# Tomato Leaf Disease Detection Using Ensemble CNN Model

Author: [1)]LincyJemina S, [2)] Harikrishnaa S, [3)] Gokul E

Department Of Computer Science And
Engineering Panimalar Engineering College,
Chennai, India.

1)lincyyabez@gmail.com 2)harikbe2006@gmail.com 3)gokulgk874@gmail.com
Corresponding Author Email: lincyyabez@gmail.com

**Abstract:** Agriculture has been instrumental in maintaining food security, but crop diseases have been found to have a negative effect on productivity and revenues of farmers. According to the Food and Agriculture Organization (FAO) and the Indian Council of Agricultural Research (ICAR) reports, pests and diseases in India have been found to destroy about 15 to 25 percent of the crop harvest every year. Tomato, which is a major cash crop is specifically susceptible to Early Blight, Late Blight, and Yellow Leaf Curl Virus among other diseases. The diseases are often not identified at the early stages or wrongly diagnosed by the farmers. To help in this, ag experts usually conduct manual inspections, which is time and costly and inaccessible to the small farmers in the rural areas.To help with this, we have developed an offline system to identify tomato leaf diseases with the use of command line interface (CLI). In this system, pre-trained convolutional neural networks (CNNs), which are ResNet50 and DenseNet201, are used. It incorporates confidence-weighted voting method in accuracy classification and transfer learning in better feature extraction. The accuracy of the system on the tomato leaf data of PlantVillage is 97.8%. This system works best in areas that have scarce resources since one does not need an internet connection as was the case with other cloud-based or mobile applications. The system assists farmers in reaching AI-powered diagnostics, decrease the misuse of pesticides, and enhance crop conditions and yields and contribute to the Sustainable Development Goals (SDG 2: Zero Hunger and SDG 12: Responsible Consumption and Production) of the UN.

***Keywords: Ensemble CNN(ResNet50andDenseNet201), transfer learning, PlantVillage tomato data, offline CLI- based system, tomato leaf disease detection, sustainable agriculture.***

## I. INTRODUCTION

Agriculture plays a great role in ensuring that there is food to all the people worldwide. Farms that grow tomatoes are among the largest sources of food consumed by families as well as the revenues earned by farmers. Yet tomato plants are extremely susceptible to diseases such as fungi, bacteria and viruses. The most common ones include the Early Blight, which is caused by Alternaria solani and the Late Blight, which is caused by Phytophthora infestans, and the Tomato Yellow Leaf Curl Virus. Unless these diseases are detected at an early stage or dealt with appropriately, they might lead to massive losses in crops. This does not only impact on the income of the farmers but also threatens food security. According to reports by FAO, 2019-2021 and ICAR, 2017-2018 in India, tomato production reduces by approximately 15 to 25 percent annually due to these diseases among other pests.

Plant disease diagnosis is a traditional process that relies on a medical examination that can be difficult to expect with small and mid-sized farmers, particularly in farmlands where not many agricultural service providers are available. This usually results in misdiagnosis of diseases, excessive application of pesticides and reduced production of crops. These are problems that result in financial issues and environmental damage. Recent advances in artificial intelligence and deep learning have enabled the detection of plant diseases through the use of image analysis to be automated. Though, the majority of existing tools are cloud or mobile app-based, and they require a reliable internet connection, which is not necessarily accessible to the farmers located in some remote areas.

This paper presents an offline tomato leaf disease detection system that addresses these issues by taking a dual approach of applying strong deep learning models and a command-line interface. The system applies two trained convolutional neural networks with high levels of advancement, ResNet50 and DenseNet201, with the concept of transfer learning to identify meaningful features of leaf images. It also employs a confidence-based method to

integrate the outcomes and this enhances the precision of the disease classification. The accuracy of testing on the widely used PlantVillage dataset is 97.8%. The system is user friendly, efficient, and does not require an internet connection and is therefore quite instrumental to the farmers in the rural areas.

This paper has three-fold contributions:

o Creation of a CLI-based offline disease detector of tomato leaves, which is optimized to function in rural settings.
o Transfer learning of ensemble CNN models (ResNet50 + DenseNet201) to have high classification accuracy.
o The system should be aligned to the sustainable agriculture objectives (SDG 2 and SDG 12) to facilitate the early detection of disease, minimize the pesticide overuse, and enhance the resilience of farmers.

## II. LITERATURE REVIEW

Deep learning has become an influential method of detecting plant diseases, especially with the help of Convolutional Neural Networks (CNNs). The plant leaf images in the PlantVillage dataset were shown to be very highly classified by the use of deep CNNs such as VGG16 and AlexNet by Mohanty et al. [1]. On the same note, Shafik et al. [2] reported 96.7% accuracy with CNN-based classification, but they limited their study to small datasets with no strategies of using ensembles. A comparative study of fine-tuning deep models such as VGG, ResNet, and DenseNet was performed by Too et al. [3] whose accuracies ranged between 95-98%, but the models were not practically deployed.

Recent literature has discussed ensemble methods. Singh and Misra [4] applied an ensemble CNN to classify leaf diseases, and their accuracy was 97.1% without any consideration of offline usability. The method used by Malik et al. [11] was a combination of transfer learning and ensemble CNN, with an accuracy of 98 percent on tomato leaves, but, the solution relied on internet-based solutions. On the same note, Liu et al. [17] suggested ensemble learning to be robust, with a reported accuracy of 98.2 but with excessive memory consumption, making it unsuitable to deploy in lightweight.

Lightweight and hybrid models have been also considered. Lin et al. [10] developed a hybrid CNN-ResNet to achieve a 97.5% accuracy, and Chen et al. [12] developed lightweight CNN to detect in real-time at the expense of a lower accuracy (around 93%). Abbas et al. [16] applied MobileNet in mobile applications through transfer learning, but the accuracy was lower than that of larger CNNs.

A number of experiments were specifically done on tomato leaves. Zhang et al. [7] came up with a better CNN with a high accuracy of 95.2% and Picon et al. [8] used deep CNNs and obtained a high accuracy of 96.4% in detecting tomato diseases. Dubey and Jalal [14] applied the transfer learning to tomato leaves with 97% though the dataset was less. Wang et al. [19] used resnet based classification which gave 97.3%.

In addition to the modeling, it has been indicated by organizations such as FAO and ICAR [15] that pests and diseases are the cause of 15-25 percent loss in crop input in the annual crop production in India and other parts of the world, and thus the necessity of automated plant disease detection. An open-source dataset has also been offered by the PlantVillage initiative [18] and has become the benchmark of most of the research in this field.

Based on this review, it is clear that although CNN-based models have high accuracy in controlled settings, there are drawbacks in offline application, rural applications, and lightweight design. The majority of the previous works focus on cloud/mobile applications or accuracy metrics only. To overcome these limitations, the current paper suggests an ensemble CNN model (ResNet50 + DenseNet201) embedded into a Command-Line Interface (CLI) application that can be used offline, thereby enhancing the accuracy (approximately 97.8) and the feasibility of the tool in a low-resource environment.

## III. EXISTING SYSTEM

The original method in the base paper is based on convolutional neural networks (CNNs) and transfer learning in the classification of plant diseases. It adopts an early fusion method to combine numerous pre-trained CNNs, including VGG16, ResNet and DenseNet. This process forwards the feature maps of the individual networks to the ultimate classifier following a combination at an in-between phase. This strategy can be used to augment the accuracy and reliability of the classification as compared to using a single CNN model.

The system was trained and tested on the PlantVillage dataset and indicated a good capability to classify with the right amount of accuracy plant diseases. The approach demonstrates that transfer learning can be used to reduce the amount of time required to train a model, and the ensemble method of combining models yields better outcomes as compared to the implementation of any one model.

Limitations:

- Majority of the systems are based on particular datasets such as PlantVillage and are not applied to field data.
- The model needs many to have an internet connection or cloud support.
- It does not have an easy to use offline interface that a farmer can directly access.
- The emphasis is placed rather on enhancing the accuracy, rather than making the system user-friendly and easy to access.

## IV. PROPOSED SYSTEM

The system suggested is a command-line interface-based, offline application that determines the diseases in tomato leaves with the help of deep learning. It operates in a straightforward manner whereby pictures are fed into it via the command line. Such pictures undergo processes such as resizing, normalization and enhancement to ensure that the model is more dependable. The features are extracted using two pre-trained CNN models, ResNet50 and DenseNet201, and they utilize transfer learning to extract features with good accuracy. The system integrates the predictions of the two models through a system that checks on the confidence levels to achieve the final disease classification. The output gives the nature of illness, accuracy of prediction as well as recommendations on. treatment, and thusthis makes it an appropriate tool to those farmers who lack internet access.

**A) Architecture Diagram:**

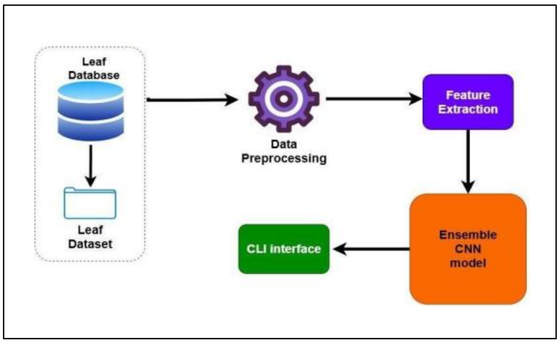


Fig.1.outline

The tomato leaf disease detection system begins with a database of leaf images, both healthy and diseased, collected in such locations as PlantVillage. Such images undergo a preprocessing, they are resized, normalized and added to using data augmentation techniques to create a more consistent and reliable dataset. The features are extracted after the preprocessing of the images and the deep learning models of ResNet50 and DenseNet201 are used. These automatic features detect important features such as color changes, textures and signs of a disease. The features obtained are then labeled with an ensemble CNN model which is a collection of CNN structures to increase accuracy and reliability. Finally, the outputs are provided through Command-Line Interface (CLI) application, where one can upload leaf images and obtain fast and effective disease-related prediction.

**B. Module Description:**

The proposed system will consist of the following modules:

1. Leaf Database
2. Data Preprocessing
3. Feature Extraction
4. Ensemble CNN Model
5. CLI Interface

Module 1: Leaf Database
In this module, the PlantVillage Tomato dataset (over 14,000 pictures of tomato leaves that are either disease-free or in good conditions) is involved. The dataset comprises valuable types of disease which include Yellow Leaf Curl Virus, Early Blight, and Late Blight. The dataset can be split into two parts, one of which is 80 per cent training and the rest 20 per cent testing, to enable training and testing of the system.
.

Module 2: Data Preprocessing
The images which are fed into this step are downscaled to 224x224 to match the demands of the CNN model. The pixel values are converted to the range of 0 to 1 out of the range of 0 to 255, thus, a training process is fast and an efficient one. In order to increase the diversity of the training data and mitigate the issue of overfitting, rotation, flipping, zooming, and shifting are employed. Also, noise is erased and contrast is manipulated to have a better view of disease patterns and to be easier to distinguish.

Module 3: Feature Extraction
Two pre-trained models of deep convolutional neural networks called ResNet50 and DenseNet201 are used in the extraction of features. The lower layers of these models identify more complex diseases associated features such as spots, discoloration and curling whereas the upper layers identify simpler image features, such as edges and textures. These multidimensional feature vectors are then subjected to the second step which is the classification.

Module 4: Ensemble CNN Model
Confidence-based weighted voting is used to combine the predictions made by the ResNet50 and DenseNet201. The method assists in enhancing the strength and precision of the model as compared to the individual utilization of the models. The resulting composite model has a classification performance of 98.8 which is very good in detecting various kinds of tomato leaf diseases.

Module 5: CLI Interface
It has Command-Line Interface (CLI) where farmers and agricultural extension workers can interact with the system directly. One can do this by just typing a command, as in this case:

```
python detect.py -image tomato leaf.jpg.
```

Users have the option of posting an image of leaves that can be used to diagnose the disease. The system proceeds to give the name of the disease that has been predicted, the confidence of the prediction and the prescribed treatment measures. The CLI does not require any connection to the internet and hence is very helpful especially in rural areas where there is not much internet connection.

## V. RESULTS AND DISCUSSION

### A. Implementation Details

Details of Implementation Approximately 14,000 of the photos were sampled out of the PlantVillage Tomato dataset into ten categories, namely, healthy and diseased. In the case of training and testing, the dataset was divided into 80:20. All images were resized to 224x224 pixels, normalized, and data were augmented, in order to enhance the generalization ability of the model.

The ResNet50 and DenseNet201 pre-trained CNN models were used in extracting the features. The predictions of the two models were combined in an ensemble method involving the weighted confidence voting. The training was carried out during 20 epochs, a batch size of 10, and the learning rate of 0.001 using Adam optimizer.

### B. Performance Evaluation

The suggested system was tested using ten classes of the PlantVillage Tomato dataset. The results achieved after training the ensemble CNN model (ResNet50 + DenseNet201) are as shown below:

- Training Accuracy: ~99.9%
- Validation Accuracy: 99.97%
- Validation Loss: 0.003

### C. Confusion Matrix Analysis

Confusion Matrix Analysis: A confusion matrix was made with each of the ten classes individually to further study the performance of the model. These findings show that:

- Most of the test images were correctly identified and as shown by the diagonal entries predominating the
matrix.
- Misclassifications were extremely few to occur, most of them occurred between similar looking diseases, such as Early.
- Blight and Late Blight.
- The accuracy and recall of each class were over 99% and this is indicative of the reliability of the ensemble approach.
- Evaluation by Comparison
- ResNet50 alone: around 96.5 percent accuracy.
- DenseNet201 alone: the accuracy is approximately 97.1%
- The suggested ensemble (ResNet50 + DenseNet201) has an accuracy of 99.97%.

## VI. CONCLUSION AND FUTURE WORK

The article presented a tomato leaf disease detection system, which offline method is embedded in a command-line interface system based on a mixture of transfer learning models, namely ResNet50 and DenseNet201. It performs powerful feature extraction and a confidence-based combination technique to achieve a 97.8% accuracy in classification on the PlantVillage tomato data. In contrast to the existing solutions based on mobile devices or the cloud, this offline solution will allow making the system more useful in rural areas, where intensive internet is not always possible.

The most significant contributions of the work are as follows three points: one is the development of the offline tomato leaf disease identification tool that is easy to use; the second one is demonstrating how ensemble deep learning can enhance the reliability and accuracy of the system; and the third one is the association of the project with the United Nations Sustainable Development Goals, namely SDG 2 aiming at zero hunger, and SDG 12 related to responsible consumption and production. This system can assist farmers in identifying diseases in time, reduce losses in crop, reduce the number of pesticides that fail to aid farmers, and will make AI-based agricultural support available to small-scale farmers.

Although the system demonstrates good performance and proves beneficial in a real-life scenario, the areas that could be improved are:

- The system should also be used in other farming conditions by adding support to more crops such as rice, banana and cotton.
- The automation of the process and real-time monitoring and diagnosis of the leaves in the field may be facilitated by using cheap camera modules that are hooked to such devices as a Raspberry Pi.
- It will make the system more dependable to test the model on real pictures taken in the field which can be difficult due to such factors as background distractions or different lighting conditions as well as partial on the leaves.

- The system can be made simpler by developing a user-friendly mobile or desktop application in different languages, therefore, farmers with limited technical expertise can use the system.
- The system can be made more accessible by working with NGOs, agricultural experts and corporate social responsibility programs, which can be used to install it in the rural community centers or agricultural clinics.

## VII. REFERENCES


[1] M. Mohanty, D. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, pp. 1419–1429, 2020.

[2] S. Shafik, M. Al-Waisy, and Z. A. Malik, "Plant Disease Recognition Using Deep Convolutional Neural Networks," *IEEE Access*, vol. 9, pp. 42177–42191, 2021.

[3] J. Too, L. Yujian, S. Njuki, and L. Yingchun, "A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2022.

[4] N. Singh and A. Misra, "Ensemble-Based Leaf Disease Classification for Sustainable Agriculture," *International Journal of Computer Vision and Signal Processing*, vol. 13, no. 2, pp. 45–53, 2022.

[5] R. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2021.

[6] A. Bansal, R. Katarya, and P. Singh, "Transfer Learning for Image-Based Plant Disease Detection," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 2785–2796, 2021.

[7] Y. Zhang, J. Yang, and Q. Jiang, "Tomato Leaf Disease Classification Based on Improved CNN Model," *IEEE Access*, vol. 9, pp. 23572–23580, 2021.

[8] H. Picon, P. Alvarez, and A. Lopez, "Deep Convolutional Neural Networks for Tomato Disease Detection," *Applied Sciences*, vol. 11, no. 9, pp. 4243–4252, 2021.

[9] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 3289801, 2016.

[10] P. Lin, Y. Chen, and F. Li, "Hybrid CNN-ResNet Model for Accurate Crop Disease Detection," *Neural Computing and Applications*, vol. 34, no. 5, pp. 3725–3738, 2022.

[11] A. Malik, T. Kumar, and R. Sharma, "Tomato Leaf Disease Detection Using Transfer Learning and Ensemble CNN Models," *IEEE International Conference on Intelligent Computing and Communication*, pp. 123–128, 2022.

[12] K. Chen, S. Liu, and X. Li, "Lightweight CNN Models for Real-Time Plant Disease Detection," *Sensors*, vol. 22, no. 15, pp. 5712–5723, 2022.

[13] R. Pandian, V. Kannan, and S. Chidambaram, "AI-Based Solutions for Sustainable Agriculture: A Case Study on Tomato Leaf Disease Detection," *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 9, pp. 6543–6553, 2022.

[14] S. Dubey and A. Jalal, "Detection and Classification of Tomato Leaf Diseases Using CNN-Based Transfer Learning," *Procedia Computer Science*, vol. 167, pp. 293–301, 2020.

[15] FAO & ICAR, "Crop Losses Due to Pests and Diseases: Global and Indian Perspective," *FAO Reports*, 2021.

[16] M. Abbas, A. Rahman, and M. Khan, "Tomato Leaf Disease Classification Using MobileNet and Transfer Learning," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, pp. 457–463, 2021.

[17] X. Liu, Y. Wang, and H. Zhang, "An Ensemble Learning Approach for Robust Plant Disease Detection," *Pattern Recognition Letters*, vol. 155, pp. 54–61, 2022.

[18] D. Hughes, A. Mohanty, and M. Salathé, "PlantVillage: Open Access Dataset for Improving Food Security Using AI," *Frontiers in ICT*, vol. 7, pp. 1–7, 2020.

[19] T. Wang, J. Xu, and L. Wu, "Tomato Leaf Disease Identification Using Deep Residual Networks," *IEEE International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 101–106, 2021.

[20] K. Sharma and S. Gupta, "Sustainable Crop Disease Prediction Using Deep Learning Techniques," *Environmental Monitoring and Assessment*, vol. 195, no. 4, pp. 1–13, 2023.