Smart Internz

# TOMATO DISEASE DETECTION USING MACHINE LEARNING



## TEAM MEMBERS

1. Ullamgunta Venkata Sri Gayatri(Team Leader)

2.Shaik Rufiya

3.Pandi Akanksha

4.Maddukuri Priyanka

# TOMATO LEAF DISEASE DETECTION USING CNN

# INTRODUCTION

## 1.1 OVERVIEW

The detection of diseases in tomato plants using Convolutional Neural Networks (CNN) is a significant advancement in agricultural technology. Here's an overview:

**Convolutional Neural Networks (CNN)** are a class of deep learning algorithms that are particularly effective for image recognition and classification tasks. In the context of tomato disease detection, CNNs can analyze images of tomato leaves to identify patterns and characteristics indicative of various diseases.

The process typically involves several steps:

1. **Data Collection**: Gathering a large dataset of tomato leaf images, including healthy and diseased samples.
2. **Preprocessing**: Enhancing image quality and applying techniques like data augmentation to increase the diversity of the dataset.
3. **Model Training**: Using the dataset to train a CNN model, which involves adjusting the model's parameters to minimize the error in disease classification.
4. **Evaluation**: Testing the model's performance on a separate set of images to assess its accuracy and reliability.

Recent studies have shown that CNN-based models can achieve high accuracy in detecting and classifying tomato diseases.

The use of CNN for tomato disease detection not only helps in early diagnosis but also supports farmers in taking timely action to prevent the spread of diseases, ultimately leading to better crop management and yield.

## 1.2 ABSTRACT

Global food production is being strained by extreme weather conditions, fluctuating temperatures, and geopolitics. Tomato is a staple agricultural product with tens of millions of tons produced every year worldwide. Thus, preserving the tomato plant from diseases will go a long way in reducing economical loss and boost output.

Technological innovations have great potential in facilitating disease detection and control. More specifically, artificial intelligence algorithms in the form of deep learning methods have established themselves in many real-life applications in a wide range of disciplines (e.g., medicine, agriculture, or facial recognition, etc.). In this paper, we aim at applying deep transfer learning in the classification of nine tomato diseases (i.e., bacterial spot, early blight, late blight, leaf mold, mosaic virus, septoria leaf spot, spider mites, target spot, and yellow leaf curl virus) in addition to the healthy state.

The approach in this work uses leaf images as input, which is fed to convolutional neural network models. No preprocessing, feature extraction, or image processing is required. Moreover, the models are based on transfer learning of well-established deep learning networks. The performance was extensively evaluated using multiple strategies for data split and a number of metrics. In addition, the experiments were repeated 10 times to account for randomness.

The ten categories were classified with mean values of 99.3% precision, 99.2% F1 score, 99.1% recall, and 99.4% accuracy. Such results show that it is highly feasible to develop smartphone-based applications that can aid

plant pathologists and farmers to quickly and accurately perform disease detection and subsequent control.

# LITERATURE SURVEY

## 2.1 EXISTING PROBLEM

Traditional methods of tomato disease detection often rely on manual inspection by agricultural experts, which can be time-consuming, labor-intensive, and prone to human error. Additionally, these methods may not provide timely detection, leading to delays in disease management and potential crop losses.

## 2.2 PROPOSED SOLUTION

Implementing an automated tomato disease detection system using deep learning techniques, specifically Convolutional Neural Networks (CNNs), offers a promising solution. By leveraging CNNs, the system can analyze images of tomato leaves and accurately identify various diseases, including early blight, late blight, bacterial spot, etc.

This approach enables rapid and objective disease detection, allowing for timely intervention and efficient management of plant health. Additionally, the system can be deployed in the form of a mobile app or web service, providing farmers with easy access to on-the-spot disease diagnosis and recommendations for treatment.

This solution not only improves the accuracy and speed of disease detection but also enhances overall crop productivity and reduces reliance on manual inspection methods.

# 3. METHODOLOGY

Figure 1 shows a diagram of all phases involved in the proposed approach. By using CNNs, performing explicit feature extraction is not required. Furthermore, there is no need for separating relevant image parts (i.e., segmentation). These steps and others are handled implicitly by the complex operations of the deep learning models. Given a generically pre-trained deep learning model, several changes need to be made to re-purpose the model to the specific application.

First, replace the classification layer to match the number of classes in the application (i.e., 10 classes for this paper).

Second, replace the learnable layer that combine features from previous layers with a new layer. This may be a fully connected layer or a convolution2d layer depending on the CNN model.

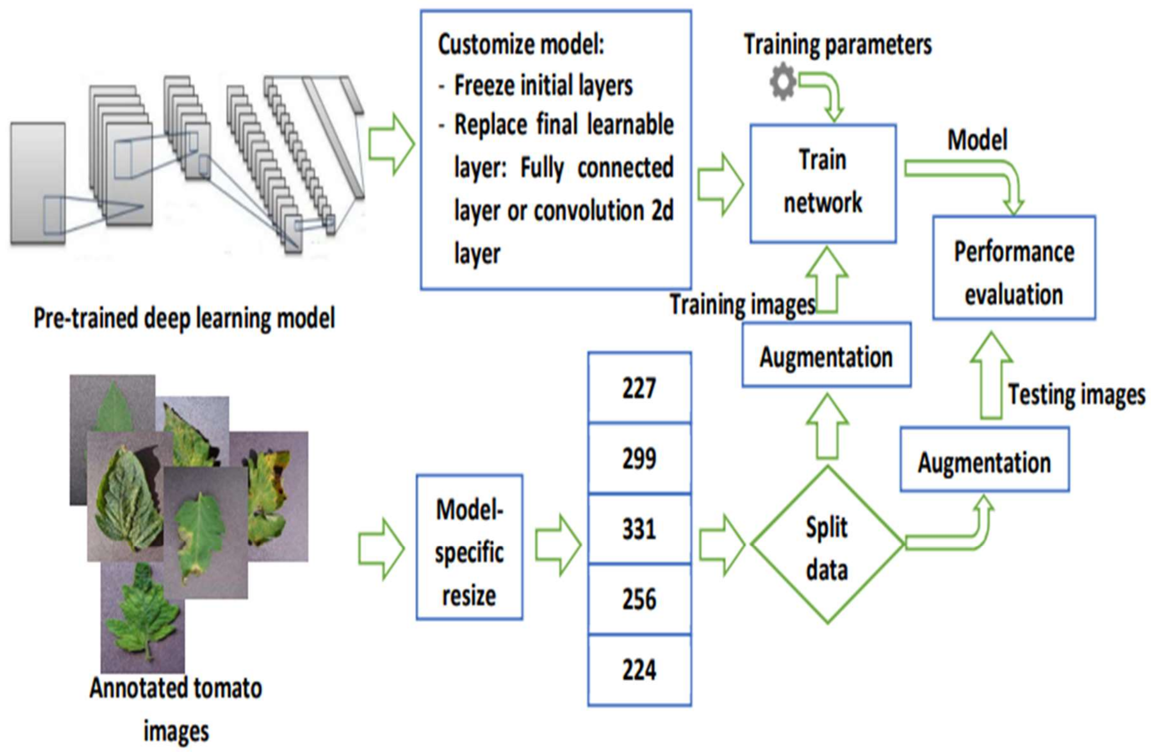



Figure 1. A diagram of all phases of the proposed approach.

Third, if training is to be made faster, then some initial layers can be frozen (i.e., layer weights will not be updated during training). The number of frozen layers can be determined empirically depending on the application and the resulting testing performance and training speed. No layers were frozen in this work as the available hardware permitted extensive training.

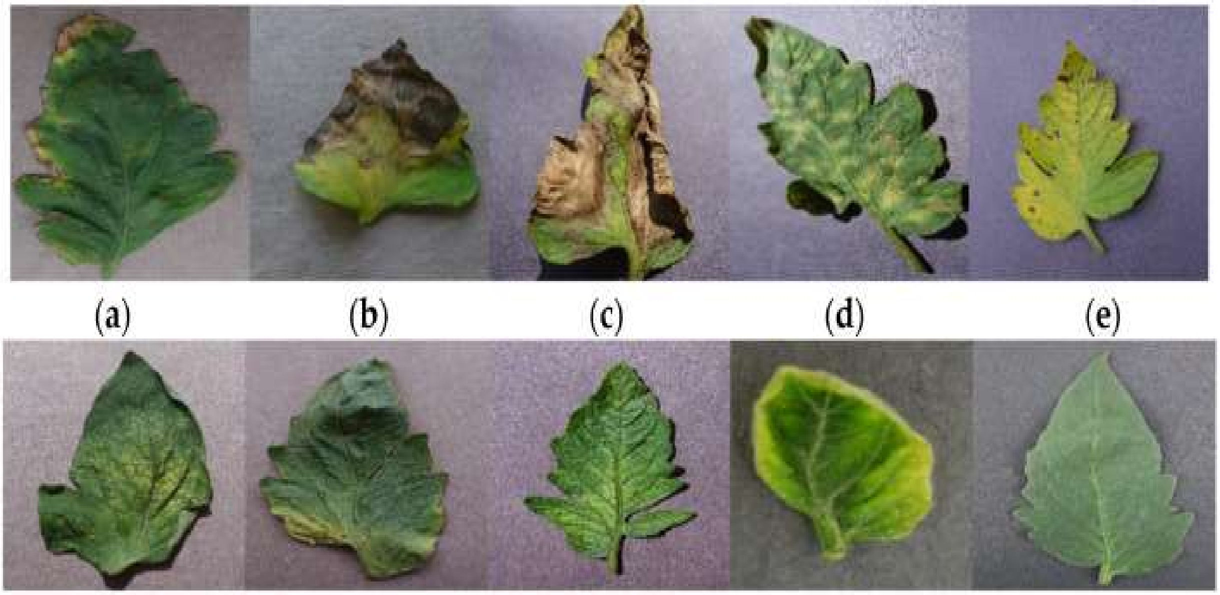
Fourth, the dataset needs to be prepared by resizing the images to fit the CNN requirements (e.g., 256 × 256 to 224 × 224). Furthermore, the data are split into training and validation subsets. In addition, image augmentation operations may be performed to introduce more variety into the dataset and improve the learning process.

Fifth, in this final step, the CNN network is retrained with the tomato dataset, and the performance is evaluated.

## 3.1 DATA SET

The dataset consists of 18,160 publicly available tomato leaf images displaying features of nine tomato diseases in addition to the healthy state. The number of images per class was as follows: 2127 bacterial spot, 1000 early blight, 1909 late blight, 952 leaf mold, 373 mosaic virus, 1771 septoria leaf spot, 1676 spider mites, 1404 target spot, 5357 yellow leaf curl virus, and 1591 healthy [23]. Each image represents a photo of a single leaf exhibiting one of the ten health classes. The photos were taken using a neutral background that appears somewhat unified for all images. In addition, each leaf appears at the center of each image.

Although the images may contain irrelevant margins displaying the background, no cropping or pre-processing were performed. The public source of the images provided the dataset in JPEG format and a 256 × 256 resolution. Samples of leaf images of the nine diseases and healthy leaves are shown in Figure 2



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

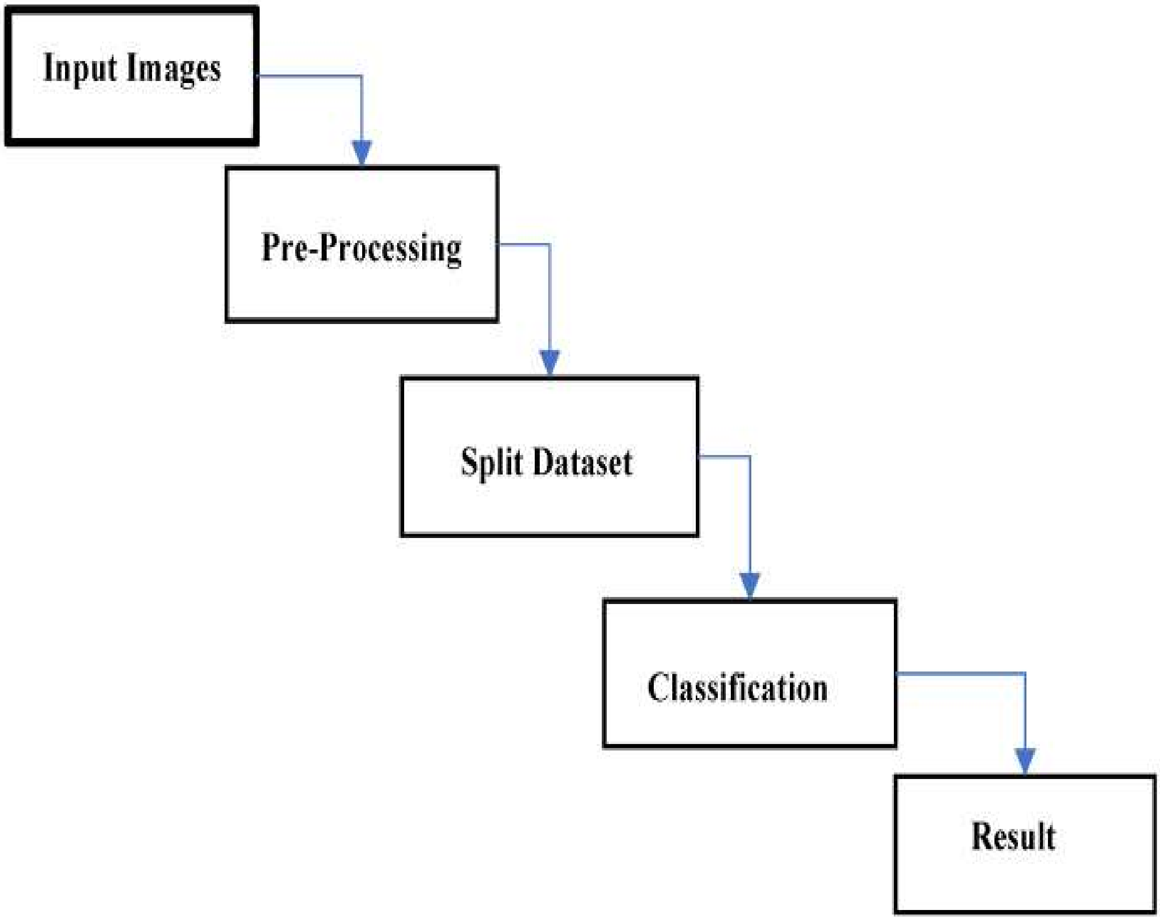
## 3.2 IMAGE PRE-PROCESSING AND LABELLING

Before training the model, image pre-processing was used to change or boost the raw images that needed to be processed by the CNN classifier. Building a successful model requires analyzing both the design of the network and the format of input data. We pre-processed our dataset so that the proposed model could take the appropriate features out of the image.

The first step was to normalize the size of the picture and resize it to 256 × 256 pixels. The images were then transformed into grey.

This stage of pre-processing means that a considerable amount of training data are required for the explicit learning of the training data features. The next step was to group tomato leaf pictures by type, then mark all images with the correct acronym for the disease. In this case, the dataset showed ten classes in test collection and training.

## 3.3 TRAINING DATASET

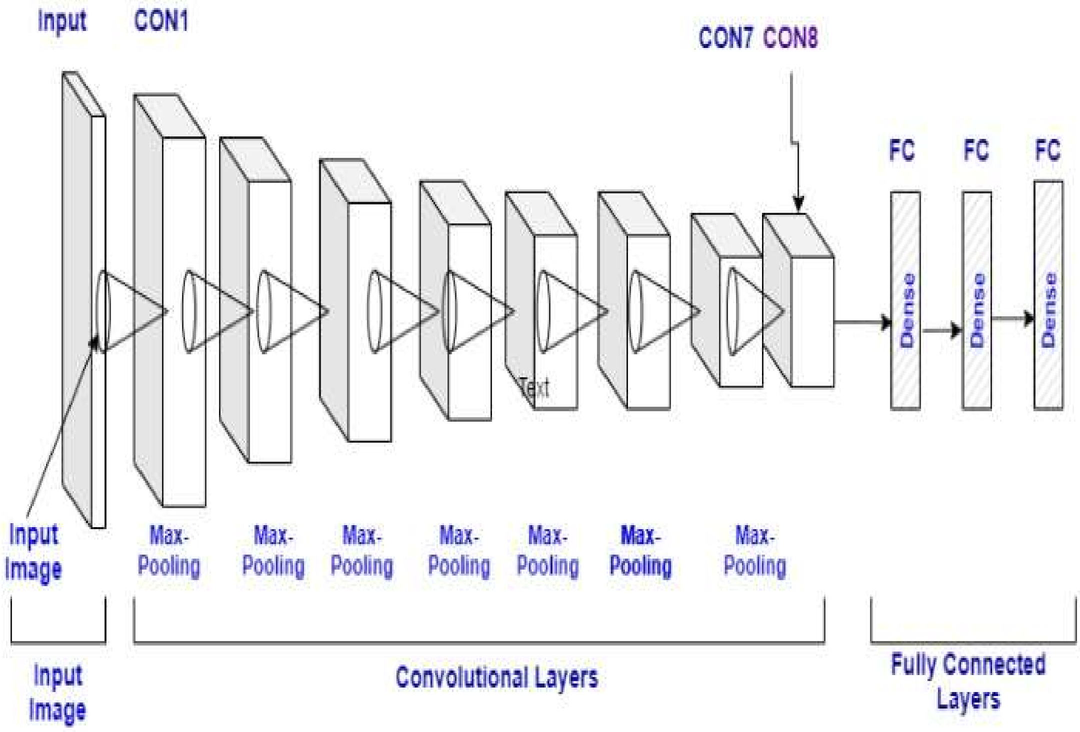



## 3.4 CONVOLUTIONAL NEURAL NETWORK

The CNN is a neural network technology widely employed today to process or train the data in images. The matrix format of the Convolution is designed to filter the pictures.

For data training, each layer is utilized in the Convolution Neural Network, including the following layers: input layer, convo layer, fully connected layer pooling layer, drop-out layer to build CNN, and ultimately linked dataset classification layer.

It can map a series of calculations to the input test set in each layer. The complete architecture is shown in Figure 3, and a description of the model is in Table 1.




Moreover, we ran the proposed model for 1000 epochs and validation and training accuracy is presented in figure 4.

For the calculation of loss, the categorical cross entropy method has been applied. The formula for calculating it represented in following equation 1. loss = − M c=1 log(po,c) (1) where M - number of classes, y - binary indicator (0 or 1) if class label c is the correct classification for observation o and predicted probability observation o is of class c.

After analyzing the performance of proposed model, we undergo for the testing. For the testing purpose, total 500 sample has been used and for the various classes testing accuracy is different and it is ranging from 76% to 100% and the average accuracy of the proposed model is 91.2%

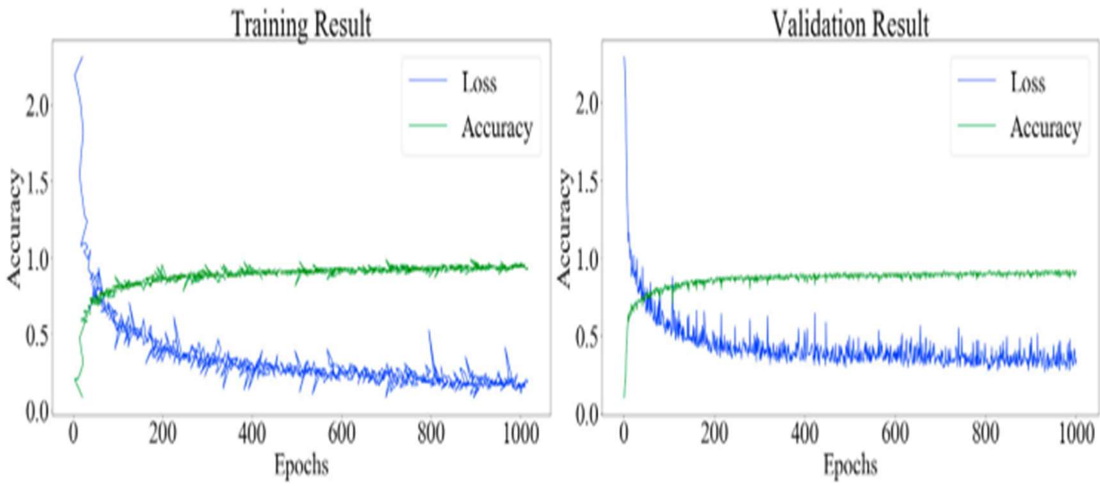



Fig. 4. (a) Training loss (b) Validation loss of proposed model.

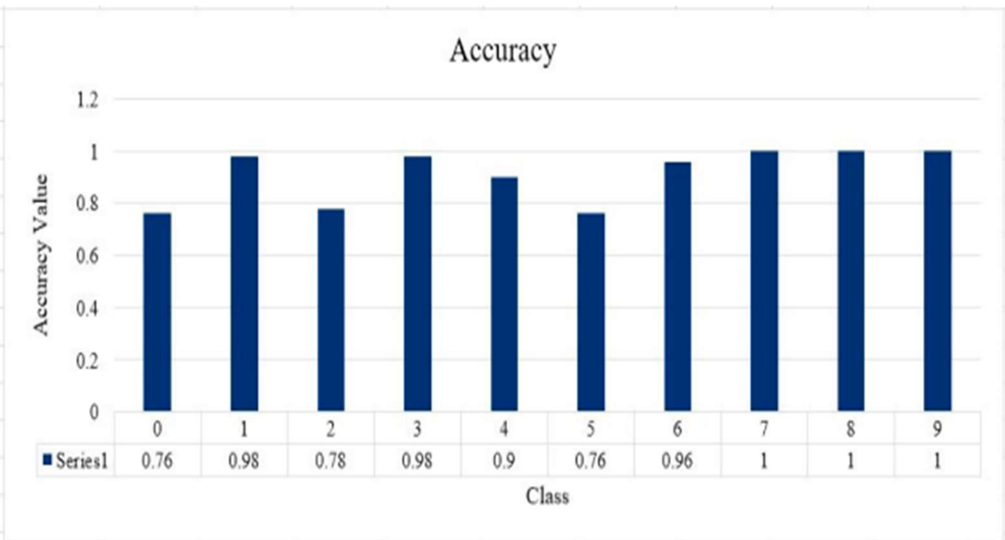



Fig. 5. Testing accuracy for the 10 classes.

# 4. ADVANTAGES

1. **Accuracy:** CNN-based disease detection systems can achieve high accuracy in identifying various tomato diseases, potentially outperforming human experts in consistency and precision.
2. **Speed:** Automated detection systems can analyze large volumes of images rapidly, providing real-time or near-real-time feedback on the health status of tomato plants.
3. **Cost-effective:** While initial development costs may exist, once deployed, automated systems can reduce the need for manual labor and expertise, ultimately saving costs over time.
4. **Timeliness:** Early detection of diseases allows for timely intervention, preventing the spread of diseases and minimizing crop losses.
5. **Scalability:** The system can be scaled to analyze large agricultural areas, offering broad coverage and support for farmers managing extensive tomato plantations.

# 5. DISADVANTAGES

**1. Data dependency:** The performance of CNN models heavily relies on the quality and diversity of the dataset used for training.

**2.Model interpretation:** While CNNs are effective at classification tasks, they often lack interpretability, making it difficult to understand how the model arrives at its decisions.

**3.Dependency on environmental conditions:** The performance of the detection system may vary based on factors like lighting conditions, camera quality, and variations in leaf appearance due to environmental factors.

**4.Deployment challenges:** Implementing and maintaining automated detection systems in agricultural settings may require infrastructure, such as reliable internet connectivity and power supply, which could be lacking in certain regions.

**5.Continual improvement:** Regular updates and retraining of the model are necessary to keep pace with evolving disease patterns and ensure continued accuracy. This process requires ongoing resources and effort.

## 6.CONCLUSION

Tomato leaf disease detection and categorization employ a variety of deep learning approaches. Compared to other transfer learning techniques such as ResNet152, VGG19, and InceptionV3, the CNN model performed well in terms of disease detection in tomato crops. The proposed model has a training accuracy of 98% and a testing accuracy of 88.17%.

Farmers can actually overcome their plant identification difficulties without the need to follow plant scientists. This will help them to cure tomato plant diseases in time, so they can improve the quality and quantity of their tomato crops, as well as profit. In the future, we hope to improve the model with a different crop. Additionally, we will try to optimize the same model on the same dataset to further improve the test accuracy.

## 7.FUTURE SCOPE

The future scope of tomato disease detection using CNNs and other advanced technologies holds significant potential for further advancements and applications:

**Multi-Spectral Imaging:** Incorporating multi-spectral or hyperspectral imaging techniques can provide additional information beyond visible light, enabling more comprehensive disease detection and characterization based on subtle spectral differences.

**Cloud-Based Solutions:** Cloud-based platforms can provide scalable infrastructure for hosting and deploying disease detection models, enabling widespread adoption and accessibility for farmers around the world.

**AI-driven Decision Support Systems:** Integration of artificial intelligence (AI) into decision support systems can provide personalized recommendations for disease management strategies based on specific environmental conditions, crop varieties, and disease prevalence.

**Disease Forecasting:** Utilizing historical data and predictive modeling techniques, it may be possible to develop disease forecasting systems that anticipate disease outbreaks and help farmers implement preventative measures in advance.

## 8.RESULT

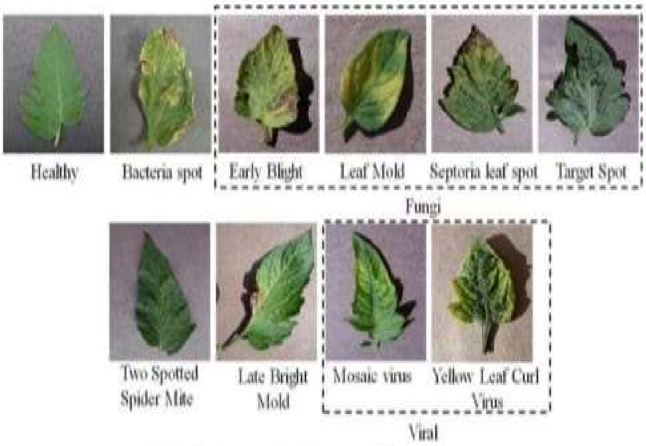
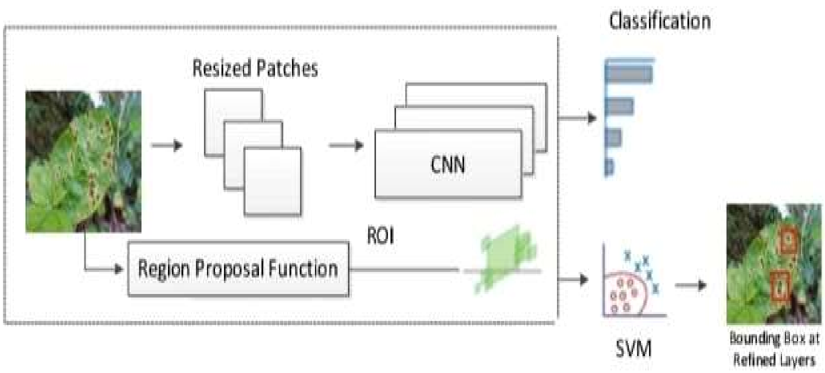




Fig. 2. Tomato Leaf Images with its Diseases.




Leaf Disease Detection
Choose File
No file chosen
Predict Leaf Disease
Original Image
Predicted Disease: Potato__Late_blight


Leaf Disease Detection
Choose File
No file chosen
Predict Leaf Disease
Original Image
Predicted Disease: Tomato__Septoria_leaf_spot