A Deep Learning based Phyto-pathological Detection for Early Diagnosis of Tomato Leaf Diseases

Fariha Ashraf[1], Fahima Tahir[1], Rabia Javed[1], Wajeeha Malik[1], Khansa Aatif[1]

[1] Department of Computer Science, Lahore College for Women University, Jail Road, Lahore 54000, Pakistan.

Abstract

Tomato (Lycopersicon esculentum Mill) is a frequently grown vegetable in various regions of the world. Tomatoes are one of the most demanded vegetables and a source of revenue for third-world countries. Tomato plant diseases can significantly reduce tomato production and impact a nation's economic development. Timely and reliable detection of tomato leaf diseases can increase tomato production, ensure food security, and sustain agricultural activities. A Convolutional Neural Network (CNN) model was employed in this study to detect and classify Fungal, Bacterial, and Viral diseases in nine types of tomato leaves. High-resolution images of plant leaves damaged by pathogens were used for training and testing the CNN model. The dataset used in Experiment 1 consisted of nine different tomato leaf classes, including one healthy leaf class and eight pathogen-affected classes. The dataset used for Experiment 2 was categorized into three classes based on the pathogens: bacteria, fungi, and viruses. Experiment 1 achieved an accuracy of 93%, and Experiment 2 achieved 96% accuracy in detecting tomato leaf diseases. The results demonstrate that the deep learning-based Convolutional Neural Network (CNN) model is an advanced and effective solution for the identification of plant leaf diseases. This research will support better crop planning and help increase overall tomato production.

Keywords: Tomato crop, Plant pathology, Deep learning, Conventional neural network (CNN), Plant disease detection, Automatic detection, Plant-pathology, Tomato leaf disease.

Introduction

Agriculture is important in the global economy due to the faster growth in the number of people and their need for food(Leach et al., 2014). Environmental issues, productivity levels, sustainable efforts, and food security are the primary challenges in the agricultural system. If a disease is not discovered early and accurately, it will be difficult to carry out prompt actions against it (Abdulridha et al., 2019). Tomatoes are the world's most consumed vegetable and an essential ingredient in numerous foods in raw or processed form (C. Wang et al., 2023). Because of its excessive global usage/ demand, Tomato is the second largest cultivated crop. And the most processed crop in the world (Sattar et al., 2024). Unfortunately, this highly demanded crop is the one that affects the maximum number of plant diseases. The tomato crop is vulnerable to almost 200 diseases during various stages of cultivation and harvesting. In the past few years, people have been eating more tomatoes each year by 3% (Caruso et al., 2022). During the past 40 years, the world's tomato crop area has expanded by 164% and the world's consumption of tomatoes has gone up by 314%. In 2009, people harvested over 154 million tons of tomatoes, and the amount increased to over 174 million tons in 2015 (Qasim et al., 2018). In the current year, 170.7 million tons of tomatoes have been produced on 5 million hectares(A. K. Singh et al., 2018). Native to Central and South America, tomatoes are eaten more frequently in the USA and Europe than in most other nations. The current need is about 100 million tons of new tomatoes in today's world. With normal temperatures of 18–25 °C and 10–25 °C in the evening, farmers usually harvest the tomatoes within 90–150 days. Pakistan is mainly an agrarian country. Most of its population lives in rural areas, and they earn through farming. The agrarian industry contributes significantly to Pakistan's economy, accounting for 18.9% of the gross domestic product and employing 42.3% of the total workforce (M. Khan et al., 2024). The country not only consumes a huge number of tomatoes every year but also exports the tomatoes to Many Asian countries. "Pakistan exported USD 0.8 million worth of ketchup to various countries in 2018 (M. Khan et al., 2024)." Tomato plant disease can significantly reduce tomato production and lead to substantial economic deficit (Ally et al., 2023).

The increased demand for tomatoes on the global market has led many countries, mostly the nearby ones, to increase their farmland and exports. The causes of plant diseases can be either living, such as fungi, bacteria, viruses, nematodes, phytoplasma, protozoa, and parasitic plants; or non-living, such as pollution or conditions in the environment. Many natural viruses and the rise of new pathogens caused a constant setback for our grain, trees, fiber, and plants in rural areas. A major issue in the field of farming occurs when plant diseases badly influence a nation's financial growth. When such diseases can be automatically identified, the effects may be reduced (Hassanien et al., 2017). The low productivity of farms results from plant diseases, disorders, or damage. As leaves are the main part of crops that are damaged, farmers put effort into discovering what causes the diseases that harm them. In most cases, plant diseases cause damage primarily to the plant leaves. Most diseases are triggered by pathogens, which refers to the area of Phytopathology. In Phytopathology, plant disorders and diseases are classified into three main factors: Biotic factors (Due to fungi, algae, bacteria, etc.), Abiotic factors (Due to environmental conditions like temperature, weather, rainfall, nutrient deficiency, humidity, etc.), and Mesobiotic factors (Due to both Biotic and Abiotic factors like viruses). The main challenge caused by biotic factors is that bacteria, fungi, animals, or viruses may impact how crops are grown and their final cost. Among the regular signs of plant diseases are leaf rust, stem rust, sclerotinia, powdery mildew, birds-eye spot on berries, seedling damping, and leaf spot. The state of the leaves reveals whether the plant has any illness. In the past, efforts were made to evaluate machine learning, and now a classifier is used to sort photos as healthy or ill. Usually, the leaves show the signs and are the area primarily examined to identify what's wrong with a plant (Venkataramana et al., 2022). Early and accurate detection of these diseases is critical for effective management and control, ensuring food security and sustainability in agricultural practices (V. K. Singh et al., 2018). Most developing country farmers depend on traditional farming, which is both physically demanding and requires much more time. Agricultural damage from diseases can be detected and identified with machine learning(Jayswal & Chaudhari, 2023).

In older days, plant diseases instigated by pathogens were identified only by visual inspection. Farmers take their affected vegetable leaves to long-distance agrarian centers where agrarian specialists identify leaf diseases by seeing them, but many plant diseases don't have any visual effect on the leaves, and it is difficult for farmers to detect plant diseases. Image devices can identify symptoms of any type of disease as they cover the entire spectrum of electromagnetic rays, ranging from gamma to radio waves. These are not visible to human eye (M. Khan et al., 2024). Deep learning methods are versatile, capable of handling the most varied types of data (Ribeiro et al., 2023). Machine learning, deep learning, and reinforcement learning techniques play significant roles in many scientific fields. Many researchers are using deep learning methods to detect and classify plant leaf disease. Deep learning methods are an effective solution to scientific problems (Aatif et al., 2024). Image processing through AI modelling is a cost-effective, accurate, fast, and automated method of plant disease detection. The use of AI-based image processing methods is an unharmful approach for plant pathogens detection, a less costly and reliable solution to the problem [6]. Early detection and reliable sorting of plant diseases depend on using image processing techniques. Collected images from around several plant samples or locations are used to train effective models that classify plant diseases. Within this study, considering all the numerous research papers that describe and address the main problems, disease threats, and strategies are as follows: Image quality and standard of plant leaves, background details and sounds are caused by the collected images, plant leaf color, size and shape differ as temperature, or climate change, vast volumes of data collection ought to be addressed, divide the exact location on a leaf into relevant disorder, and identifying significant diseases through segmented location.

Early and correct diagnosis of plant leaf diseases is an essential requirement of third-world countries like Pakistan, where the economy heavily depends on agrarian income. Lack of accurate and timely diagnosis is one of the major reasons for extensive damage to agricultural countries' economies. This research presents a non-invasive, reliable, and cost-saving method for the diagnosis of tomato plant leaf diseases. This research aims to detect tomato leaf diseases from RGB images by using a Deep Learning method. This will help increase the healthy tomato production and achieve high revenue goals.

The main contributions of this research methodology include:

- The Plant Village dataset is used to capture high-resolution images of different kinds of plant leaves that are damaged by pathological causes under different factors.
- Improved the CNN-based method to detect and classify plant illnesses such as rust, viruses, bacteria, fungus, nutrient deficiencies, and atmospheric conditions.
- The detection of tomato leaf diseases caused by three pathogens: fungi, bacteria, and viruses using deep learning.
- Proposed a benchmark solution for nine tomato diseases with transfer learning capabilities, offering an adaptable framework for regional adaptation and agricultural extension.

LITERATURE SURVEY

Many image processing research fields have mainly been influenced by deep CNN models. The reason comes from the fact that CNN can automatically pull features from images by running through its stack of convolutional layers. Several CNN architectures have been compared for their effectiveness in detecting plant diseases on potato and mango leaves (Arya & Singh, 2019; Wang et al., 2017). Deep learning is an assemblage of machine learning techniques. Instead of relying on a fixed characteristic order, like many machine learning methods do, deep learning can do without one. Other approaches to plant disease work have brought impressive outcomes, but they take more time and cannot be easily modified (Kwabena et al., 2020). The proposed study explains that CNN was more effective than traditional machine learning in terms of accuracy. It appears that applying deep learning to detect plant diseases could be the best and most scalable option for better management of farmlands and gains in production, as the world needs more food (Ronihal et al., 2025). The proposed study explained how Mask R-CNN is used to detect plant diseases using photographs. Since so many diseases in plants are dangerous for agriculture, the authors chose Mask R-CNN, a computer vision system, to detect early signs of new problems and precisely mark and classify affected parts of leaves. The study proposed a detailed approach for finding apple leaf disease using deep convolutional neural networks. Producing such pathological objects and a complex AlexNet network that can recognize apple leaf diseases. The model was designed to spot the presence of four types of apple diseases using a total of 13,689 images. As a result of experiments, the proposed use of a convolutional neural network with leaf disease identification reaches 97.62 % accuracy, uses less memory than the regular Alex Net model and improves the accuracy of pathological image generation by 10 %. According to the research, the suggested deep learning model improves the precision and speed of disease management for apples, and the method of image generation would improve the accuracy of the model. Tomato yellow leaf curl virus (TYLCV) is the most damaging virus, found in many parts of the world with tomato yellow leaf curl disease (TYLCD). It also causes the tomato leaf to curl and turn yellow (B. Liu et al., 2017). The use of different computational methods together leads to better and smarter monitoring of plants in precision farming (Dhingra et al., 2019). The proposed study explains an automatic method that diagnoses citrus diseases based on their symptoms using color and texture. Images are prepared, color histogram and texture info from GLCM calculated, and lastly k-NN, SVM and Random Forest classifiers are used. The use of color and texture details in combination allows the model to distinguish different diseases that seem alike. It proves that AI-enabled systems that use images help to promptly find some common diseases in citrus (Ali et al., 2017). The proposed study illustrates tomato plant leaf diseases using KNN and PNN algorithms, so the disease detection system works better and more efficiently (K. & Rao, 1 C.E.). The proposed study examined a system for detecting and categorizing tomato leaf diseases with the help of CNNs (Tm et al., 2018). The proposed study used a sophisticated tomato detection system that was created using YOLOv3 (You Only Look Once version 3). The main goal was to improve the existing object detection methods used in real-time farming by handling situations where objects in the way and complex backgrounds are problems. YOLOv3 was selected because it detects objects quickly and accurately, both of which are important for smart farming. All in all, the research found that YOLOv3 in YOLO-Tomato is a strong tool for crop monitoring that helps advance AI-enabled modern farming (G. Liu et al., 2020).

Materials and Methods

The proposed method detects and classifies tomato leaf diseases brought on by different environmental factors using a deep learning method, as shown in Figure 1. The dataset is collected from the online, freely available website ‘Plant Village’. The tomato crop is the most common plant worldwide, so this work also examines tomato plant diseases. This work used data from 14233 Tomato leaves and their diseases among the 54306 total plant images in our library. In the system, a Sequential CNN is trained from scratch to identify diseases on tomato plant leaves. In this research, a CNN model is used to detect Fungal, Bacterial, and Viral diseases using nine different tomato leaf classes, including one Healthy class, five Fungal disease classes, one bacterial disease class, and two viral disease classes, based on two Experiments. Fungal and bacterial diseases are biotic factors, while viral diseases are due to Mesobiotic Factors (both Biotic and Abiotic). Sequence of images are labeled including 1: Bacterial spot disease, 2: Septoria leaf spot disease, 3: Early blight disease 4: Spider mites disease, 5: Healthy leaves, 6: Target spot disease, 7: Late blight disease, 8: Mosaic virus disease, 9: Leaf mold disease, 10: Yellow leaf curl virus disease. In Experiment 1, the dataset is classified into nine classes, from which eight are unhealthy and one belongs to healthy images of tomato leaves. In Experiment 2, the whole dataset is categorized into three classes based on their pathogens as either bacteria, fungi, or viruses. Results reveal that both Experiments had very good accuracy in both training and testing.

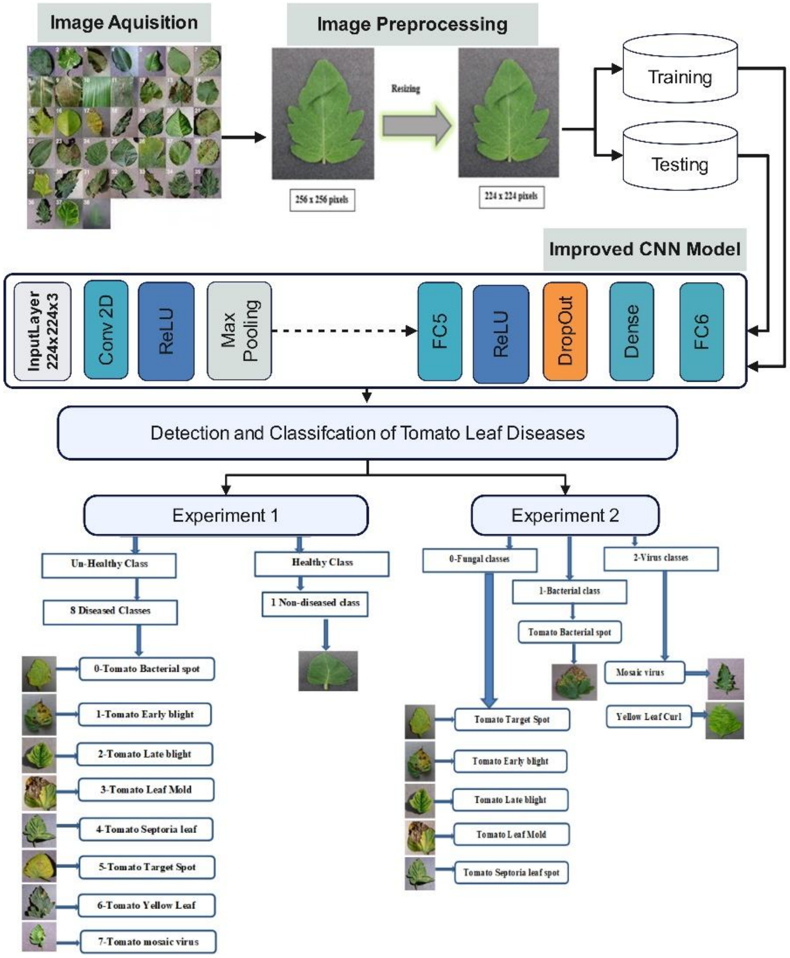


Figure 1: Proposed Research Methodology

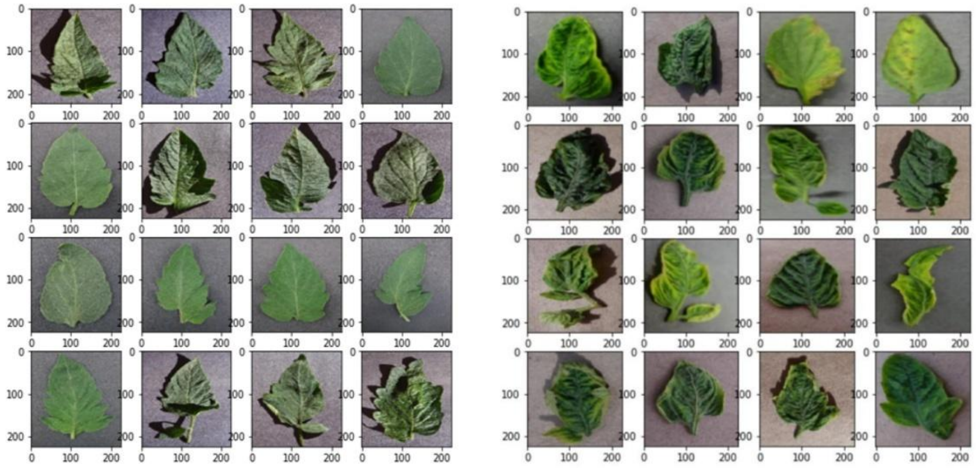
Image Acquisition

The PlantVillage dataset(Hughes & Salathé.; PlantVillage Dataset), a well-established benchmark in plant disease detection research, was used in this study. The dataset is carefully curated, and its class labels have been validated with the assistance of plant pathology experts, making it a reliable resource for training and evaluating deep learning models. Its standardized images provide a consistent foundation for benchmarking and comparing methods across studies. Building on this foundation, future research can extend the approach by incorporating real-time field images and further collaboration with plant pathologists to enhance adaptability and practical deployment in diverse agricultural environments.

This work is based on two types of experiments using the same dataset with different numbers of image divisions. In this research, ‘Plant Village Dataset’ is used as shown in Figure 2, which consists of 54306 healthy and unhealthy images of different leaves divided into 38 categories by 14 species and 26 diseases with an image size of 256 × 256 pixels. The tomato crop is used for this work, along with healthy and unhealthy images that have a total of 18160 images.

Figure 2: 38 Different Crops from Plant Village Dataset

The dataset is present in three different versions in the raw directory: color of Original RGB images, Gray-scaled images, and RGB images with color corrected and leaf segmented. This research analyzed only Tomato crops from the Plant Village dataset, containing ten different leaf classes. Across all experiments, used color version of the Plant Village dataset and the same dataset was used for both Experiments. 14233 images were set aside for this research. Two experiments are done as shown in figure 3, 7560 images are used for the first experiment, and 6673 are used for the second experiment.



Experiment 1 Dataset Images **Experiment 2** Dataset Images

Figure **Error! No text of specified style in document.**: Sample images from Plant Village Dataset of Experiments 1 and 2

This research used only nine classes as a dataset, out of which eight classes are unhealthy, and one is healthy, as described in Table 1.

Table 1:Datasets and Their Description for Experiments 1 and 2

Experiments	Dataset	Instances	Classes	Resource	Type/Format
Experiment 1	Tomato Crop	7200	8 Un-Healthy	PlantVillage	.JPG
	Tomato Crop	360	Healthy		
	Total	**7560**			
Experiment 2	Tomato Crop	4500	Fungal		
	Tomato Crop	900	Bacterial		
	Tomato Crop	1273	Viral		
	Total	**6673**			

Class Distribution for Model

The label sequence of images is used as 0-8 and belongs to 9 unhealthy and healthy classes for Experiment 1 and 0-2 for Experiment 2, which belong to Fungal, Bacterial, and Viral courses, including only unhealthy classes of images as shown in Table 2. The Dataset used in the model is divided into three parts: training, validation, and test evaluation, with 70% for training 15% for Testing, and 15% for Validation. The test dataset is used to provide an unbiased assessment of a final model that fits the training dataset. The validation dataset is used when modifying the hyperparameters for the procedure to include an unbiased model fit test on the research data collection. The analysis becomes more skewed when the expertise on the Test Dataset is considered. The testing dataset is introduced into the model configuration.

Table 2: Image distribution for training, testing, and validation tests in experiments 1 and 2

Experiments	Classes	Training Set	Testing Set	Validation Set
Experiment 1	Tomato Bacterial spot	630	135	135
	Tomato Early blight	630	135	135
	Tomato Late blight	630	135	135
	Tomato Leaf Mold	630	135	135
	Tomato Septoria leaf spot	630	135	135
	Tomato Target Spot	630	135	135
	Tomato Yellow Leaf Curl Virus	630	135	135
	Tomato mosaic virus	252	54	54
	Tomato healthy	630	135	135
	Total	**5292**	**1134**	**1134**
Experiment 2	Fungal	3150	675	675
	Bacterial	630	135	135
	Viral	895	189	189
	Total	**4675**	**999**	**99**

Image Preprocessing

Preprocessing applies to all the transformations on the raw data until they are fed into the machine learning algorithm or deep learning algorithm. Data pre-processing is an important phase in machine learning or deep learning because the accuracy of the data and the valuable knowledge that can be obtained from it directly influences the model's capacity to learn; thus, it is highly important to pre-process the given data before feeding it into the model. In both experiments, the images in the dataset are resized to 224 x 224 dimensions, which speeds up the training process to make the measurement of the model testing practical as presented in figure 3.

The method of either maximizing the input or goal variables allows for speeding up the movement of trains.

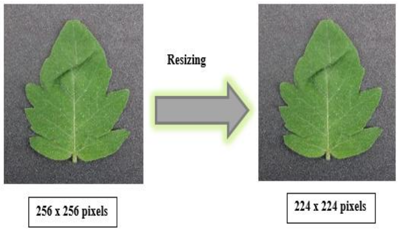


Figure 3 Image Resizing

Deep learning, a branch of artificial intelligence, mimics the human brain by using artificial neural networks to analyze unstructured data and extract meaningful features. Its hierarchical architecture allows automated feature extraction and classification, making it particularly effective for image-based tasks. Convolutional Neural Networks (CNNs) are widely used in computer vision for object recognition and image classification due to their ability to capture spatial features through convolution, pooling, and activation layers, while dropout and batch normalization help prevent overfitting and stabilize training. In this research, a sequential CNN model was developed from scratch with 21 layers, including Conv2D, MaxPooling, Batch Normalization, Dropout, Flatten, and Dense layers. Filters of 3×3 were applied with increasing node sizes (32, 64, 128), ReLU activation was used in convolutional layers, and SoftMax in the output layer for probability-based classification across nine disease categories. This architecture, combined with dropout regularization (0.25) and pooling strategies (2×2 and 3×3), achieved effective feature extraction while minimizing overfitting, enabling accurate classification of tomato leaf diseases. The layer structure of the CNN model for experiments 1 and 2 is shown in Tables 3 and 4.

Table 3: CNN Layers constructed for Experiment 1

Layer		**Shape**	**Kernel Size**	**Parameters**	**Activation Function**
Input	**Image**	**224x224x3**	**----**	**0**	**----**
1	conv2d	224x224x32	3x3	896	Relu
2	batch normalization	224x224x32	----	128	----
3	max_pooling2d	74x74x32	3x3	0	----
4	Dropout	74x74x32	----	0	----
5	conv2d_1	74x74x64	3x3	18496	Relu
6	batch_normalization_1	74x74x64	----	256	----
7	conv2d_2	74x74x64	3x3	36928	Relu
8	batch_normalization_2	74x74x64	----	256	----
9	max_pooling2d_1	37x37x64	2x2	0	----
10	dropout_1	37x37x64	----	0	----
11	conv2d_3	37x37x128	3x3	73856	Relu
12	batch_normalization_3	37x37x128	----	512	----
13	conv2d_4	37x37x128	3x3	147584	Relu
14	batch_normalization_4	37x37x128	----	512	----
15	max_pooling2d_2	18x18x128	2x2	0	----
16	dropout_2	18x18x128	----	0	----
17	Flatten	41472	----	0	----
18	Dense	1024	----	42468352	Relu
19	batch_normalization_5	1024	----	4096	----
20	dropout_3	1024	----	0	----
21	dense_1	9	----	9225	Softmax

Table 4: CNN Layers constructed for Experiment 2

Layer		**Shape**	**Kernel Size**	**Parameters**	**Activation Function**
Input	**Image**	**224x224x3**	**----**	**0**	**----**
1	conv2d_1	224x224x32	3x3	896	relu
2	batch_normalization_1	224x224x32	----	128	----
3	max_pooling2d_1	74x74x32	3x3	0	----
4	dropout_1	74x74x32	----	0	----
5	conv2d_2	74x74x64	3x3	18496	Relu
6	batch_normalization_2	74x74x64	----	256	----
7	conv2d_3	74x74x64	3x3	36928	relu
8	batch_normalization_3	74x74x64	----	256	----
9	max_pooling2d_2	37x37x64	2x2	0	----
10	dropout_2	37x37x64	----	0	----
11	conv2d_4	37x37x128	3x3	73856	Relu
12	batch_normalization_4	37x37x128	----	512	----
13	conv2d_5	37x37x128	3x3	147584	Relu
14	batch_normalization_5	37x37x128	----	512	----
15	max_pooling2d_3	18x18x128	2x2	0	----
16	dropout_3	18x18x128	----	0	----
17	flatten_1	41472	----	0	----
18	dense_1	1024	----	42468352	Relu
19	batch_normalization_6	1024	----	4096	----
20	dropout_4	1024	----	0	----
21	dense_2	3	----	3075	Softmax

The next step is the compilation of the model. There are three parameters for compiling the model: Optimizer, Loss, and Metrics. Optimizer controls and adjusts the rate of learning in the model throughout the training. Adam optimization is a method of stochastic gradient descent based on adaptive estimation of moments of first and second order. In particular, the learning rate is a configurable hyperparameter used in neural network training, having a small positive value, sometimes between 0.0 and 1.0. Optimizer used in this proposed model for both experiments is "Adam" and is instantiated with the *compile () function*. Adam is typically a strong optimizer to be found in certain situations. Here, the learning rate determines how fast the optimal weights are calculated for this model. Loss functions are meant to measure the sum that a model will seek to reduce during preparation. The loss function used for both experiments is categorical_crossentropy, a common function for classification, and shows better results if the value of scores is lower. A metric is a function that is used to calculate the output of the layout. Metric functions are like loss functions, so that the results of calculating a metric are not considered in evaluating the algorithm. The metric function used for both experiments is 'accuracy' for the precision value on the validation set while training the model.

Training and testing the model

To train a network means to get kernels for the convolution layers and to find weights for the linked layers that, when working on a training dataset, make the difference between predictions and ground truth as small as possible. For preparing our algorithm, we used 70 % of the total data and set up the network to focus on the set of extracted features from plants' images. The more epochs you run, the better the model's results, although there is a limit. Beyond the first epoch, the model would no longer develop. Experiment one had 20 epochs, and experiment two had just 10. The test predictions used the prediction function with an array of nine numbers, which corresponds to experiment 1. The numbers given are probabilities for each digit (0–8) to be the output for experiment 1 and 4 numbers for experiment 2, with the probabilities for each digit (0–3) to be the output. How well models predict is measured by choosing the array with the highest index. All the arrays add up to 1. For the second experiment, 3 pathogen classes of tomato crops with labeling 0 for fungal class, 1 for bacterial class, and 2 for virus class are used.

Results and Discussion

The proposed model is implemented and executed in Google Collaboratory, also known as Google Colab. In Experiment 1 and Experiment 2, TensorFlow functions as the backend during training. This work uses Keras as part of the TensorFlow library. A Convolutional Neural Network (CNN) was chosen to implement the model in this paper. A new Sequential CNN model is made from scratch with these layers. It consists of different CNN layers and includes in total 21 layers for both Experiments. Dropout, layers for MaxPooling2D, Dense Layers, and a Flatten Layer are used. Model output should be judged mainly on how accurate it is. Different numbers of leaf images are included for each class, and a total of 14233 images are used in two experiments that form the proposed model. Sorted the data into a 70:30 ratio so that 70% was used for training and 15% each for testing and validating the classifier in both experiments. Accuracy for Experiment 1 was 93%, using 8 types of Tomato disease and one healthy Tomato class in 20 epochs as shown in figure 4.

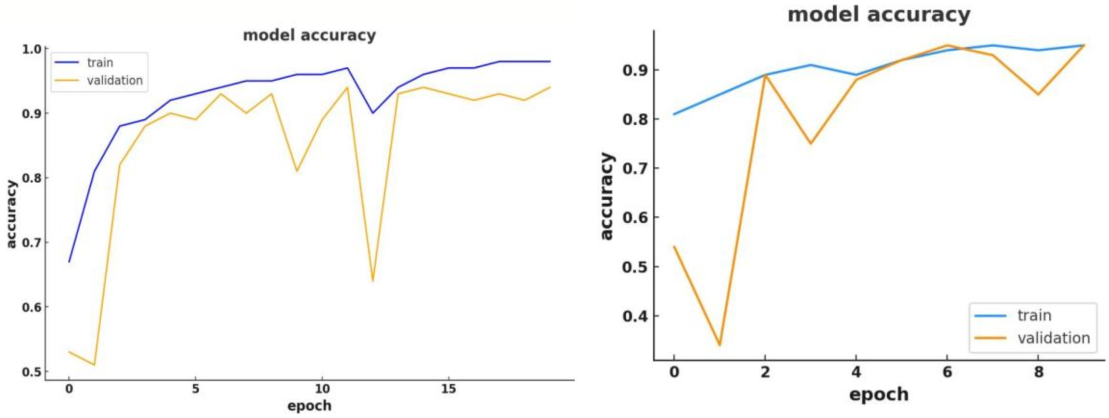


Figure **Error! No text of specified style in document.**: Plots of Accuracy against Epochs in Experiments 1 and 2

Experiment 2 achieved 96% accuracy, based on the detection of disease pathogens, i.e., fungi, Bacteria, and viruses, in tomato plants using 10 epochs. Model loss is a scalar value during training of the model that is attempted to minimize. The predictions are closer to the true labels, and the loss is lower, as shown in Figure 5.

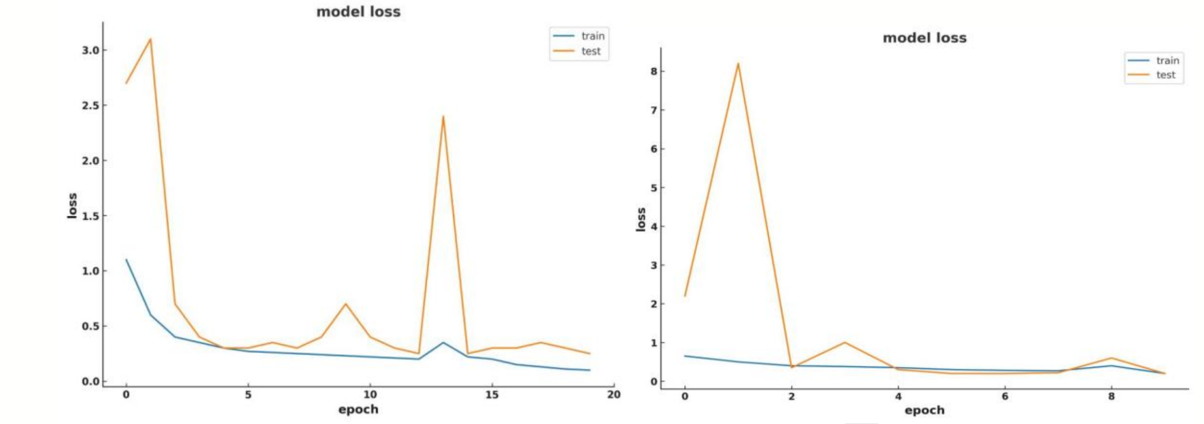


Figure 5: Plots of Loss against Epochs in Experiments 1 and 2

In Experiment 1, the model's overall performance was evaluated in classifying tomato leaf illnesses across nine different categories. According to the confusion matrix as shown in Figure 6, although accuracy varied by class. The high recall (>0.94) and precision of Leaf mold, Yellow Leaf Curl Virus, Target spot, and Healthy leaves were confirmed by strong diagonal dominance in the matrix, demonstrating that these classes were repeatedly accurately identified. Although the confusion matrix revealed considerable overlap with other bacterial-like symptoms, which was reflected in its lower precision (0.92), bacterial spot also demonstrated a high recall (0.97). Early blight had the lowest recall (0.73) due to the most obvious misclassifications with Target Spot and Septoria Leaf Spot. Early blight and Septoria both produce concentric necrotic lesions and similar chlorotic halos at overlapping developmental stages, making visual separation difficult even for human experts. Likewise, the Mosaic virus displayed a high number of false positives, which accounted for its lower precision (0.88), often being confused with Yellow Leaf Curl Virus. As these both are viral diseases and can induce mosaic patterns and leaf curling, creating overlapping visual cues. Overall, Experiment 1's confusion matrix shows that although the model was able to accurately represent most diseases, several categories, especially early blight and Mosaic virus, were more likely to cause misunderstanding due to their overlapping symptoms and complex biological pathology.

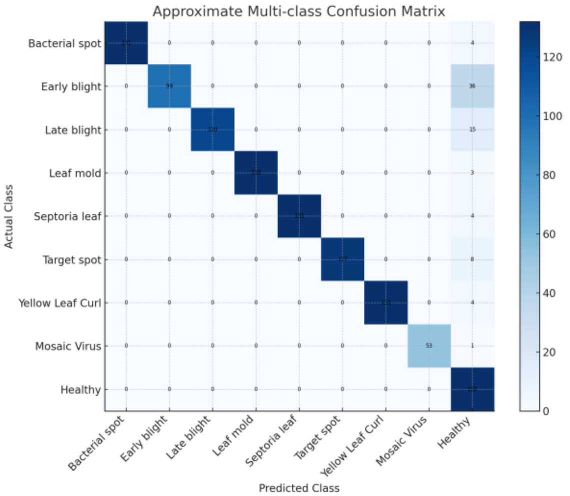


Figure 6: Confusion Matrix for Experiment 1

In contrast to Experiment 1, grouping the same images into three broader pathogen categories (Fungal, Bacterial, Viral) simplified decision boundaries and produced 96% accuracy in only 10 epochs. The confusion matrix for Experiment 2, as shown in Figure 7, displayed greater diagonal values and fewer off-diagonal errors, indicating a sharper class distinction. With a recall of 0.98 and a precision of 0.97, fungal diseases dominated the diagonal and achieved nearly perfect classification. Although the confusion matrix showed sporadic misclassification with fungal samples, which accounted for the somewhat lower precision (0.90), bacterial illnesses were also reliably diagnosed. The matrix demonstrated high precision (0.96) for viral infections; however, certain instances that were incorrectly diagnosed resulted in a lower recall (0.91). Experiment 2's confusion matrix shows better robustness and lower total misclassification than Experiment 1 since the grouping technique made decision boundaries simpler. However, because differences between fungal or viral infections were no longer reflected, this increase in matrix clarity comes at the price of more precise disease-level identification.

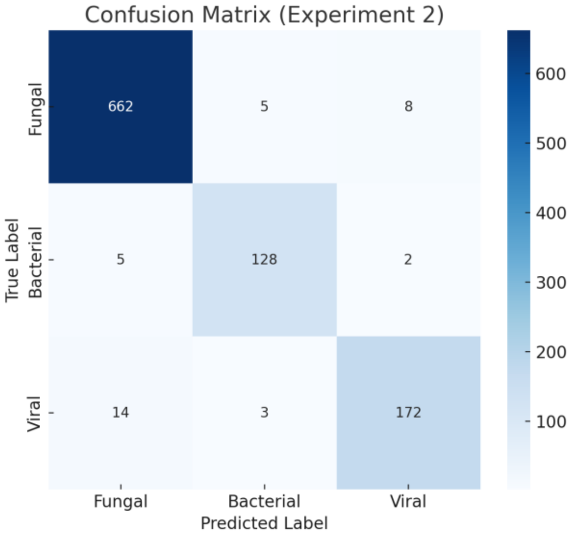


Figure 7: Confusion Matrix for Experiment 2

The results obtained after constructing a CNN from scratch for the Classification and Detection of Tomato leaf Diseases affected by Fungal, Bacterial, and Viral Pathogens, based on two Experiments, are shown in Table 5. The preliminary purpose of this research is to detect the cause of tomato plant diseases.

Table **Error! No text of specified style in document.**: Results for Experiments 1 and 2

Experiments	**Class**	**Recall**	**Precise**	**Specificity**	**F1-Score**
Experiment 1	Bacterial spot	0.97	0.92	0.98	0.94
	Tomato Early blight	0.73	0.95	0.99	0.82
	Tomato Late blight	0.89	0.85	0.97	0.87
	Tomato Leaf mold	0.98	0.95	0.99	0.96
	Tomato Septoria leaf	0.97	0.89	0.98	0.93

	Tomato Target spot	0.94	0.96	0.99	0.95
	Tomato Yellow Leaf Curl	0.97	0.98	0.99	0.97
	Tomato Mosaic Virus	0.98	0.88	0.99	0.92
	Tomato Healthy	0.98	0.99	0.99	0.97
Experiment 2	Fungal	0.98	0.97	0.95	0.97
	Bacterial	0.95	0.90	0.98	0.93
	Viral	0.91	0.96	0.99	0.93

Deep learning–based tomato leaf disease detection offers significant advantages over conventional Phyto-pathological tools such as ELISA, PCR, and expert visual scouting. Unlike laboratory assays, which are highly specific but time-consuming, costly, and limited to targeted pathogens, deep learning models provide rapid, non-invasive, and scalable diagnosis using simple leaf images. Compared to visual scouting, which is prone to human bias and limited accuracy, CNN-based systems objectively analyze multiple diseases simultaneously with competitive accuracy approaching laboratory standards. Thus, while ELISA and PCR remain indispensable for confirmatory testing, deep learning presents a more practical solution for large-scale, real-time field monitoring and early detection, especially in resource-constrained agricultural contexts(Khakimov et al., 2022). When benchmarked against state-of-the-art deep learning methods, the proposed model delivers satisfactory accuracy in both experimental settings as shown in Table 6.

Table 6: Accuracy Comparison with state-of-the-art deep learning methods

Author(s), Year	**Algorithm (Classification)**	**Dataset Detail**	**Specie Name**	**Accuracy**
(Fuentes et al., 2017)	R-CNN	9 types of tomato diseases and pests	Tomato Plant	83%
(Sibiya & Sumbwanyambe, 2019)	CNN	Fungal images (Plant Village dataset)	Apple Specie	90.4%
(Sun et al., 2017)	CNN	10,000 of 100 species (BJFU100 dataset)	Ornamental plant species	91.78%
(Ashqar & Abu-Naser, 2019)	CNN	600 images (Plant Village dataset)	Maize Plant	92%
(Tang et al., 2023)	CNN	1,909 images (Plant Village dataset)	Tomato Plant	94.5%
(R. Khan et al., 2024)	CNN	2,700 images (Plant Village dataset)	Tomato Plant	92.3%
Proposed Method 2025 **Experiment 1**	CNN	7560 images (Plant Village dataset)	Tomato Plant	93%

	Experiment 2	CNN	6673 images (Plant Village dataset)	Tomato Plant	96%

The results of this study confirm the potential of deep learning–based image analysis for tomato leaf disease detection, yet their significance extends beyond accuracy metrics. From a plant pathological perspective, the proposed CNN model contributes to disease management by offering rapid, non-invasive, and scalable detection. Accurate differentiation between diseases such as Septoria leaf spot, Early blight, and Target spot remains challenging due to overlapping symptoms, but the model's performance demonstrates that computer vision can support early recognition and reduce scouting errors that often limit Integrated Pest Management (IPM) programs. When integrated with existing management frameworks, such systems can provide timely decision support, alerting farmers and extension services before disease outbreaks intensify. However, limitations must be acknowledged. The dataset was curated under controlled conditions, raising risks of bias and overfitting that may hinder generalization in real farm environments with variable lighting, occlusion, or mixed infections. Reliance solely on visual symptoms also restricts precision at early disease stages when lesions are not distinct. Therefore, field validation and integration with complementary tools such as weather-based forecasting or confirmatory ELISA/PCR remain necessary for robust deployment. Looking forward, integrating CNN-based detection with smartphone applications, drone imaging, and weather-driven disease forecasting could provide powerful early-warning tools. A hybrid framework—using deep learning for rapid field screening and molecular assays such as ELISA or PCR for confirmatory testing would align with IPM principles, combining speed, scalability, and accuracy for sustainable tomato disease management.

CONCLUSION

Agriculture is a vital source of income in our country, and protecting major crops like tomatoeses from diseases is essential for sustaining productivity and economic growth. This study developed a 21-layer CNN to identify nine tomato leaf diseases using the PlantVillage dataset. Two experimental settings demonstrated strong performance: disease-level classification achieved 93% accuracy, while pathogen-type classification (fungal, bacterial, viral) achieved 96%. These results highlight the potential of deep learning models as practical, low-cost, and user-friendly tools for crop monitoring. By enabling early and accurate detection, such approaches can reduce dependency on chemical treatments, support environmentally sustainable practices, and strengthen agricultural diagnostics for farmers.

References

Aatif, K., Fahiem, M. A., & Tahir, F. (2024). Forecasting Floods Using Deep Learning Models: A Longitudinal Case Study of Chenab River, Pakistan. *IEEE Access*, *12*, 115802–115819. https://doi.org/10.1109/ACCESS.2024.3445586

Abdulridha, J., Ehsani, R., Abd-Elrahman, A., & Ampatzidis, Y. (2019). A remote sensing technique for detecting laurel wilt disease in avocado in presence of other biotic and abiotic stresses. *Computers and Electronics in Agriculture*, *156*, 549–557. https://doi.org/10.1016/J.COMPAG.2018.12.018

Ali, H., Lali, M. I., Nawaz, M. Z., Sharif, M., & Saleem, B. A. (2017). Symptom based automated detection of citrus diseases using color histogram and textural descriptors. *Computers and Electronics in Agriculture*, *138*, 92–104. https://doi.org/10.1016/J.COMPAG.2017.04.008

Ally, N. M., Neetoo, H., Ranghoo-Sanmukhiya, V. M., & Coutinho, T. A. (2023). Greenhouse-Grown Tomatoes: Microbial Diseases and their Control Methods: A Review. *International Journal of Phytopathology*, *12*(1), 99–127. https://doi.org/10.33687/phytopath.012.01.4273

Arya, S., & Singh, R. (2019). A Comparative Study of CNN and AlexNet for Detection of Disease in Potato and Mango leaf. *IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques, ICICT 2019*. https://doi.org/10.1109/ICICT46931.2019.8977648

Ashqar, B. A. M., & Abu-Naser, S. S. (2019). Identifying Images of Invasive Hydrangea Using Pre-Trained Deep Convolutional Neural Networks. *International Journal of Control and Automation*, *12*(4), 15–28. https://doi.org/10.33832/IJCA.2019.12.4.02

Caruso, A. G., Bertacca, S., Parrella, G., Rizzo, R., Davino, S., & Panno, S. (2022). Tomato brown rugose fruit virus: A pathogen that is changing the tomato production worldwide. *Annals of Applied Biology*, *181*(3), 258–274. https://doi.org/10.1111/AAB.12788;WGROUP:STRING:PUBLICATION

Dhingra, G., Kumar, V., & Dutt Joshi, H. (2019). *A novel computer vision based neutrosophic approach for leaf disease identification and classification*. https://doi.org/10.1016/j.measurement.2018.12.027

Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. *Sensors 2017, Vol. 17, Page 2022*, *17*(9), 2022. https://doi.org/10.3390/S17092022

Hassanien, A. E., Gaber, T., Mokhtar, U., & Hefny, H. (2017). An improved moth flame optimization algorithm based on rough sets for tomato diseases detection. *Computers and Electronics in Agriculture*, *136*, 86–96. https://doi.org/10.1016/J.COMPAG.2017.02.026

Hughes, D. P., & Salathé, M. (n.d.). *An open access repository of images on plant health to enable the development of mobile disease diagnostics*. Retrieved September 25, 2025, from http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf!!!

Jayswal, H. S., & Chaudhari, J. P. (2023). Plant Leaf Diseases Detection and Classification Using Spectroscopy. *Smart Innovation, Systems and Technologies*, *324*, 473–483. https://doi.org/10.1007/978-981-19-7447-2_42

K., B., & Rao, M. (1 C.E.). Tomato Plant Leaves Disease Classification Using KNN and PNN. *Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/IJCVIP.2019010104*, *9*(1), 51–63. https://doi.org/10.4018/IJCVIP.2019010104

Khakimov, A., Salakhutdinov, I., Omolikov, A., & Utaganov, S. (2022). Traditional and current-prospective methods of agricultural plant diseases detection: A review. *IOP Conference Series: Earth and Environmental Science*, *951*(1), 012002. https://doi.org/10.1088/1755-1315/951/1/012002

Khan, M., Gulan, F., Arshad, M., Zaman, A., & Riaz, A. (2024). Early and late blight disease identification in tomato plants using a neural network-based model to augmenting agricultural productivity. *Science Progress*, *107*(3). https://doi.org/10.1177/00368504241275371

Khan, R., Ud Din, N., Zaman, A., & Huang, B. (2024). Automated Tomato Leaf Disease Detection Using Image Processing: An SVM-Based Approach with GLCM and SIFT Features. *Journal of Engineering*, *2024*(1), 9918296. https://doi.org/10.1155/2024/9918296

Kwabena, P. M., Weyori, B. A., & Mighty, A. A. (2020). Gabor Capsule Network for Plant Disease Detection. *International Journal of Advanced Computer Science and Applications*, *11*(10), 388–395. https://doi.org/10.14569/IJACSA.2020.0111048

Leach, J. E., Leung, H., & Tisserat, N. A. (2014). Plant Disease and Resistance. *Encyclopedia of Agriculture and Food Systems*, 360–374. https://doi.org/10.1016/B978-0-444-52512-3.00165-0

Liu, B., Zhang, Y., He, D. J., & Li, Y. (2017). Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks. *Symmetry 2018, Vol. 10, Page 11*, *10*(1), 11. https://doi.org/10.3390/SYM10010011

Liu, G., Nouaze, J. C., Mbouembe, P. L. T., & Kim, J. H. (2020). YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. *Sensors (Switzerland)*, *20*(7). https://doi.org/10.3390/S20072145,

PlantVillage Dataset. (n.d.). Retrieved September 25, 2025, from https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset/data

Qasim, M., Farooq, W., & Akhtar, W. (2018). *Policy and Institutional Reforms to Improve Horticultural Markets in Pakistan (ADP/2014/043) Preliminary Report on the Survey of Tomato Growers in Sindh, Punjab and Balochistan*.

Ribeiro, H. V., Lopes, D. D., Pessa, A. A. B., Martins, A. F., da Cunha, B. R., Gonçalves, S., Lenzi, E. K., Hanley, Q. S., & Perc, M. (2023). Deep learning criminal networks. *Chaos, Solitons & Fractals*, *172*, 113579. https://doi.org/10.1016/J.CHAOS.2023.113579

Ronihal, S. S., Selvam, S., & Kusuma, P. (2025). Enhancing Crop Yield: A Deep Learning Approach to Detect Tomato Leaf Diseases. *Proceedings of the 3rd International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, IITCEE 2025*. https://doi.org/10.1109/IITCEE64140.2025.10915243

Sattar, S., Iqbal, A., Parveen, A., Fatima, E., Samdani, A., Fatima, H., Iqbal, M. S., & Wajid, M. (2024). Tomatoes Unveiled: A Comprehensive Exploration from Cultivation to Culinary and Nutritional Significance. *Qeios*. https://doi.org/10.32388/CP4Z4W.2

Sibiya, M., & Sumbwanyambe, M. (2019). A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks. *AgriEngineering 2019, Vol. 1, Pages 119-131*, *1*(1), 119–131. https://doi.org/10.3390/AGRIENGINEERING1010009

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends in Plant Science*, *23*(10), 883–898. https://doi.org/10.1016/J.TPLANTS.2018.07.004, Singh, V. K., Singh, A. K., Singh, P. P., & Kumar, A. (2018). Interaction of plant growth promoting bacteria with tomato under abiotic stress: A review. *Agriculture, Ecosystems & Environment*, *267*, 129–140. https://doi.org/10.1016/J.AGEE.2018.08.020

Sun, Y., Liu, Y., Wang, G., & Zhang, H. (2017). Deep Learning for Plant Identification in Natural Environment. *Computational Intelligence and Neuroscience*, *2017*(1), 7361042. https://doi.org/10.1155/2017/7361042

Tang, Z., He, X., Zhou, G., Chen, A., Wang, Y., Li, L., & Hu, Y. (2023). A Precise Image-Based Tomato Leaf Disease Detection Approach Using PLPNet. *Plant Phenomics*, *5*. https://doi.org/10.34133/PLANTPHENOMICS.0042/SUPPL_FILE/PLANTPHENOMICS.0042.F1.DOCX

Tm, P., Pranathi, A., Saiashritha, K., Chittaragi, N. B., & Koolagudi, S. G. (2018). Tomato Leaf Disease Detection Using Convolutional Neural Networks. *2018 11th International Conference on Contemporary Computing, IC3 2018*. https://doi.org/10.1109/IC3.2018.8530532

Wang, C., Li, M., Duan, X., Abu-Izneid, T., Rauf, A., Khan, Z., Mitra, S., Emran, T. Bin, Aljohani, A. S. M., Alhumaydhi, F. A., Thiruvengadam, M., & Suleria, H. A. R. (2023). Phytochemical and Nutritional Profiling of Tomatoes; Impact of Processing on Bioavailability - A Comprehensive Review. *Food Reviews International*, *39*(8), 5986–6010. https://doi.org/10.1080/87559129.2022.2097692

Wang, G., Sun, Y., & Wang, J. (2017). Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Computational Intelligence and Neuroscience*, *2017*. https://doi.org/10.1155/2017/2917536,