*Grenze International Journal of Engineering and Technology, June Issue*

# Automated Identification of Tomato Plant Disease using CNN based Deep Learning Model

Disha Wankhede[1], Parth Kamalakar[2], Vaidehi Unhale[3], Sejal Revanwar[4] and Yash Khandagale[5]
[1-5]VIT, Computer Engineering, Pune, India
Email: disha.wankhede@vit.edu. parth.1252070002@vit.edu, vaidehi.1252070012@vit.edu, sejal.1252070014@vit.edu, yash.1252070019@vit.edu

***Abstract*— Tomato plants can catch a lot of different diseases very easily. Early identification of these diseases plays an important role in effective agricultural management. The traditional detection method depends mainly on manual inspection by experts, which not only takes long time but also has a high chance of errors. An automated tomato plant disease identification system is proposed based on Convolutional Neural Network (CNN) in this paper. The model recognizes between healthy and diseased plants from images of tomato leaves. Training uses a publicly available dataset and evaluating the model, the experimental results show that the CNN based approach gives reliable and efficient disease classification. This method helps in identifying tomato plant diseases at an early stage and it can support farmers for taking better agricultural decisions, improving crop health and reducing losses in farming.**

***Index Terms*— Tomato Plant Disease, Convolutional Neural Network, Deep Learning, Leaf Disease Classification.**

## I. INTRODUCTION

Agriculture plays a very important role in the global economy and tomato is one of the crops that people grow in many parts of the world. It is used in lots of everyday food dishes and also has good nutritional value. But tomato plants easily get affected by different diseases. Many of these problems appear on the leaves such as Early Blight, Late Blight and Leaf Mold. When farmers do not notice these diseases at the start, the infection spreads fast and the quality and quantity of the crop can go down.

Usually farmers or agricultural experts try to find plant diseases by simply looking at the plants. This visual checking method has been used for many years. Still it can take a lot of time and effort and it mostly depends on the experience of the person who is checking the crop. Small farmers especially may not always have access to experts or proper tools that help in identifying diseases quickly. Because of this the disease is sometimes discovered late and by that time the plants are already damaged. So there is a clear need for a simple, easy and automatic system that can help farmers notice diseases earlier.

In the last few years people have started using artificial intelligence in agriculture. Deep learning and computer vision are becoming useful for solving many practical problems. One well known method is the Convolutional Neural Network (CNN). This type of model is good at understanding images and finding patterns in them. It can also study things like the colour of the leaf, the texture and the shape. By learning these patterns the model can understand whether the leaf looks healthy or if it shows signs of disease.

*Grenze ID: 01.GIJET.12.2.513*

In this research a CNN based deep learning model is used to identify diseases and infections in tomato leaves with the help of digital images. The system takes the image of a leaf and processes it step by step. After that it tries to decide if the leaf is healthy or affected by disease. The main idea of this work is to create a system that can help farmers notice plant diseases earlier and reduce crop loss. It also shows how modern deep learning methods can be used for real agricultural problems and help farmers make good decisions in farming.

## II. LITERATURE REVIEW

Automated plant disease detection using leaf images has become a topic many researchers are exploring in recent years. Earlier studies mostly used traditional image processing methods. In those methods features like colour, texture and shape were taken from the leaf images and then different machine learning algorithms such as Support Vector Machine, k-Nearest Neighbors and Decision Tree were used for classification [1]. These techniques worked to some extent but the results were not always consistent. Changes in lighting, background or even the position of the leaf sometimes affected the accuracy of the system.

Chebrolu et al. 2025 worked on tomato leaf disease detection using both deep learning and traditional machine learning methods. Their study showed that deep learning models can provide better accuracy and make the detection process more automatic compared with earlier approaches [1]. Santhiya et al. 2024 studied different CNN based models such as GoogLeNet, AlexNet and a custom CNN architecture for detecting tomato leaf diseases. Their results indicated that CNNs are capable of giving reliable classification results when trained with proper leaf image datasets [2].

Karthigaiveni et al. 2024 presented a method based on transfer learning for plant disease detection. Using pre-trained deep learning models can improve classification accuracy while reducing training time compared with building a model from scratch [3]. Sharma et al. 2023 suggested a CNN based method focusing on texture-related features for classifying different plant diseases. Deep features learned by CNN models can improve the performance of multiclass tomato plant disease classification [4].

Yadav et al. 2025 also developed an automated system using CNN for detecting diseases in tomato leaves. Their work demonstrated that CNNs can successfully identify multiple types of tomato leaf diseases when the model is trained on a suitable dataset [5]. Sladojevic et al. showed that CNN based models can classify several plant diseases from leaf images and their performance was better than many traditional approaches [2].

One widely used dataset in this area is the PlantVillage dataset which was created to support research in image based plant disease prediction [3]. Mohanty et al. used this dataset with deep CNN architectures and achieved very high classification accuracy under controlled conditions [4]. Their research also mentioned that models trained only on laboratory-style images may not always perform equally well in real field environments.

To improve model performance and reduce training effort, many researchers started using transfer learning with pre-trained models such as VGG16, ResNet and MobileNet. Too et al. carried out a comparison of several fine-tuned deep learning models and found that transfer learning can achieve good accuracy even when the training dataset is not very large [5]. Ferentinos also evaluated different CNN models for plant disease recognition and confirmed that deep learning approaches can work effectively for different crops and disease types [6].

Apart from transfer learning some researchers have also proposed lightweight CNN models that try to maintain a balance between accuracy and computational cost. Data augmentation methods like rotating, flipping or scaling images are often used during training so that the model can generalise better and avoid overfitting. The performance of these systems is usually measured using accuracy, precision, recall, F1-score and confusion matrix [6][7]. Because of these challenges there is still a need for a practical and efficient tomato leaf disease detection system.

## III. DATASET DESCRIPTION

Fig. 1. illustrates the distribution of images across different tomato leaf disease categories and the healthy class in the dataset. Each class contains 10,000 images, resulting in a perfectly balanced dataset with equal representation of all categories. Such a balanced distribution helps prevent model bias toward any particular class, improves learning consistency, and contributes to better classification accuracy and generalization performance across all disease types.

Automated plant disease detection using leaf images has become a topic many researchers are exploring in recent years. Earlier studies mostly used traditional image processing methods. In those methods features like colour, texture and shape were taken from the leaf images and then different machine learning algorithms such as Support Vector Machine, k-Nearest

Automated plant disease detection using leaf images has become a topic many researchers are exploring in recent years. Earlier studies mostly used traditional image processing methods. In those methods features like colour, texture and shape were taken from the leaf images and then different machine learning algorithms such as Support Vector Machine, k-Nearest

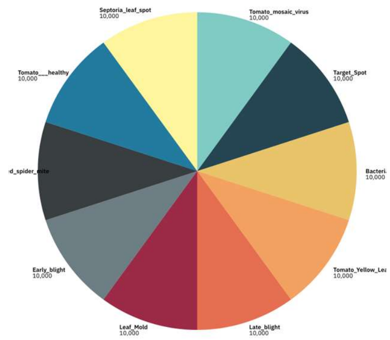


Fig. 1. Dataset Distribution

## III. METHODOLOGY

The proposed system is designed to identify diseases in tomato leaves automatically using a CNN based deep learning method. The work is carried out in several steps such as collecting the dataset, preparing the images, applying augmentation, designing the CNN model, training the model and then evaluating its performance. Each step is done carefully so that the system can detect diseases in tomato leaves in a reliable way. Fig 2. Shows the Architecture and methodology of the system.

### *A. Dataset Collection*

For this work a dataset of tomato leaf images was used which contains both healthy leaves and diseased leaves. The images were taken from publicly available plant disease datasets such as the PlantVillage dataset. This dataset contains common tomato leaf diseases like Early Blight, Late Blight and Leaf Mold. Having different types of leaf images helps the model understand the patterns that appear when a leaf becomes infected.

### *B. Image Preprocessing*

Before giving the images to the CNN model some preprocessing steps were carried out. All images were resized to the same dimension so that the model receives uniform input. Image normalisation was also applied so that pixel values remain in a suitable range during training. Basic image improvement methods were used to make the leaf spots, colour changes and damaged areas easier to notice.

### *C. Data Augmentation*

To increase the size of the training dataset and make the model more stable, some data augmentation techniques were used. The training images were rotated, flipped and slightly zoomed. In some cases the images were shifted a little as well. These steps help the model learn from different variations of the same image which is similar to real situations where leaf images may look different because of lighting or orientation.

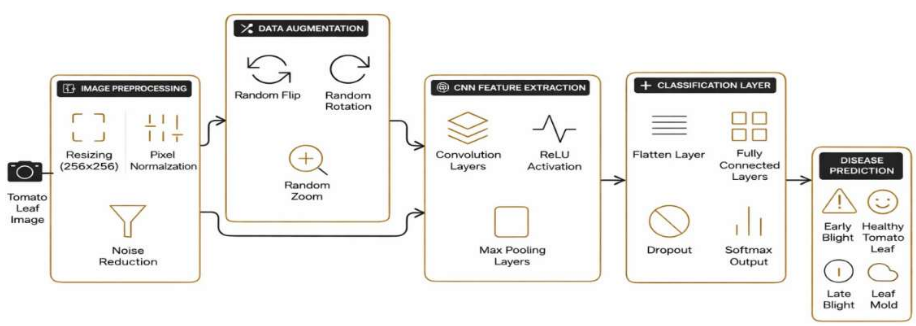


Fig. 2. Architecture

### D. CNN Model Architecture

A custom CNN model was built for classifying tomato leaf diseases. The model includes several convolution layers along with activation functions and max pooling layers. These layers help in extracting useful patterns from the images such as texture, shape and colour distribution of the leaf. After this the extracted features are flattened and passed to fully connected layers. In the final layer a Softmax function is used to decide the disease category of the leaf image.

### E. Model Training

The prepared dataset was divided into training data and validation data. The CNN model was trained using a suitable optimiser and loss function so that the prediction error becomes smaller. Training was carried out for several epochs while observing the training accuracy and validation accuracy. Some parameter tuning and early stopping were also used to improve the model performance and reduce overfitting.

### F. Performance Evaluation

After the training process the model was tested using a separate test dataset. The performance was measured using different evaluation metrics such as accuracy, precision, recall, F1-score and confusion matrix. These measurements help in understanding how correctly the model identifies the diseases in tomato leaves.

### G. System Deployment

After testing, the trained CNN model was connected to a simple interface so that it can be used easily. In this system a user can upload a tomato leaf image and the system will provide the predicted disease result. This shows how the model can be used in real situations where farmers or users may want to check plant health quickly.

## IV. EXPERIMENTAL RESULTS

This CNN-based model was built using TensorFlow and Keras and trained on a dataset of tomato leaf images. The dataset is divided into two parts: a training set and a validation set. The model was trained for 40 epochs with a batch size of 32. Before training, image preprocessing steps like resizing, normalisation and data augmentation were applied to help the model learn better and improve performance.

During the training process the model showed a steady improvement in accuracy. In the beginning, training accuracy increased quickly and then slowly became stable. At the end of training the model reached 95–96% accuracy. This shows that CNN is able to learn useful features from the tomato leaf images like colour, patterns, texture and disease spots.

The validation accuracy followed a similar pattern to training accuracy. It changed slightly in some epochs because of data augmentation and variety in datasets but it stayed above 90% in later training. This indicates that the model works well on new images and does not overfit too much on training data. The loss graph also supports these results. The training loss decreased smoothly during the training process showing stable learning. The validation loss had minor fluctuations in some epochs but overall it also decreased and stayed low.

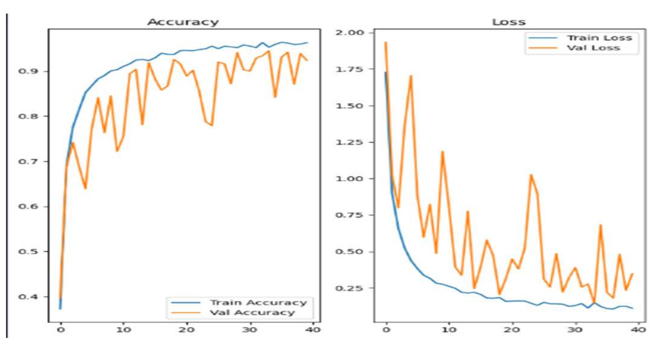


Fig. 3. Accuracy and Loss

Fig. 3. show that the model learns effectively during training: training accuracy steadily increases while training loss continuously decreases, indicating successful learning. The validation accuracy remains high (around 80–90%) but fluctuates, and the validation loss is more unstable, suggesting slight overfitting and sensitivity to the validation data, although the model still generalizes reasonably well.

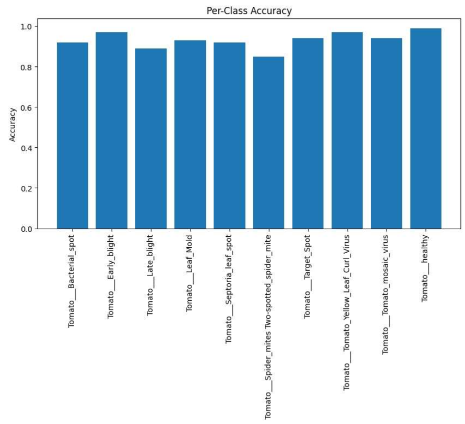


Fig. 4. Per Class Accuracy

Fig. 4. shows consistently high performance across all classes, with accuracies ranging from approximately 85% to 99%. Most classes achieve over 90% accuracy, indicating that the model effectively distinguishes between different tomato diseases and healthy leaves. The lowest accuracy is observed for Spider Mites (~85%), suggesting it is the most challenging class for the model.

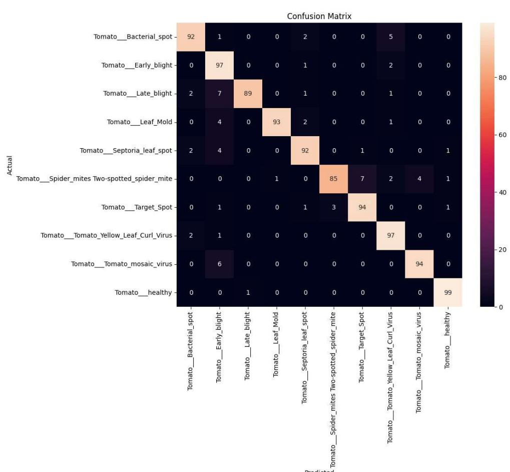


Fig. 5. Confusion Matrix

Fig. 5. shows strong classification performance, with most predictions concentrated along the diagonal, indicating correct classification of tomato leaf diseases. Only a few samples are misclassified between similar disease categories, while the Healthy class achieves near-perfect accuracy. Overall, the model demonstrates high reliability and effective disease differentiation across all classes.

To test the model further, some sample predictions were made using validation images. The model correctly identifies the disease pattern in tomato leaves and can classify unseen images correctly after training. The final model is saved to be used later for deployment.

Table 1. demonstrates strong model performance, achieving an overall accuracy of 92% on 1,000 test samples. Most classes show high precision, recall, and F1-scores above 0.90, indicating effective identification and classification of various tomato leaf diseases.Tomato – Leaf Mold achieved the highest F1-score (0.96), while Tomato – Healthy attained a perfect recall of 1.00, meaning all healthy leaf samples were correctly identified. Tomato – Late Blight also showed excellent detection capability with a recall of 0.99. However, Tomato – Early Blight recorded the lowest recall (0.84), suggesting that some Early Blight samples were misclassified as other diseases. The macro-average and weighted-average F1-scores of 0.92 indicate balanced performance across all classes, confirming that the model generalizes well and maintains consistent classification accuracy for both diseased and healthy tomato leaves.

TABLE I. MODEL PERFORMANCE METRICS (CLASSIFICATION REPORT)

| Class Name | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Tomato – Bacterial Spot | 0.95 | 0.94 | 0.94 | 100 |
| Tomato – Early Blight | 0.91 | 0.84 | 0.88 | 100 |
| Tomato – Late Blight | 0.82 | 0.99 | 0.90 | 100 |
| Tomato – Leaf Mold | 0.98 | 0.95 | 0.96 | 100 |
| Tomato – Septoria Leaf Spot | 0.95 | 0.88 | 0.91 | 100 |
| Tomato – Spider Mites | 0.91 | 0.90 | 0.90 | 100 |
| Tomato – Target Spot | 0.94 | 0.92 | 0.93 | 100 |
| Tomato – Yellow Leaf Curl Virus | 0.99 | 0.90 | 0.94 | 100 |
| Tomato – Mosaic Virus | 0.97 | 0.92 | 0.94 | 100 |
| Tomato – Healthy | 0.86 | 1.00 | 0.93 | 100 |
| Overall Accuracy | | | 0.92 | 1000 |
| Macro Average | 0.93 | 0.92 | 0.92 | 1000 |
| Weighted Average | 0.93 | 0.92 | 0.92 | 1000 |

## V. CONCLUSIONS

In this work a CNN-based deep learning model was developed to automatically identify diseases in tomato leaves using image data. The model learns important visual patterns from leaf images and classifies them into healthy and diseased categories. Image preprocessing and data augmentation techniques help to improve the reliability and performance of the system.

The results show that this model can detect tomato leaf diseases with good accuracy using standard evaluation methods. Compared to traditional methods this approach reduces the need for manual checking and gives faster and more accurate results. Overall, the study shows how deep learning is useful in agriculture. This model is simple and cost effective, helping farmers detect plant disease at an early stage. Early detection helps farmers take proper decisions and manage crops more effectively.


## ACKNOWLEDGMENT

The authors wish to thank the Department of Computer Engineering, VIT Pune for their support. This work was supported by publicly available datasets from the PlantVillage repository.


## REFERENCES


[1] M. Chebrolu, K. Garikapati, Y. Veeramachaneni, J. L. Annabathina, S. K. Mishra, and S. K. Mishra, "Tomato leaf disease detection using deep learning and machine learning," in Proc. 2025 Int. Conf. on Artificial Intelligence and Machine Vision (AIMV), IEEE, 2025.

[2] S. S. T. Santhiya, T. Senthil Kumar, P. Jayadharshini, P. V. Rohit, V. Sriram, and N. S. Sreepriyanth, "Automated detection of tomato leaf disease using convolutional neural networks: Evaluation of GoogLeNet, AlexNet, and custom CNN architectures," in Proc. 4th IEEE Int. Conf. on ICT in Business, Industry and Government (ICTBIG), IEEE, 2024.

[3] S. Karthigaiveni, C. Madhubala, G. Avinash, E. Saranya, P. Rohini, and G. Yogapriya, "Plant disease detection and classification using transfer learning," in Proc. Int. Conf. on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), IEEE, 2024.

[4] G. Sharma, V. Anand, and S. Gupta, "Texture features retrieval using convolutional neural networks for multiclass classification of plant diseases," in Proc. 2023 Int. Conf. on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), IEEE, 2023.

[5] H. K. Yadav, V. Srivastava, S. Banoth, R. Rohini, D. Rawat, and P. Rawat, "Automated detection of tomato leaf diseases using convolutional neural networks," in Proc. 2025 Int. Conf. on Automation and Computation (AUTOCOM), IEEE, 2025.

[6] J. G. A. Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," SpringerPlus, vol. 2, no. 1, pp. 1–12, 2013.

[7] 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, Art. ID 3289801, 2016.

[8] D. P. Hughes and M. Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," arXiv preprint arXiv:1511.08060, 2015.

[9] S. P. Mohanty, D. P. Hughes, and M. Salathe, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, Art. no. 1419, 2016.

[10] E. C. Too, Y. J. Li, S. Njuki, and Y. C. Liu, "A comparative study of fine-tuning deep learning models for plant disease identification," Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.

[11] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.
[12] A. Kamilaris and F. X. Prenafeta-Boldu, "Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.