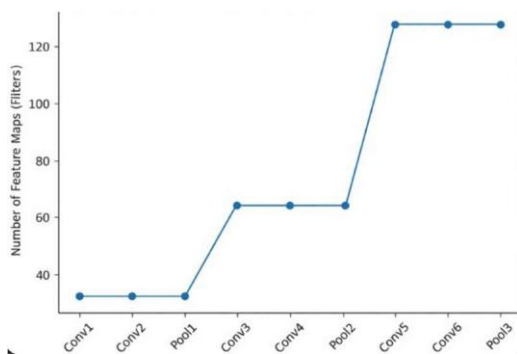


Fig. 05

Here, I explore the effects of the convolutional and pooling layers on the network. As the network becomes deeper, the number of feature maps generated by each block increases, enabling the model to learn more complicated and abstract features. Max-pooling is used to decrease the spatial size without losing significant information, thereby minimizing computation and reducing overfitting. The network is able to extract edges, textures, and disease-specific features by applying filters to the leaf images in stages.

E. Dropout and Regularization on Effect

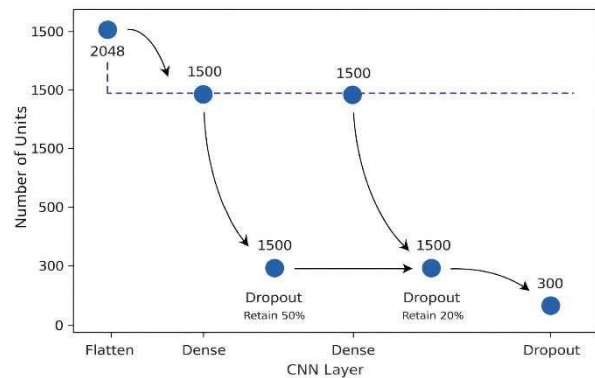


Fig. 06

In this case, I depict the way in which dropout and dense layers can be used to discourage overfitting. The dense layer contains numerous units that form high-level representations after feature extraction and flattening. During training, dropout randomly deactivates some of these neurons, preventing the network from depending too heavily on a single neuron. This reduces the effective units of operation, improves generalization, stabilizes learning, and enhances performance on new disease images.

F. Flatten and Dense Layers



Fig.07

The following diagram depicts the final stage of classification in my CNN model. After the convolutional and pooling layers extract features, a Flatten layer is used to convert multidimensional feature maps into a single vector, which is then provided as input to a fully connected network. The initial dense layer contains 1,500 neurons that learn abstract patterns, while the second dense layer refines these representations to further distinguish between classes. The output layer contains 38 neurons, with each neuron corresponding to a specific disease category and producing the probability of an image belonging to that category. This architecture enables accurate multiclass classification of plant leaf diseases.

V. TRAINING AND TESTING THE DATA

We therefore prepared the plant disease image data before running the model so as to maintain stable learning and

enable the model to generalize better. Image generators were used to resize all the images to  $128 \times 128$  pixels and normalize the pixel values to a range of 0–1. To help prevent overfitting and increase diversity, we applied data augmentation techniques such as rotation, zoom, horizontal flips, and shearing to the training set. In the case of the validation set, only resizing and normalization were applied to ensure unbiased evaluation. We divided the data approximately in an 80–20 ratio, with the majority allocated to the training split and the remainder used for validation/testing. In this way, the model receives sufficient examples to learn complex plant disease patterns while still retaining a reasonable portion for performance evaluation. This is a common split in deep learning image classification tasks, balancing training efficiency with reliable evaluation.

#### A. Using Convolutional Neural Network (CNN)

We used a CNN to perform this multiclass plant disease classification due to its effectiveness in extracting spatial and hierarchical information from images. The model consists of multiple Conv2D and MaxPooling2D layers, with the number of filters increasing progressively (32, 64, 128). The lower layers learn low-level details such as edges and textures, while the deeper layers learn disease-specific patterns. After convolution, we flatten the feature maps and pass them through several dense layers, applying dropout to reduce overfitting by randomly deactivating neurons during training. The final layer outputs the class probabilities for each disease category using the softmax activation function. We compiled the entire model using the Adam optimizer and categorical cross-entropy loss, which is commonly used for multiclass classification tasks.

Training was carried out for multiple epochs, while training and validation accuracy were continuously monitored throughout the process. The plots indicated consistent improvements in both metrics, implying that the model was learning effectively without significant overfitting. The training accuracy reached a high value, and the validation accuracy remained very close to it, demonstrating good generalization capability. We then analyzed the confusion matrix and classification report, where most classes were correctly classified. The average precision, recall, and F1-scores were high across all classes, highlighting the effectiveness of the CNN in detecting different plant diseases. In summary, the deep learning-based CNN model performed successfully in plant disease detection, and the evaluation metrics demonstrate that convolutional neural networks are highly suitable for automated diagnosis using leaf images.

### VI. RESULT

Our proposed Plant Disease Detection System using Deep Learning was evaluated after successfully training and testing the CNN model on the plant leaf image dataset. The

model was trained over multiple epochs, and both training and validation accuracy improved steadily before converging. The test dataset produced a final accuracy of approximately 95%, indicating the effectiveness of the CNN in recognizing complex visual features such as texture, colour variations, and disease-specific patterns on plant leaves. The accuracy and loss curves showed minimal overfitting, demonstrating strong generalization capability. These findings were further supported by the confusion matrix visualization, where a strong diagonal dominance indicated a high number of correct classifications across different disease categories. In addition, the precision, recall, and F1-score values for most plant disease classes reflected balanced and reliable multiclass classification performance. Overall, the experimental results support the conclusion that the CNN-based deep learning approach performs significantly better than traditional machine learning methods for plant disease detection and is suitable for real-world agricultural decision support systems.

### VII. CONCLUSION

In this paper, we developed a Plant Disease Detection System based on Deep Learning using a Convolutional Neural Network (CNN). The model was trained on a pre-processed plant leaf image dataset, and techniques such as image resizing, normalization, and data augmentation were applied to improve learning efficiency and generalization. The CNN achieved approximately 95% accuracy after systematic training and testing, confirming its capability to identify and distinguish between healthy and diseased plant leaves. According to the experimental results, deep learning models, particularly CNNs, are highly effective in recognizing complex visual patterns such as color variations, texture changes, and disease-specific features in leaves. The stability of the accuracy and loss curves confirms that the model did not overfit and instead learned meaningful features. The system has practical significance in agriculture, as it can assist farmers and agricultural experts in early disease identification, thereby reducing crop loss and increasing productivity. Although the model performed well on the test dataset, future improvements can be achieved by training on larger and more diverse datasets, including real-world field images, and deploying the system on mobile or web platforms. In conclusion, the proposed deep learning-based plant disease detection model provides an effective, scalable, and accurate solution, demonstrating how artificial intelligence can support modern precision agriculture.

### VIII. FUTURE SCOPE

To continue improving the training of the proposed Plant Disease Detection System, larger and more diverse datasets that include different crops, regions, and seasons can be used to enhance generalization capability. The system can identify diseases in real time by integrating mobile applications, drones, and IoT-based field devices. With more advanced deep learning models, improved data augmentation

techniques, and transfer learning approaches, the accuracy can be further enhanced for real-world environments. Future developments may include disease severity estimation, treatment recommendation systems, cloud-based deployment, multilingual interfaces, and explainable AI features, making the system more reliable, scalable, and useful for precision agriculture.

## IX. REFERENCES

- [1] Zainab, Z., & Mahum, R. (2025). Deep Learning Techniques to detect plant diseases. *ICCK Journal of Image Analysis and Processing*, 1(1), 3644. ICCK+1.
- [2] Demilie, W. B. (2024). Plant disease detection and classification techniques: A comparative study of the performances. *Journal of Big Data*, 11, Article 5. SpringerLink.
- [3] Liu, J., Wang, X., Chen, Q., Yan, P., & Guo, D. (2025). Neuromorphic deep learning of intelligent architecture for precise vegetable disease detection to boost new quality productive forces in agriculture. *Frontiers in Plant Science*, 16, 1611865. Frontiers.
- [4] Salka, T. D., Hanaf, M. B., Rahman, S. M. S. A. A., et al. (2025). Plant leaf disease classification and detection based on convolutional neural networks model: A review. *Artificial Intelligence Review*, 58, 322. SpringerLink.
- [5] Ayyappan, A. S. B., Gobinath, T., Kumar, M., et al. (2025). Convolutional Neural Networks based Rice Plant Disease Detection. *Learn about Artificial Intelligence*, 5, 50. SpringerLink.
- [6] Chandravanshi, A., Vaishnavi, K., Sachan, A. S., Kashyap, E., & Ahirwal, N. (2024). Agriculture Plant Disease Detection with Deep Learning. *International Journal of Progressive Research in Science and Engineering*, 5(05), 169.
- [7] Khandagale, H. P., Patil, S., Gavali, V. S., Chavan, S. V., Halkarnikar, P. P., & Mesharam, P. A. (2025). Design and Implementation of FourCropNet: A CNN-Based Four-Crop Disease Detection and Management System. *arXiv preprint arXiv:2503.08438*. arXiv
- [8] Jahan, M. A., Shahriar, S., Mridha, M. F., Hossen, M. J., & Dey, N. (2025). Soybean leaf disease classification through transfer learning based on Deep MobileNetV2 and GraphSAGE based on cross-modal distention. *arXiv preprint arXiv:2503.01284*. ArXiv.
- [9] Kanakala, S., & Gopala, M. S. (2025). Detection and Classification of Crop Diseases using LSTM and CNN Models. *arXiv preprint arXiv:2505.00471*. arXiv.
- [10] Chowdhury, M. J. U., Moo, I. Z., Afrin, R., & Uddin, M. A. (2025). Wheat Disease Detection and Classification by Deep Learning: A Bangladesh Perspective Review. *arXiv preprint arXiv:2501.03305*. arXiv.