

Enhancing Crop Health and Early Detection of Tomato Leaf Diseases Using Deep Learning Techniques

Prashant Vaishnav¹, Amit Kumar Saxena^{1*}

¹ Department of CSIT, Guru Ghasidas Vishwavidyalaya Bilaspur, India

 OPEN ACCESS

Received: 13/02/2026

Accepted: 09/04/2026

Published: 22/04/2026

Citation: Vaishnav P, Saxena AK (2026) Enhancing Crop Health and Early Detection of Tomato Leaf Diseases Using Deep Learning Techniques. Indian Journal of Science and Technology 19(15): 1017-1028. <https://doi.org/10.17485/IJST/19i15.218>

* Corresponding author.

amitsaxena65@rediffmail.com

Funding: None

Competing Interests: None

Copyright: © 2026 Vaishnav & Saxena. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.indjst.org/))

ISSN

Print: 0974-6846

Electronic: 0974-5645

Abstract

Background/Objectives: One of the most widely produced and consumed crops in the world, tomatoes are often threatened by various leaf diseases, including early blight, late blight, and leaf mold, which can result in significant yield losses if not detected and managed promptly. Due to reliance on manual inspection, tomato leaf diseases are often detected late, resulting in significant crop loss. The goal of this research is to develop a deep learning (DL) model that accurately classifies tomato leaf diseases. The model is trained using supervised learning on the PlantVillage dataset, which includes labelled images of tomato leaves under different conditions. **Method:** Gaussian Blurring-Gaussian Mixture Models (GB-GMM) are a preprocessing method used to enhance the image quality. While EfficientNet and VGGNet architectures are utilised for accurate classification, data augmentation is used to boost model robustness. A standard train-validation-test split is used to assess the model. **Findings:** The results demonstrate that the proposed method, implemented in Python, performs better than the others in terms of accuracy (94.3%), precision (93.7%), recall (92.5%), and F1-score (93.1%) in VGGNet architectures. These results indicate that the proposed model is very effective overall and produces balanced predictions. In the future, the system may be integrated with drone and IoT technologies for automatic disease warnings and real-time field surveillance. **Novelty:** The Multivariable Grey Prediction Evolution Algorithm (MGPEA) is included for illness trend forecasting to enhance predictive power further. This technology facilitates large-scale, sustainable agricultural management, minimizes human inspection, and enables prompt disease response.

Keywords: Tomato Leaf Diseases; Crop Health; Gaussian Mixture Models; Multivariable Grey Prediction Evolution Algorithm; EfficientNet; VGGNet

1 Introduction

The cultivation of tomatoes traverses essential global agricultural systems for food supply and economic stability. The production and quality of tomato crops are significantly reduced by various kinds of leaf diseases caused by bacteria, fungi, and viruses^{(1), (2)}. To manage a field and avoid significant crop loss, early disease diagnosis and classification are essential. Conventional disease detection techniques are time-consuming, tedious, and sometimes inaccurate⁽³⁾. Use of machine learning algorithm help to identify different features⁽⁴⁾. The potential for automating detection using image analysis is demonstrated by the increasing maturity of DL^{(5), (6)}. This study applies DL to classify and categorise tomato leaf disease at an early stage to maintain healthy crops, reduce losses during production, and encourage sustainable agriculture⁽⁷⁾. Various leaf diseases affect tomato plants, significantly reducing yields and quality and affecting both food security and economic value⁽⁸⁾. The procedure of manually identifying these illnesses is time-consuming and needs specialised expertise⁽⁹⁾. Inaccurate analysis is more likely to occur in the early phases. Treatment errors or delays might cause the illness to spread quickly. Misdiagnosis leads to increased pesticide usage, which raises production costs. Developing an effective, accurate, and automated method for the early detection and categorisation of illnesses is the challenge⁽¹⁰⁾. DL, namely, convolutional neural networks, may be the answer to identifying disease trends in leaf image analysis. Machine Learning algorithm may help to predict in early stages⁽¹¹⁾.

The challenge is to create a reliable, accurate DL-based model that can learn to classify various tomato leaf diseases. This would enable prompt intervention and increase overall crop yield. The urgent need to improve food security and sustainability in the face of climate change and a growing global population led to the development of a DL-based technique for TLD identification⁽¹²⁾. Food production is seriously threatened by tomato diseases, which affect not only the supply of affordable food but also farmers' livelihoods^{(13), (14)}. The shortcomings of traditional methods highlight the need for innovative solutions. DL offers a powerful approach to overcoming these limitations by leveraging the vast amounts of visual data collected from tomato fields^{(14), (15)}. The potential to significantly lower crop losses, improve resource utilisation, and promote more environmentally friendly agricultural practices is a compelling reason to pursue this study^{(16), (17)}. Use of technical innovation IoT devices to overcome the TLD dilemma by primary data⁽¹⁸⁾. The primary objective is to create a robust deep learning model specifically designed to recognise tomato leaf diseases from photos⁽¹⁹⁾. The model must reliably distinguish among many TLDs to achieve high classification accuracy, which is necessary for targeted treatment. To identify diseases before their external symptoms spread and enable preventive intervention, the project also aims to demonstrate how the model can be utilised for early disease detection. Create a system that uses this model and may be incorporated into real-world agricultural environments, enabling farmers to access and use this technology in their day-to-day work.

In order to identify tomato leaf disease, Patwal et al.⁽²⁰⁾ used Colour Moments to collect Skewness, Mean, and Standard Deviation. Images of leaves with Septoria Leaf Spot, early blight, and late blight are used to extract and segment features. The results provide a framework for classifying plant diseases and demonstrate the adaptability of Colour Moments for image indexing. Additionally, accurate and timely diagnosis is made possible by the image processing, which is crucial for disease management. To develop treatments, Ramu et al.⁽²¹⁾ used machine learning algorithms to identify early and late blight in potato and tomato. The findings demonstrate that ML technology can help address various agricultural challenges, such as reducing the impact of tomato and potato diseases.

2 Methodology

This section describes the proposed method, dataset description, model development, and experiment parameters.

2.1 Proposed Method

Tomato leaf disease classification thus involves a DL-based approach incorporating image segmentation and neural network modelling for crop health and productivity enhancement. High-resolution images will be considered based on agricultural environments involving both healthy and diseased samples of tomato leaves. Following this, image preprocessing will be undertaken to develop image models using resizing, normalization, noise reduction, and any other necessary image enhancement. A CNN architecture will be designed and trained on the labelled dataset to notice and classify shared tomato leaf diseases, are early blight, late blight, and leaf mold. The intention is to construct a robust, scalable system for real-time diagnosis, which will help farmers in taking timely and informed decisions, and as a result, improve tomato crop yield and reduce losses due to disease.

The classification of tomato leaf diseases using DL involves a structured pipeline that starts with data collection, primarily from the PlantVillage dataset, as shown in Figure 1. Next, apply Preprocessing methods such as GB-GMM to efficiently separate tomato leaves by removing background noise. Advanced DL architectures such as VGGNet and EfficientNet are employed for

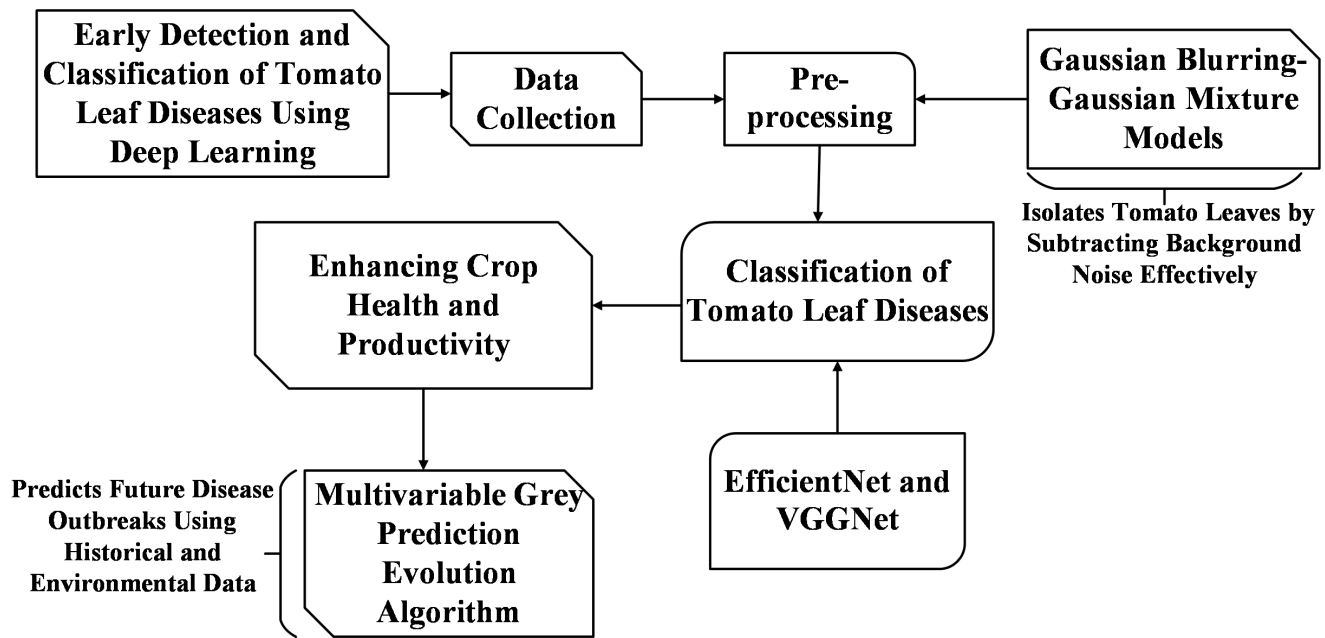


Fig 1. Block Diagram of the Proposed Work

classification due to their high computational efficiency and accuracy in large-scale image tasks. To further enhance crop health and productivity, the MGPEA is integrated to predict future disease outbreaks using historical and environmental data, enabling timely and proactive disease management.

2.2 Data Collection

The data collection process involves obtaining images from the Plant Village dataset [<https://www.kaggle.com/datasets/naveedgull/tomato-leaf-disease/data>], which serves as the primary source. This dataset, which is a collection of labelled images, is extremely popular. Every picture is classified as depicting a specific illness or as healthy. The information includes common tomato plant illnesses, such as leaf mold and, early and late blight. The dataset's images include a wide range of visual conditions, including different lighting, leaf orientations, and disease stages, all of which should support the deep learning model across a broad range of potential real-world situations. The dataset's enormous richness makes it ideal for developing a reliable deep learning model that can correctly identify and classify illnesses in practical settings. This freely available, well-labelled dataset provides a solid basis for the supervised learning model.

Figure 2 indicates the prevalence of diseases in the dataset indicates the number of different conditions the tomato plants have. Yellow leaf curl virus is the most common, with 5,332 samples, thereby being highly frequent. Bacterial spot (2,127 samples), late blight (1,909 samples), and septoria leaf spot (1,771 samples) also have high presence in the dataset and are hence among the more common diseases in the tomato crop. Early blight (1,000 samples), leaf mold (952 samples), and target spot (1,404 samples) diseases have relatively moderate frequencies. Spider mites (1,676 samples) are also part of the overall dataset, and these pests infest the health of the plants. Meanwhile, viral diseases such as mosaic virus (373 samples) are less common. Finally, the dataset is composed of healthy tomato leaf samples (1,591), which can act as a control for comparison.

2.3 Preprocessing

Since the DL model requires clean, consistent, and information-rich inputs, preprocessing the images is an essential next step after data collection. Pixel values are resized and normalised to 0-1, which increases model efficiency during training and speeds up convergence. Depending on the situation, noise reduction methods such as GB or histogram equalisation might be used when picture quality is compromised, or the backdrop is not required—background subtraction using GB-GMM helps separate the leaf from unimportant background noise. By focusing exclusively on the relevant areas of the picture, GB-GMM can precisely separate the foreground (the tomato leaf) from the background. Additionally, data augmentation methods, including brightness adjustments, zooming, horizontal and vertical rotation, and random rotation, are used.

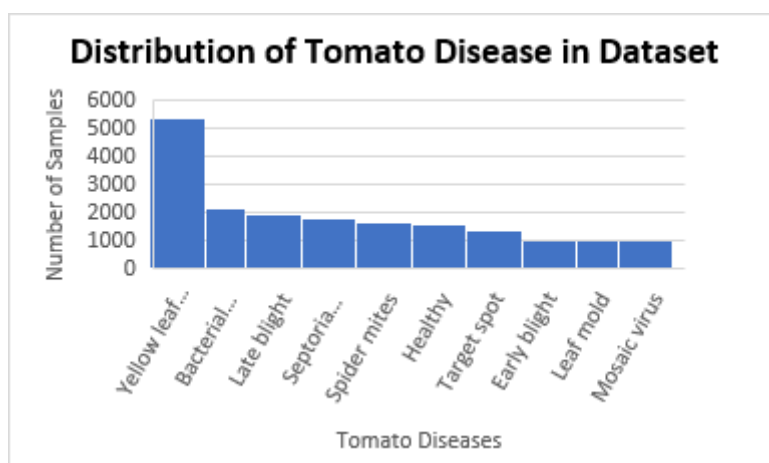


Fig 2. Approximate Distribution of Images 10 Classes

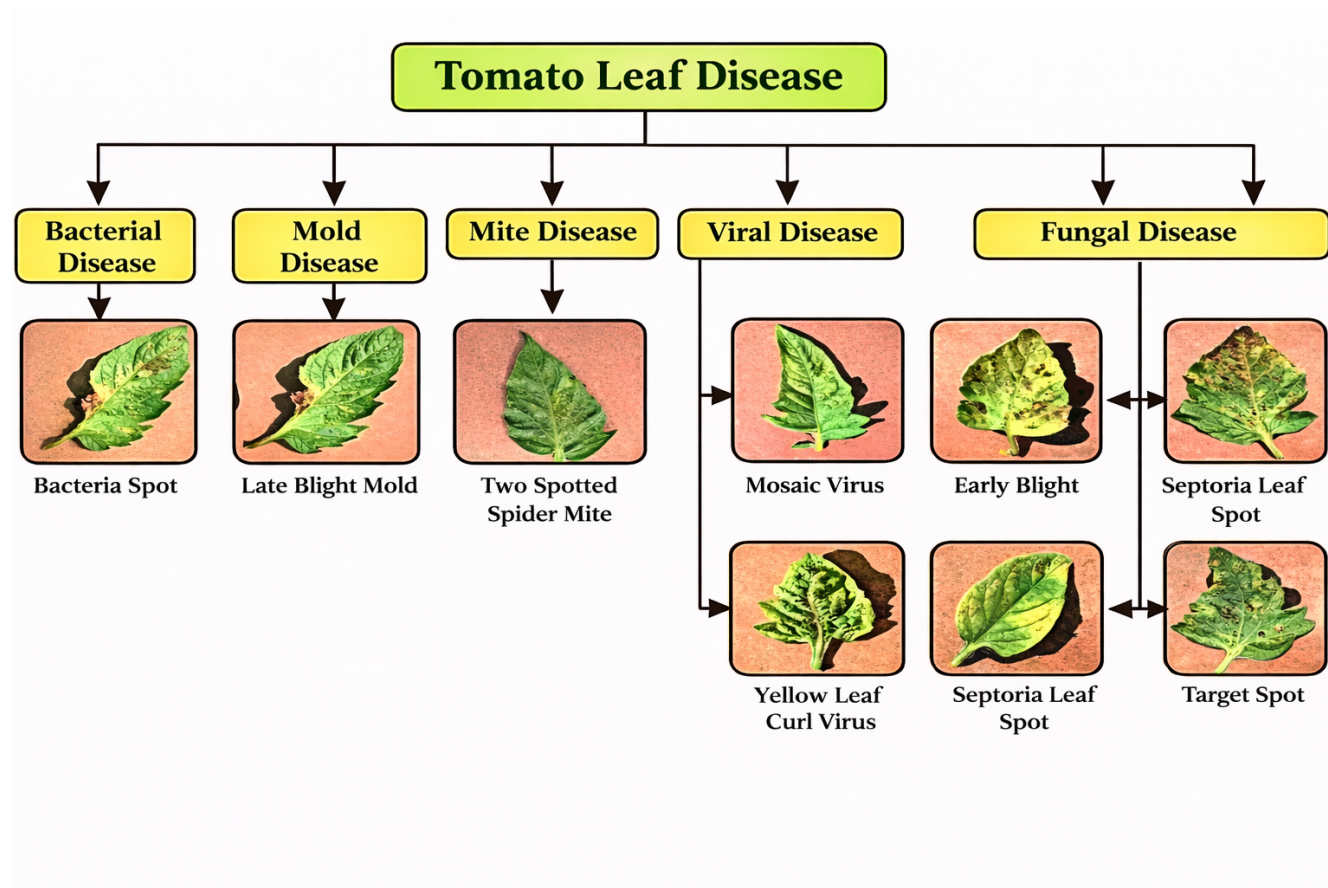


Fig 3. Hierarchical Classification of Tomato Leaf Diseases

A hierarchical classification of TLD is provided in Figure 3, which divides them into five primary categories: bacterial illness, Mold disease, mite disease, viral disease, and fungal disease. Examples of certain illnesses and accompanying pictures of tomato leaves affected by them are used to highlight each category better. Fungal diseases include early blight, Septoria leaf spot, leaf Mold, and target spot; bacterial diseases exhibit bacterial spots; mold diseases show late blight Mold; mite diseases show damage from two-spotted spider mites; and viral diseases include yellow leaf curl virus and mosaic virus. Based on the causative agents and distinctive symptoms on the foliage, this visual categorization helps distinguish between different forms of TLD.

2.3.1. Gaussian Blurring-Gaussian Mixture Models (GB-GMM)

Improving agricultural production and health requires early detection and identification of TLD. This can be accomplished, for example, by combining DL models with advanced image processing techniques, such as GB-GMM. These techniques provide a well-defined approach to informing or enhancing the accuracy of detection systems while enabling efficient image preprocessing of plant disease pictures.

GB is used for image processing in medical and agricultural fields to reduce noise and enhance details. This work uses GB to smooth the image and assist in identifying pertinent features for detecting sick spots in a TLD detection procedure. Using a Gaussian kernel, a matrix generated from the Gaussian function is used to convolve a picture. The Gaussian function is defined as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (1)$$

Where (x, y) are the positions of the kernel, and σ is the standard deviation, which controls the amount of blurring. The kernel is applied to every pixel in the image, and the output is a smoothed version where high-frequency noise and small details are removed.

The main aim of GB in the disease detection pipeline is to preprocess images by reducing noise, which is important for segmentation and feature extraction. This agrees with the model to categorize and focus on larger structures. Otherwise, the model may focus on irrelevant details within the image, which could affect its performance.

GMM are a probabilistic model that assumes that all the data points in a dataset are created from a mixture of several Gaussian distributions. GMM is applied to model the pixel intensity distributions of healthy and diseased areas in the image. The GMM is described by the following equation:

$$p(x) = \sum_{i=1}^K \pi_i N(x | \mu_i, \Sigma_i) \quad (2)$$

Where the probability density function $p(x)$ represents the likelihood of a data point x . In a Gaussian Mixture Model (GMM), K denotes the number of Gaussian components, and π_i is the weight of the i^{th} Gaussian component. Each component follows a normal distribution $N(x | \mu_i, \Sigma_i)$, with mean μ_i and covariance Σ_i .

Each component in the mixture represents a different cluster of pixel intensities, and the mixture model gives us a framework for modelling the complex texture and colour variations to distinguish diseased from healthy areas. Once the image has been preprocessed and GMM has been used to model the image, the following step is to extract features using DL. DL often employs CNN, which can learn hierarchical features of an image. Edge, texture, shape, and other features will help differentiate healthy from diseased tomato leaves.

The CNN applies multiple layers of convolutions, activations, and pooling operations to extract increasingly abstract features. The production from the CNN can then be passed into a dense layer to perform classification. The final step is to categorize the leaf images into separate categories, healthy or diseased, by training a classifier, or a DNN, on top of the features obtained from the images. The classifier will use features extracted from both the GMM and the features obtained using the CNN to categorize the tomato leaves disease. The classification problem can be formalized as:

$$y = f(x) \quad (3)$$

Where y is the predicted class, x is the feature vector extracted from the image using CNNs.

The model can identify early disease symptoms on tomato leaves by using GB-GMM and DL. Rapid disease detection enables you to enhance health and production early, whether by using fungicides or altering environmental factors. The primary advantage of GB-GMM is that it combines GB and GMM to reduce image noise and model the complex distribution of leaf texturing, resulting in better agricultural decision-making and increased disease-detection accuracy, especially for mild or early-stage symptoms. When combined with DL models, the GB-GMM approach offers a powerful option for TLD categorization. GB can improve feature extraction while lowering noise.

The probabilistic modelling framework that GMM uses allows it to model intricate patterns in the data. The procedures of feature extraction and classification are automated using DL. All three methods may contribute to crop yield and health, resulting in sustainable agriculture.

2.4 Classification of Tomato Leaf Diseases

The model can categorize tomato leaf diseases into multiple groups after training. When a fresh picture of a tomato leaf is fed into the algorithm, it either labels the leaf as healthy or forecasts the most likely disease class. By using cutting-edge DL architectures such as EfficientNet and VGGNet, classification performance is improved. Due to its excellent accuracy and efficient use of computation and parameters, EfficientNet is used in large-scale image categorisation, including disease diagnosis. In comparison, VGGNet provides a deep network with simple yet effective layers that aid in extracting fine information from photos and is quite good at detecting minute patterns in tomato leaf diseases. By using these structures, the model may improve its illness classification predictions by learning discriminative features from leaf pictures. When combined, these methods deliver strong performance in classifying tomato leaf diseases, ensuring the system's reliability and effectiveness in practical applications.

2.4.1. EfficientNet

Google AI's EfficientNet deep learning architecture was designed with the dual goals of achieving high accuracy and low computing cost. It is the ideal model for agricultural applications, such as TLD categorisation, because of its focus on efficiency. By using EfficientNet in the field to identify plant illnesses early, farmers may improve crop health and productivity while using fewer pesticides. Network depth, breadth, and resolution are all individually scaled using EfficientNet's novel scaling technique. One or two of these factors are successfully scaled by existing models, which often results in a sharp rise in computing cost. On the other hand, EfficientNet balances all three using a compound-scale approach, enabling much more efficient resource use. The compound scaling rule for EfficientNet is given by:

$$\text{Scaling factor} = \phi^\alpha, \text{ where } \alpha \in \{w, d, r\} \quad (4)$$

Where, ϕ is a base scaling factor, and w , d , and r represent breadth, distance, and resolve scaling, respectively. This allows EfficientNet to scale more effectively compared to traditional models, optimizing performance without excessive computation.

2.4.2. VGGNet

The classification of diseases in crops, specifically tomato plants, is important for keeping crops healthy and increasing productivity. A very successful method that is being applied to do this is DL, and the most successful models for image classification are VGGNet. VGGNet is a CNN and has been used in multiple areas of research with application to disease recognition in plants. The structural design of VGGNet is simple yet effective and is applicable for the automatic classification of leaf diseases which allows for early intervention in bringing down crop health to improve crop health.

VGGNet was created through the Visual Geometry Group, and it uses a more extended architecture than most other deep networks, with many stacked convolutional layers, meaning it can extract very complex hierarchical features of images. From a design standpoint, the underlying assumption of this architecture is that deeper networks can potentially model complex representations and patterns. While VGGNet is typically recognized by its use of very small receptive fields (3x3 filters), VGGNet is most notably characterized by a longer network comprised of a massive amount of convolutional layers. The architecture includes convolutional layers, max-pooling layers, dense layers and a softmax output layer, whereas the softmax layer outputs the probabilities of each possible disease class being classified (i.e. diseased and healthy).

The primary step in VGGNet is the convolutions of the image input. Filters, or kernels that participate in this operation, find very simple features like edges and textures, which will later be combined in deeper layers to find very complex structures. The convolution operation is as follows:

$$S(x, y) = \sum_{i=-1}^1 \sum_{j=-1}^1 I(x+i, y+j) W(i, j) \quad (6)$$

where $W(i,j)$ is the filter weight and $I(x,y)$ is the image intensity at point (x,y) . By moving the filter over the picture, this formula computes the output feature map $S(x, y)$. To detect infections early, the filters in the initial layers usually capture low-level features of tomato leaves, such as edges and textures.

A nonlinear activation function is applied to the output of each convolutional layer after the convolution process. The following is the definition of the Rectified Linear Unit activation function that VGGNet uses:

$$ReLU(x) = \max(0, x) \quad (7)$$

ReLU helps the model discover more intricate patterns in the data by adding non-linearity. This is crucial for distinguishing between healthy and diseased tomato leaves, which may differ only slightly in texture or colour. Since VGGNet uses expensive feature maps, max pooling is also used to downsample them and improve computational efficiency. Max pooling preserves the most significant observations while reducing the spatial resolution of the feature map. One way to represent the pooling procedure is:

$$P(x, y) = \max_{i,j \in \text{pool region}} F(x+i, y+j) \quad (8)$$

Where $P(x,y)$ is the pooled feature map and $F(x,y)$ is the feature map after convolution. By reducing the number of parameters and processing cost, the pooling method improves network efficiency while preserving essential feature for illness diagnosis. The output feature maps are flattened into a vector after a few convolutional and pooling layers, and then they go through dense layers. The feature eliminated by the previous levels are combined in these layers using linear transformations. One way to depict a completely connected layer is as:

$$h = W \cdot x + b \quad (9)$$

Where W is the weight matrix, x is the input feature vector, b is the bias term, and h is the layer's output. By combining learnt characteristics, the fully connected layers aid the final classification decision. A probability distribution is produced by the softmax layer, the last layer of VGGNet, from the output of the last fully connected layer. The definition of the softmax function is:

$$P(y = i | x) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (10)$$

where C is the total number of classes and z_i is the class I score. The tomato leaf is classified as healthy or ill using the softmax layer's output, which indicates the likelihood that a given image belongs to each class. VGGNet's deep architecture enables it to learn complex representations from images of tomato leaves automatically.

The architecture of a CNN, likely VGG16, used for tomato leaf disease classification. Starting with an input image of size 224x224x64, the network consists of multiple convolutional layers interspersed with Max Pooling layers (indicated by red blocks), which progressively extract features and decrease the spatial dimensions of the feature maps. The quantity of feature maps increases through the convolutional blocks. Following the convolutional and pooling layers, the high-dimensional feature representation is flattened and fed into a series of fully connected layers, with a decreasing number of neurons. While the fully connected layers, also using ReLU (shown in green), perform the final classification into various tomato leaf disease categories, the convolutional layers, in conjunction with ReLU activation (shown in blue), learn hierarchical features from the input image. This is represented by the output of the final fully linked layer (fc8) with 1000 features, which is usually followed by a softmax layer for probability output.

2.5 Enhancing Crop Health and Productivity

The categorisation scheme directly impacts refined crop yield and health. Early disease detection by farmers enables prompt actions like targeted pesticide treatment, optimised irrigation, or selective pruning, all of which reduce crop loss and increase total yields. Additionally, the system saves time and money by eliminating the need for expensive, time-consuming human examinations, thanks to its high categorisation accuracy. By accurately forecasting possible disease outbreaks based on crop conditions, environmental variables, and past disease data, the MGPEA may further improve the system. By modelling and forecasting patterns using historical data, MGPEA provides farmers with essential insights into the potential evolution of illnesses, enabling them to take preventative action. Together with the disease detection system, this predictive capacity helps farmers manage their crops more sustainably and effectively. The technology may be used to monitor tomato fields at scale, enabling simultaneous management of many fields and increasing operational efficiency. Adoption of this integrated technology may lead to more productive farming methods, better tomato yields, and lower costs, all of which will enhance the overall prosperity and sustainability of agricultural enterprises.

2.5.1. Multivariable Grey Prediction Evolution Algorithm (MGPEA)

For crops like tomatoes, early disease identification and categorisation are essential to improving crop health and yield. DL and advanced optimisation techniques, such as the MGPEA, are very beneficial for disease prediction. Effective modelling and prediction of illness development based on a range of ecological and image-based factors are enabled by MGPEA, which

combines evolutionary algorithms with the power of grey prediction models. By providing accurate, timely information for better disease control, this hybrid approach increases crop productivity and quality.

A well-liked technique for forecasting systems with insufficient or ambiguous data is the Grey Prediction Model (GM). It is based on the idea that a few prior observations can be used to forecast future data patterns. Grey system theory, in which the system state is unknown or only partly understood, is the fundamental concept of grey prediction. Grey prediction is used in MGPEA to estimate disease progression by analyzing leaf attributes from photos or prior illness data. The grey prediction equation is given by:

$$x^{(0)}(k) = [x(1), x(2), \dots, x(k)] \quad (11)$$

Where $x^{(0)}(k)$ represents the original data sequence, and $x(k)$ is the observation at time k .

The main goal of the grey model is to use this data sequence to develop a prediction model. The original data sequence is subjected to the accumulated generating operation (AGO) to reduce the model's uncertainty. The AGO is defined as:

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad (12)$$

Where $x^{(1)}(k)$ is the accumulated data sequence. This stage helps smooth the data and improve its suitability for forecasting. By accounting for a range of variables that may affect the development of tomato plant illnesses, the multivariable grey prediction model extends the classical grey model. These variables include the surrounding environment and various metrics of leaf features derived from photos. The following is an expression for the multivariable grey prediction equation:

$$x^{(1)}(k) = \sum_{i=1}^K x_i^{(1)}(k) \quad (13)$$

Where $x^{(1)}(k)$ is the accumulated sequence of the multivariable data, and K represents the number of variables considered for prediction. This multivariate approach increases the accuracy of sickness detection and prediction by enabling a more comprehensive model that accounts for interactions among multiple disease-related factors.

3 Results and Discussion

Thousands of annotated photos of different disease types, including blight and leaf mold, as well as healthy leaves, were used in the evaluation for the TLD classification. The images were preprocessed and augmented to improve generalization, and a CNN model was trained and tested using an 80:20 data split. Comparative analysis with traditional ML algorithms confirmed the superior performance of the proposed DL approach. The results support the model's potential as a practical tool for real-time, automated diagnosis of TLD, enabling early intervention, reducing crop loss, and ultimately enhancing crop health and output.

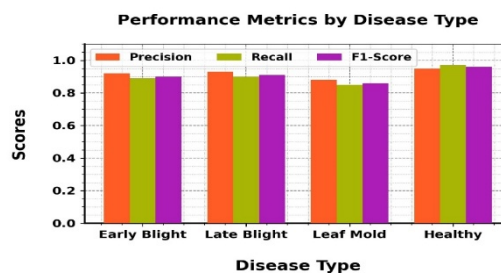


Fig 4. Monitoring the Disease Type in Leaf

Figure 4 demonstrates the optimization of disease type classification of tomato leaves using important metrics. Precision is a metric used to compute the accuracy of positive predictions from the model. The value is greater, showing fewer false positives. The value of Recall demonstrates how well the model does at identifying the true positives, and a large value shows fewer false negatives. F1-Score is the vocal mean of Recall and Precision and is a middle value of both names. The diseases as named in the table are Leaf Mold, Early Blight, Late Blight, and Healthy. Early Blight has a very high Precision (0.92), i.e., most of its correctly classified instances, and Recall (0.89), i.e., most of its instances. Its F1-score of 0.90 confirms that its Precision and Recall are optimally balanced. Late Blight is the closest with a Precision of 0.93 and a Recall of 0.90, yielding an F1-score of 0.91. Leaf Mold lags by a mere F1-Score of 0.86 and can be better. The healthy leaves provide the highest Precision (0.95) and Recall (0.97) with

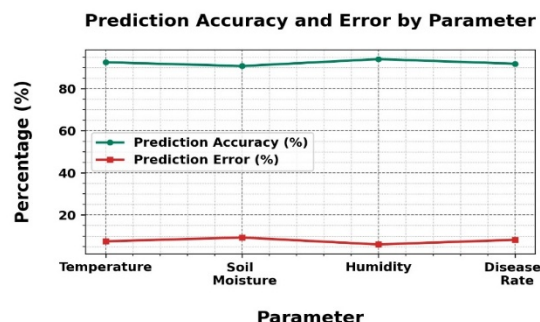


Fig 5. Prediction of Accuracy and Error Parameter

an F1 Mark of 0.96, which confirms the high potential of the modules to precisely classify healthy leaves. Its chart represents the potential of the model to differentiate among various classes of tomato leaf illness.

Figure 5 displays the MGPEA model graph which demonstrates the prediction accuracy and error rates for four important parameters: Temperature, Soil Moisture, Humidity, and Disease Rate. Every parameter is linked to the highest degree of accuracy of prediction, implying the model's capability to predict these factors with high reliability. The temperature model has a 7.5% forecast error and a 92.5% prediction accuracy. The soil moisture parameter has an accuracy of 90.7% and an error rate of 9.3%. Humidity shows the highest degree of accuracy, with a prediction error of 6.0% and a veracity of 94.0%. The illness Rate model has an 8.2% error rate and a 91.8% accuracy rate in predicting the illness rate. The graph demonstrates how the MGPEA model is helpful for agricultural or environmental monitoring systems, as it can forecast certain environmental features with a small error.

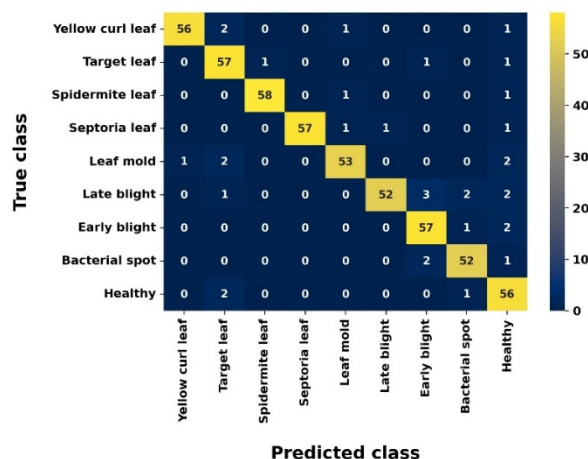


Fig 6. Confusion Matrix of Predicted Class and True Class

Figure 6 displays the confusion matrix graph which represents the presentation of an arrangement model by linking the predicted class labels with the true class labels. It helps to visualise how fine the models are at classifying different categories. In this situation, the matrix corresponds to a model predicting leaf diseases with the following categories: Yellow Curf Leaf, Target Leaf, Spider Mite Leaf, Septoria Leaf, Leaf Mold, Late and Early Blight and Healthy. The numbers provided indicate the number of samples in each session. For example, Yellow Curf Leaf has 56 samples, Target Leaf has 57, and so on, with the Healthy class also having 56 samples. For each cell in the matrices, indicate how many samples from the true class (rows) were predicted as each predicted class (columns). Diagonal cells indicate correct predictions, while off-diagonal cells represent misclassifications.

Figure 7 presents the performance metrics of classification methods for leaf of tomato disease identification at various threshold percentages. The metrics include Sensitivity, Specificity, and Accuracy. Sensitivity refers to the capability to correctly identify diseased leaves (True Positives). As the threshold increases from 85% to 98%, sensitivity decreases, indicating that the model becomes less sensitive to detecting disease as the threshold rises. The model's ability to correctly detect healthy leaves

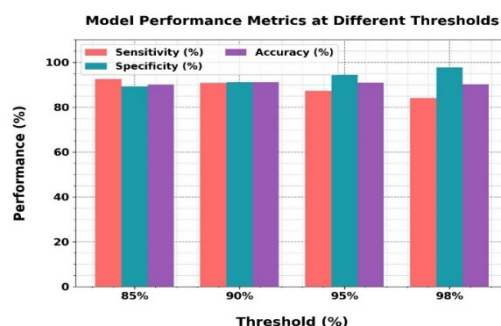


Fig 7. Performance Metrics of the Suggested Method

(True Negatives) is called specificity. Accuracy is the overall correctness, while specificity increases with a higher threshold, indicating the model's growing ability to detect healthy leaves correctly. At around 90%, it stays relatively constant, dipping a little at higher levels. The graph often illustrates a trade-off between specificity and sensitivity. Although raising the threshold increases specificity, it decreases sensitivity, and accuracy remains largely constant.

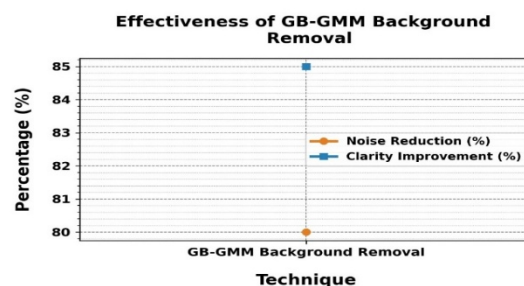


Fig 8. Proposed Method of GB-GMM

The GB-GMM Background Removal approach produces an average noise reduction of 80% and an increase in clarity of 85%, as seen in Figure 8, which depicts the methods used in leaf disease diagnosis. By eliminating extraneous background information and enhancing the visibility of essential elements, such as disease spots on leaves, these measurements demonstrate how effectively this strategy improves picture quality. Practically speaking, noise reduction is the process of removing extraneous information that might mislead the identification algorithm, such as dirt, shadows, or other plant parts. An 80% decrease indicates that the majority of distracting factors have been effectively eliminated. In the meantime, the 85% increase in clarity implies that the key sections of the picture, such as the sick areas, become more transparent and more straightforward for both ML algorithms and people to understand. Because of its capabilities, GB-GMM is especially well-suited for preprocessing images in automated agricultural systems, thereby improving the reliability and diagnostic accuracy of tomato disease diagnosis. Better training data for machine learning models is further supported by its high clarity improvement ratio, which eventually improves system performance.

The four segmentation techniques, Proposed Method, U-Net, SegNet, and FCN, across six performance indicators are shown in Figure 9. The Proposed Method demonstrates strong overall efficacy and balanced forecasts, as evidenced by its superior performance in accuracy (94.3%), precision (93.7%), recall (92.5%), and F1-score (93.1%). Additionally, its overfitting rate is the lowest (10%), indicating high generalization. Despite having the shortest training period (4 hours), U-Net has the most overfitting (35%) and the lowest accuracy (78.9%), indicating mediocre model reliability. With modest scores and 28% overfitting, SegNet outperforms the U-Net's measures but still falls short of the Suggested Way. Despite requiring the most extended training period (7 hours), FCN provides a solid balance, outperforming SegNet and U-Net in terms of accuracy (88.2%, F1-score 85.9%) and overfitting (22%). Overall, the graph shows that the Proposed Method is the most successful of the four models, offering the best trade-off among accuracy, productivity, and generalization.

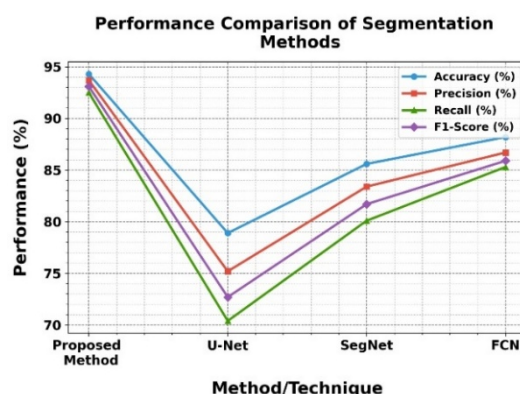


Fig 9. Comparison of Proposed Vs Existing Methods

4 Conclusion

Research on the early detection and management of TLD with DL has demonstrated tremendous potential to improve agricultural productivity and health. The system accurately detected and classified common illnesses, including early blight, late blight, and leaf mold, using a CNN trained on a variety of tomato leaf image datasets. This automated system provides a dependable and expedient alternative to conventional manual inspection techniques, which are often time-consuming, labour-intensive, and prone to error. With an accuracy of 94.3%, precision of 93.7%, recall of 92.5%, and an F1-score of 93.1%, the findings demonstrate that the proposed technique, implemented using Python, outperformed others, showing good overall performance and well-balanced forecasts in VGGNet architectures. Furthermore, the system is scalable across mobile platforms, IoT-integrated agricultural systems, and innovative farming applications, thanks to its ability to adapt to real-world conditions. The research also highlights the need for early disease diagnosis to minimize crop damage, maximize pesticide use, and reduce financial losses. This approach can be applied to many plants and illnesses as a foundation for widespread agricultural innovation, even though they now being used to deal with tomatoes. In the future, the model may be improved to withstand various weather conditions and integrated with edge computing devices for real-time field use. In conclusion, deep learning is very potent and revolutionary for contemporary agriculture, with broad uses in food security and ecological farming.

5 Acknowledgements

The research was not funded by any external funding agency.

References

- Sharma J, Al-Huqail AA, Almogren A, et al. Deep learning-based ensemble model for accurate tomato leaf disease classification by leveraging ResNet50 and MobileNetV2 architectures. *Scientific Reports*. 2025;15:13904. <https://doi.org/10.1038/s41598-025-98015-x>.
- Bouacida I, Farou B, Djakhadjakha L, Seridi H, Kurulay M. Innovative deep learning approach for cross-crop plant disease detection: A generalized method for identifying unhealthy leaves. *Information Processing in Agriculture*. 2025;12(1):54–67. <https://doi.org/10.1016/j.inpa.2024.03.002>.
- Joshi K, Hooda S, Sharma A, Sonah H, Deshmukh R, Tuteja N, et al. Precision diagnosis of tomato diseases for sustainable agriculture through deep learning approach with hybrid data augmentation. *Current Plant Biology*. 2025;41:100437. <https://doi.org/10.1016/j.cpb.2025.100437>.
- Patel D, Saxena AK. Feature Selection in High Dimension Datasets using Incremental Feature Clustering. *Indian Journal of Science and Technology*. 2024;17(32):3318–3326. <https://doi.org/10.17485/IJST/v17i32.2077>.
- Assaduzzaman M, Bishshash P, Nirob MAS, Marouf AA, Rokne JG, Alhaji R. XSE-TomatoNet: An explainable AI-based tomato leaf disease classification method using EfficientNetB0 with squeeze-and-excitation blocks and multi-scale feature fusion. *MethodsX*. 2025;14:103159. <https://doi.org/10.1016/j.mex.2025.103159>.
- Gunasekaran H, Rajkumar S, Kirubhadharsini BL. Lightweight deep learning for tomato disease detection: trends, challenges, and edge AI perspectives. *Frontiers in Plant Science*. 2026;16:1737208. <https://doi.org/10.3389/fpls.2025.1737208>.
- Dolatabadian A, Neik TX, Danilevicz MF, Upadhyaya SR, Batley J, Edwards D. Image-based crop disease detection using machine learning. *Plant Pathology*. 2025;74(1):18–38. <https://doi.org/10.1111/ppa.14006>.
- Das A, Pathan F, Jim JR, Ouishy MR, Kabir MM, Mridha MF. XLTLDisNet: A Novel and Lightweight Approach to Identify Tomato Leaf Diseases with Transparency. *Heliyon*. 2025. <https://doi.org/10.1016/j.heliyon.2025.e42575>.
- Sharma R, Singh A, Jhanjhi NZ, Masud M, Jaha ES, Verma S. Plant Disease Diagnosis and Image Classification Using Deep Learning. *Computers, Materials & Continua*. 2025;71(2). <https://doi.org/10.32604/cmc.2022.020017>.
- Imtiaz A, Swapnil FBI, Masud SR, Karmaker D. Tomato leaf dataset: A dataset for multiclass disease detection and classification. *Data in Brief*. 2025;60:111520. <https://doi.org/10.1016/j.dib.2025.111520>.

- 11) Prasad R, Saxena A. A Machine Learning Approach for an Early Prediction of Parkinson's disease. *Indian Journal of Science and Technology*. 2024;17:3410–3418. <https://doi.org/10.17485/IJST/v17i33.2091>.
- 12) Yildiz MB, Hafif MF, Koksoy EK, Kurşun R. Classification of diseases in tomato leaves using deep learning methods. *Intelligent Methods in Engineering Sciences*. 2024;3(1):22–36. <https://doi.org/10.1177/00368504241275371>.
- 13) Tolba A, Mostafa NN, Ali Y. Innovative Approach for Early Detection and Diagnosis of Tomato Leaf Diseases. *Optimization in Agriculture*. 2024;1:40–55. <https://doi.org/10.61356/j.oia.2024.1197>.
- 14) Gupta P. A Comprehensive Review on Tomato Leaf Disease Detection using Deep Learning. *TEJAS Journal of Technologies and Humanitarian Science*. 2025;4:33–39. <https://doi.org/10.63920/tjths.43003>.
- 15) Khan R, Din NU, Zaman A, Huang B. Automated Tomato Leaf Disease Detection Using Image Processing: An SVM-Based Approach with GLCM and SIFT Features. *Journal of Engineering*. 2024;2024(1):9918296. <https://doi.org/10.1155/2024/9918296>.
- 16) Ghosh H, Rahat IS, Emon MMR, Mashrafi MJ, Tanzin MAAA, Mohanty SN, et al. Advanced neural network architectures for tomato leaf disease diagnosis in precision agriculture. *Discover Sustainability*. 2025;6(1):312. <https://doi.org/10.1007/s43621-025-01149-1>.
- 17) Arafat IS, Aswath S, Nishath SH. Automatic Early Detection of Tomato Leaf Disease using IoT and Deep Learning. 2025. <https://doi.org/10.21203/rs.3.rs-3958296/v1>.
- 18) Bahaa M, Hesham A, Ashraf F, Abdel-Hamid L. A Smart AIoT-Based Mobile Application for Plant Disease Detection and Environment Management in Small-Scale Farms Using MobileViT. *AgriEngineering*. 2026;8. <https://doi.org/10.3390/agriengineering8010011>.
- 19) Lestari CAD, Anam S, Sa'adah U. Tomato Leaf Disease Classification with Optimized Hyperparameter: A DenseNet-PSO Approach. Atlantis Press. 2024. https://doi.org/10.2991/978-94-6463-413-6_23.
- 20) Patwal P, Chauhan R, Bhatt C, Devliyal S. Automated tomato disease detection and classification using image processing and machine learning for precision agriculture. CRC Press. 2025. <https://doi.org/10.1201/9781003559085-84>.
- 21) Ramu J, Darshini NKP, Senthilkumar G, Sathyakala S, Mahendran G, Maranan R. Machine learning approaches for early detection and prevention of late blight and early blight diseases in potato and tomato plants. CRC Press. 2025. <https://doi.org/10.1201/9781003559115-12>.