Journal of Intelligent Systems and Internet of Things
Vol. 14, No. 02, PP. 213-228, 2025
DOI: https://doi.org/10.54216/JISIoT.140217

Classification of Tomato Diseases Using Deep Learning Method

Adnan M. A. Shakarji[1,*], Adem Gölcük[2]

[1]Institute of Sciences, Selcuk University, Konya, Turkey
[2]Computer Engineering Department, Selcuk University, Konya, Turkey

Emails: kirkuk_adnan@yahoo.com; adem.golcuk@selcuk.edu.tr

Abstract

With an average annual intake of almost 20 kilograms per person, tomatoes are the most consumed vegetable worldwide. Diseases brought on by dangerous organisms are among the most important factors adversely affecting tomato production's output and quality. Depending on the climate and environmental conditions, tomatoes can be afflicted by a variety of illnesses throughout the planting and growing phases. It is essential for tomato growers to identify possible infections and take the appropriate preventative measures. Applications of artificial intelligence have grown in popularity recently. AI is being used in agriculture to identify plant illnesses. This research uses deep learning, a branch of artificial intelligence, to categories common tomato diseases. In the beginning, samples of frequently seen tomato illnesses were gathered from tomato growers in Kirkuk. Once there were enough data, the system developed with image processing algorithms produced meaningful images. Using a CNN-based GoogleNet deep learning system, the resulting dataset was trained and diseases were classified. The results show that the deep learning system that was constructed has a high degree of success and dependability when it comes to tomato disease classification.

Received: April 03, 2024 Revised: July 19, 2024 Accepted: November 05, 2024

Keywords: Deep Learning; Tomato Disease Detection; Image Processing; GoogleNet; Artificial Intelligence

1. Introduction

One of the main sources of income for people is agriculture. Turkey's geographical and biological advantages make it a perfect place to develop a wide variety of agricultural products [1]. The growing global population is causing a number of issues, including a shortage of food. Scientists have been activated and studies to boost yield have begun because today's agricultural lands cannot give enough resources and the yield received from the unit area is low. One of the most widely grown, consumed, and traded agricultural items worldwide, tomatoes are essential to human nutrition.

Tomatoes are a rich source of vitamins, minerals, organic acids, essential amino acids, and dietary fibre. It is also a good source of potassium, vitamin A, and vitamin C, and it contains minerals like iron and phosphorus. Tomatoes are minimal in fat and calories and a rich source of dietary fibre without cholesterol. Tomatoes are considered a protective plant because of its unique nutritional content, which includes lycopene, beta carotene, and flavonoids. Because of its anti-oxidant and anti-cancer qualities, lycopene has gained a lot of popularity, especially in recent years. Both tomato production and consumption are steadily increasing as a result of these causes. The most significant polysaccharides found in ripe, fresh tomato fruit are pectins, xylans, arabinoxylans, and cellulose. They contain trace amounts of sucrose along with substantial levels of glucose and fructose. In fresh tomato juice, glutamic acid makes up 45% of the free amino acids, with aspartic acid coming in second. The most prevalent organic acid is citric acid, however malic acid is also found. Thirty of the more than 400 compounds that have been identified as influencing tomato flavour and scent have a significantly greater impact on aroma production. Tomato taste is formed by a combination of organic acids, sugars, free amino acids, mineral salts, and volatile and changeable fragrance compounds. The carotenoids, particularly lycopene, are the metabolite that gives tomatoes their red hue. The species, maturity stage, climate, temperature, light, soil, fertilisation, irrigation, and other treatments during cultivation, harvest, and storage conditions all affect the amount of each of these metabolites.

Plants that thrive in warm, hot regions, particularly those with high temperatures, produce ripe, dark-coloured fruits that are high in sugar. Tomato plants, however, do not reach full maturity in cold and cool climes, and they may even perish at temperatures as low as -3 °C. Furthermore, tomatoes are very vulnerable to environmental stressors like drought, high salinity, harsh temperatures, and pollution [2]. Analysis of the global causes of agricultural output losses reveals that diseases account for 9.1%, pests for 11.2%, and weeds for 14.7%. This rate is nearly equal to one-third of global agricultural output. The overall loss rate exceeds 50% when post-harvest losses—which can range from 6 to 12%—are included [3]. They created an embedded system that can identify leaf illnesses in real time. By using this system to run an algorithm on RGB (Red-Green-Blue) photos and carry out parallel processing, they hoped to expedite the illness identification process. As a result of the study, they created an automatic system that can identify diseases on real-time photographs [4]. They used a database of cucumber leaves to do illness identification. With an accuracy rate of 86%, the researchers used the k-means clustering algorithm to partition the diseased leaves, extracted the shape and colour data to identify diseases, and then used the sparse representation approach to classify these diseased leaves [5]. To detect rice diseases, it employs deep convolutional neural network techniques. For training, 500 images of both healthy and diseased rice leaves were used. The trained model detects ten different common disorders. According to the experimental results, 95.48% of disease cases are correctly identified [6]. They examined several methods for classifying and diagnosing plant diseases. This study evaluated the designs of support vector machines (SVM), artificial neural networks (ANN), fuzzy logic classifiers, k-nearest neighbour algorithms (k-NN), and CNN. CNN architecture was found to be more accurate than the other four methods in identifying a greater number of disorders [7]. They trained a CNN network using images captured by smartphone cameras in order to detect and classify maize leaf diseases. CNN's suitability for this market was proved by the study's average accuracy percentage of 92.85% [8]. Al-Amin created a Convolutional Neural Network (CNN) model to identify disease in potato leaves, and he discovered that it had a high accuracy rate of 98.33% [9].

Another study focused on the creation and evaluation of an image-based dataset for the classification of diseases in the culture of Agaricus bisporus (J.E. Lange) Imbach. The dataset contains both healthy images and images of different kinds of illnesses. This research is thought to generate a dataset that could be used for the identification and classification of fungal diseases, in addition to facilitating the use of deep learning or other machine learning techniques that will allow the automatic identification and classification of diseases. The portable mushroom imaging system created for the study was used to visit mushroom farms during the dataset creation process. Approximately 3000 images of each lighting environment were obtained, including 1800 images of healthy mushrooms and 7250 images of diseased mushrooms. We found four types of illnesses that are prevalent in grown mushrooms. Three distinct lighting settings were used to picture each fungus [10]. One of the most widely grown and consumed cereal crops in the world is wheat. To categorise bread wheat variations, a pre-trained hybrid model based on convolutional neural networks (CNNs) is suggested. A bread wheat image dataset for deep learning was created by acquiring and separating images of five distinct registered bread wheat kinds using image processing techniques. The Xception model, one of the pre-trained CNN models, was then refined by transfer learning to classify the photos. Xception CNN models and BiLSTM (Bidirectional Long Short-Term Memory) algorithms hybrid (Xception + BiLSTM) models were developed to improve the classification success. The Xception + BiLSTM model achieved the highest classification success with 97.73% as a consequence of the classifications. The findings demonstrated the applicability of the suggested techniques for both the automatic extraction of pure wheat varieties and the classification of bread wheat variants [11]. Furthermore, 8354 photos of certified "Ayten Abla," "Bayraktar 2000," "Hamitbey," "Şanlı," and "Tosunbey" bread wheat varieties—among the staple foods—were used to construct a data set. Four steps were involved in the classification of wheat genotypes utilising pictures of bread wheat genotypes. Using an image processing and feature selection method, 90 colour (C), 4 shape (S), and 12 morphological (M) features were first recovered from the images in this dataset. The retrieved traits were merged in various ways in the second stage. Using the Artificial Bee Colony (ABC) algorithm, feature selection was done from all of the combined features in the third stage to determine which features are beneficial in classifying performance. Ultimately, machine learning algorithms, including Support Vector Machines (SVM), Decision Trees (DT), and Quadratic Discriminant (QD) classifiers, were utilised to classify bread wheat genotypes based on these features discovered in three steps [12].

Plant disease is mostly caused by bacteria, viruses, fungus, and unfavourable environmental circumstances. Vital plant processes like photosynthesis, pollination, fertilisation, and germination can be harmed by diseases. For this reason, early disease detection is crucial. The amount and quality of the items produced are reduced when plants are diseased. Early disease detection and therapy implementation are crucial to averting this circumstance. For the farmer, identifying and treating plant diseases can take a lot of time. Expert assistance may be required in situations where the disease cannot be identified. The farmer is further burdened financially as a result. Faster and more precise techniques are required to save the farmer from this burden and loss. These days, it is possible to identify whether a plant has a disease and, if so, what kind of sickness it is by using technical gadgets rather than an outside expert. Using image processing and artificial intelligence algorithms, operations like object detection and categorisation through images are highly successful due to the improvement in image acquisition quality of technological instruments. In certain instances, image processing and artificial intelligence algorithms operating on such technological gadgets can readily expose a detail that the human eye is unable to discern. The idea of precision agriculture has arisen as a result of these technical advancements and agricultural practices. Precision agriculture techniques fall into a number of categories, including remote data collection with technology, mapping the soil, measuring temperature and moisture, determining whether agricultural products are intact, rotten, overripe, etc., counting products, irrigation, and spraying. Through precision agriculture, the goal is to minimise environmental harm while simultaneously increasing product yield and revenue at a low cost. Applications of precision agriculture are utilised to identify plant diseases. In the agriculture sector, deep learning techniques—a branch of artificial intelligence—are employed to identify plant diseases and their varieties.

In this study, common tomato diseases were classified by deep learning method. Examples of common tomato plant diseases were obtained by conducting a study with tomato growers operating in Kirkuk province. To obtain more samples for the study, diseased tomatoes were grown as much as possible. After reaching enough samples, a data set was prepared by obtaining meaningful images for the computer system using image processing techniques. The deep learning system was trained with the created dataset and tomato diseases were successfully classified.

2. Tomato Diseases

A detailed review of the literature on tomato diseases was the first step in the project. Diseases that can be identified using visual analysis techniques were chosen as a consequence of the literature review. Visual perception-based disease investigations were examined. To detect diseases that may be visually diagnosed, a new artificial intelligence model was created.

A. Tomato Early Leaf Disease (Alternaria Solani)

This disease can be encountered at any stage of the growing process. It is especially seen as spots on leaves, stems and fruits. It also causes root rot or root collar blight in seedlings. The spots first start small and dark in color and then grow larger. They turn dark gray and form spots in concentric rings. If the disease is severe, all leaves dry up and fall off. Flowers and fruit stalks can also be affected by this disease. In the fruits, dark, sunken, mostly bordered spots usually occur on the part where the stem is attached. The disease kills the plant in a short time. The disease is caused by a fungus called Alternaria Solani [13]. Figure 1 shows a tomato with Alternaria Solani disease.

Figure 1. Tomato Early Leaf Disease (Alternaria Solani)

B. Tomato Leady Mildew Disease (Botrytis Cinerea)

The causal agent of lead mildew has a very wide host range and causes a different disease picture in each host. It causes stem, fruit and leaf infections in tomato. The lesions, which are as small as a pinhead at first, develop under the epidermis, expand and spread to the tissues. The epidermis cracks and causes dehydration of the host. Lesions on the stem and fruit stalk may cause fruit drop. The petals of the host at flower time are very susceptible to the disease. The fungus enters the fruit through these parts and initiates fruit rot. The disease is caused by the fungus Botrytis cinera [14]. Figure 2 shows a tomato with Botrytis cinerea disease.

Figure 2. Tomato Lead Blight Disease (Botrytis Cinerea)

C. Tomato Mildew Disease (Phytophthora Infestans)

The disease starts with small, pale green or yellowish spots on the leaves, then the color of the spots changes from brown to black as the disease progresses. In favorable weather, the disease progresses to the petioles, branches and stems of the plant. In humid weather, a white or ash-colored cover is formed on the underside of the spots. In advanced stages, the spots tear, dry and sometimes rot. The disease can also spread to fruits. Brown spots on the fruit are distinguished from the normal red part by a green frame when the tomato is reddened. The disease is caused by Phytophthora infestans [15]. Figure 3 shows a tomato with Phytophthora infestans disease.

Figure 3. Tomato Mildew Disease (Phytophthora Infestans)

D. Bacterial Spot Disease of Tomato (Xanthomonas Vesicatoria)

The disease can be seen in all parts of the plant (stem, leaf, fruit). It can be seen in all parts of the plant (stem, leaf, fruit). The spots on the leaves, which are initially in the form of oil drops and surrounded by a yellow halo, turn into a brownish black color as the disease progresses. In the later stages of the disease, these spots grow and merge and cause the leaf to turn yellow and dry. If the disease agent infects the plants during the seedling period, the seedlings and young plants become scorched and die. The first symptoms on the fruit stem are similar to the leaf symptoms and are in the form of long lines. The proliferation and spread of symptoms on the stem causes the flowers to die. The first symptoms on the fruit are small, greenish-white and surrounded by halos. As the disease progresses, these spots enlarge, turn brownish black and become crater-like and sunken. The disease is caused by Xanthomonas vesicatoria [15]. Figure 4 shows a tomato with Xanthomonas vesicatoria disease.

Figure 4. Bacterial Spot Disease of Tomato (Xanthomonas Vesicatoria).

E. Tomato Bacterial Cancer and Wilt Disease (Clavibacter michiganensis subsp.michiganensis)

The first symptoms of the disease are mostly seen as inward curling, browning and wilting of the leaflets at a single point of the plant. The first symptoms are mostly seen as inward curling, browning and wilting of the leaflets at a single point of the plant. In infections occurring during the seedling stage, young seedlings may remain weak and stunted or may wither and die rapidly. When conditions are not favorable, seedlings may not show symptoms until the beginning of flowering. When plants approach the flowering stage, the disease usually starts with wilting of the lower leaves and the wilting progresses upwards. Sometimes one-sided wilting may develop. The wilted parts dry up after a short time. In advanced stages, the color of the conduction bundles turns brown and cracks called cancers occur on the trunk and side branches. Since the conduction bundles of wilted plants are blocked, water and nutrients cannot reach the leaves and the leaves look as if they are parched from thirst. Fruit spots usually appear on green fruits, 2-3 mm in diameter, round, with a dark brown center and surrounded by a white halo. These spots are typical symptoms of the disease and are described as "bird's eye spots". The disease is caused by Clavibacter michiganensis subsp [16]. Figure 5 shows a tomato with Clavibacter michiganensis disease.

Figure 5. Tomato Bacterial Cancer and Wilt Disease (Clavibacter Michiganensis).

F. Bacterial Spot Disease of Tomato (Pseudomonas Syringae pv.tomato)

Spots appear on all of the plant's aboveground organs as a result of the illness. During the seedling stage, symptoms begin to show themselves, with numerous brown-black spots developing on the seedlings' leaves and stems. Over time, the entire seedling dries up as a result of these patches. Similar to the spots that form during the seedling phase, the disease causes brown to black spots to form on the leaves, stems, flowers, and fruit stalks throughout the field period. These areas typically have a golden halo. As time passes, the spots combine, deforming and drying the leaf. Pseudomonas syringae pv. tomato is the causative agent of the disease [17]. A tomato with Pseudomonas syringae disease is seen in Figure 6.

Figure 6. Bacterial Spot Disease in Tomato (Pseudomonas Syringae).

3. Deep Learning

Big data and computer processing power are used in deep learning, a well-liked subfield of machine learning that is based on artificial neural networks. Despite having historical roots, deep learning has just recently gained popularity. Deep learning techniques have witnessed a sharp rise, especially as the number of data and graphics processing units (GPUs) in the computing industry has increased and GPUs have been more reasonably priced. Data processing units and matrices are processed by GPUs in deep learning. Multiple calculations and processes can be carried out simultaneously using GPUs. When acquiring and analyzing data, deep learning techniques don't need any human involvement. They are thus easily distinguished from traditional machine learning techniques. The ability to automatically perform feature extraction and find high-level features is one of the most crucial components of deep learning [18].

Consistent classification and predictions are made possible by deep learning models' strong hierarchical structure and extensive learning capability. Deep neural networks, which are multi-layered and multi-neuron variants of artificial neural networks (ANNs), are characterized by their ability to automatically extract features based on the problem and build these features through network learning. Image recognition is one of the most creative and effective applications of computer science and artificial intelligence. In the last decade, this field has made significant strides, especially with the introduction of Convolutional Neural Networks (CNN) technology. CNNs are used in a wide range of applications, such as image and video analysis, facial recognition software, automotive technology, and medical image analysis.

A. Convolutional Neural Networks (CNN)

One of the most significant and ground-breaking uses of artificial intelligence and computer science is image recognition. This discipline has advanced tremendously in the past ten years, particularly with the advent of Convolutional Neural Networks (CNN) technology. CNNs are employed in many different fields, including medical image analysis, automotive technology, facial recognition systems, and image and video analysis [19].

CNN's core convolution layer is utilized to identify local characteristics in a picture. To carry out the convolution process, a tiny filter or kernel is used. This filter traverses the entire image and generates an activation map by multiplying and summing each pixel value of the region that overlaps with the filter. This process is repeated to extract various features of the image (edges, texture, shapes, etc.). Figure 7 shows the general structure of the CNN architecture.

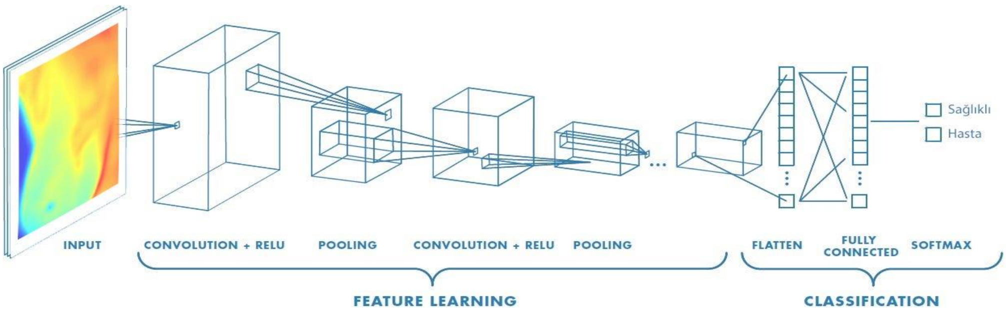

Figure 7. General Structure of CNN Architecture.

Typically, an activation function is used to the values that emerge from the convolution layer. Rectified Linear Unit, or ReLU, is the most widely used activation function. ReLU preserves positive values while setting negative values to zero. Through this process, the network's capacity for non-linear learning is enhanced, enabling the model to pick up increasingly intricate features.

While minimizing the size of the activation maps generated by the convolution layer, the pooling layer seeks to maintain significant features. Max pooling and average pooling are the two most popular pooling techniques. While average pooling takes the average of the data inside a particular window, max pooling takes the maximum value within that frame. Typically, the CNN architecture's last layers are fully connected. These layers generate an output (classification or regression) by combining all of the features from the convolution and pooling layers.

CNNs can be used in many different sectors. CNN technology are used in the automobile industry for road sign recognition and environmental sensing in autonomous vehicles. By examining medical images like X-rays, MRIs, and ultrasounds, it is crucial in the field of medical imaging to detect anomalies. It is employed in security systems to track human movements and recognize faces. It finds product flaws for production line quality control in industrial automation. It is utilized in image and video processing for tasks including object recognition, event detection from video data, and picture classification. It is also utilized for social media platform content analysis and filtering. It is essential to robotics because it gives robots the ability to sense their surroundings and identify objects, allowing them to carry out intricate jobs and support people. It is employed in agriculture to analyze plant development, identify pests, and detect plant illnesses.

CNNs can identify significant features in small image segments and develop local connections, which enables the network to weed out irrelevant data and analyze just the pertinent information. Furthermore, every convolutional filter employs the same weights across the input image because of the parameter sharing feature. This allows for efficient training with less data and drastically lowers the number of parameters the model must learn. Additionally, CNNs acquire more robust and generic characteristics by being less sensitive to slight picture shifts and changes. This enables the model to accurately classify images obtained from a variety of perspectives and situations.

However, CNNs also have certain limitations and challenges. Successfully training them often requires a large amount of labeled data, which can be a barrier, especially in fields where labeling costs are high. Moreover, CNN models are often described as "black boxes" because it is difficult to understand how and why the model arrives at specific decisions. This can lead to issues, particularly in critical applications like medical diagnosis. Finally, deep CNN models require substantial computational power during training and inference, increasing the need for high-performance GPUs and raising costs, making advanced hardware requirements a significant challenge.

4. Material And Methods

In this study, a two-stage approach was adopted to determine tomato diseases. As a result of the comprehensive literature review, detailed information was obtained about the factors that cause tomato diseases, the symptoms of these diseases and the control methods. In addition, tomato cultivation was carried out and the development of the plants was monitored daily and the causes of disease emergence and control methods were observed in practice. This two-way approach was supported by both theoretical knowledge and practical experience, creating a solid basis for the detection of tomato diseases and the determination of control methods. Visuals from the development process of the grown tomatoes are shown in Figure 8.

Figure 8. The Growth Process of Tomato Seedlings.

Data on tomato plant diseases were gathered for this study from a tomato greenhouse located in Kirkuk, Iraq. A Redmi 10 T AI four-camera system, comprising a 64-megapixel primary camera with f/1.89, an 8-megapixel ultra-wide-angle camera with f/2.2, a 2-megapixel depth camera with f/2.4, and a 5-megapixel macro camera with aperture and focal length f/2.4, was used to take pictures of tomato plants.

Below are the specifications of the computer used for deep learning in this study:

Processor (CPU): Intel Core i7-11800H (2.3GHz, 8 cores, 16 threads, 4.6GHz turbo frequency) , Graphics Card (GPU): NVIDIA 3050ti (4GB GDDR6 VRAM), RAM: 16GB DDR4 3200MHz , Storage: 500GB NVMe SSD, Display: 15.6-inch, 1920x1080 (Full HD), 144Hz refresh rate, Operating System: Windows 10 , Ports: USB 3.2, USB 2.0, HDMI, Ethernet, Headphone/Microphone jack, Battery: 48WHr

A. CNN-based GoogleNet Model

In this study, the detection and classification of tomato diseases using CNN-based GoogleNet model was realized. The construction of the GoogleNet model is a fundamental component for artificial intelligence and deep learning applications in MATLAB. There are different inception modules in between the 27 layers that make up this model's 22-layer depth structure. By mixing different filter sizes in a single layer, Inception modules enable the model to capture different feature scales. This structure has the benefit of enabling the model to identify more intricate traits in a broader context. Convolution filters of sizes 1x1, 3x3, and 5x5 are included in these modules, which help the model learn both local and global features [20]. The GoogleNet function in MATLAB makes it simple to load and make the GoogleNet model available. With this feature, users can load the model directly and carry out activities like visual classification. The model's effectiveness is especially noticeable in challenging image identification tasks and big datasets. The general layout of the GoogleNet model is depicted in Figure 9.

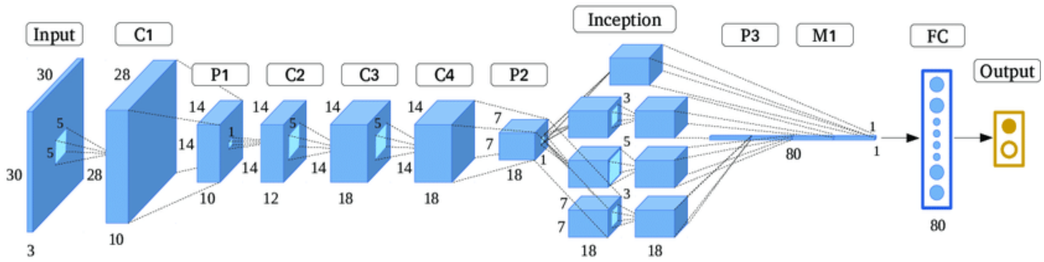

Figure 9. GoogleNet Model.

The model's training effectiveness and the precision of the outcomes depend heavily on the data preparation procedure. For this process, MATLAB provides a number of tools and functions, including data loading, preprocessing, labelling, and data splitting. Gathering data from several sources and transferring it into the MATLAB environment is the first step in the data loading process, which is optimized, particularly for large datasets. Resizing photos, normalizing color channels, and, if required, lowering noise are all done during the preprocessing stage. These procedures can be simply implemented by users thanks to the tools that MATLAB offers to automate them. The labelling step is essential to guaranteeing precise labelling, especially in applications such as disease detection; each image is given a label according to the associated illness or medical condition, and the model uses these labels to train. Lastly, three sets of data are usually separated: training, validation, and testing. To appropriately assess the model's capacity for generalization and to avoid problems like overfitting, this separation is carried out. By offering tools that enable the partition of data according to criteria or at random, MATLAB facilitates the data splitting process. The trainNetwork function in MATLAB is used to train models. The pre-trained weights of a complex model, such as GoogleNet, can be used to begin training. This can greatly improve the model's performance, particularly for extremely specific tasks or when working with limited amounts of data. To help the model better fit the current dataset, data-specific fine-tuning is done. Through this process, the model learns new information in addition to what it already knows and adjusts to certain tasks [21]. Metrics including accuracy, precision, recall, and F1 score are used to assess the model's performance in delicate applications like disease detection. MATLAB facilitates a thorough evaluation of the model and offers tools for computing these measures. These resources provide valuable information about how the model will function in practical settings. In conclusion, MATLAB and GoogleNet-based illness identification systems have the potential to completely transform medical imaging and diagnosis. High accuracy rates can be attained by these systems, making them useful instruments for crucial applications like early illness detection. Each of the critical processes in this process—data preparation, model training, and performance evaluation—directly affects the system's overall performance. The flowchart of the steps is displayed in Figure 10.

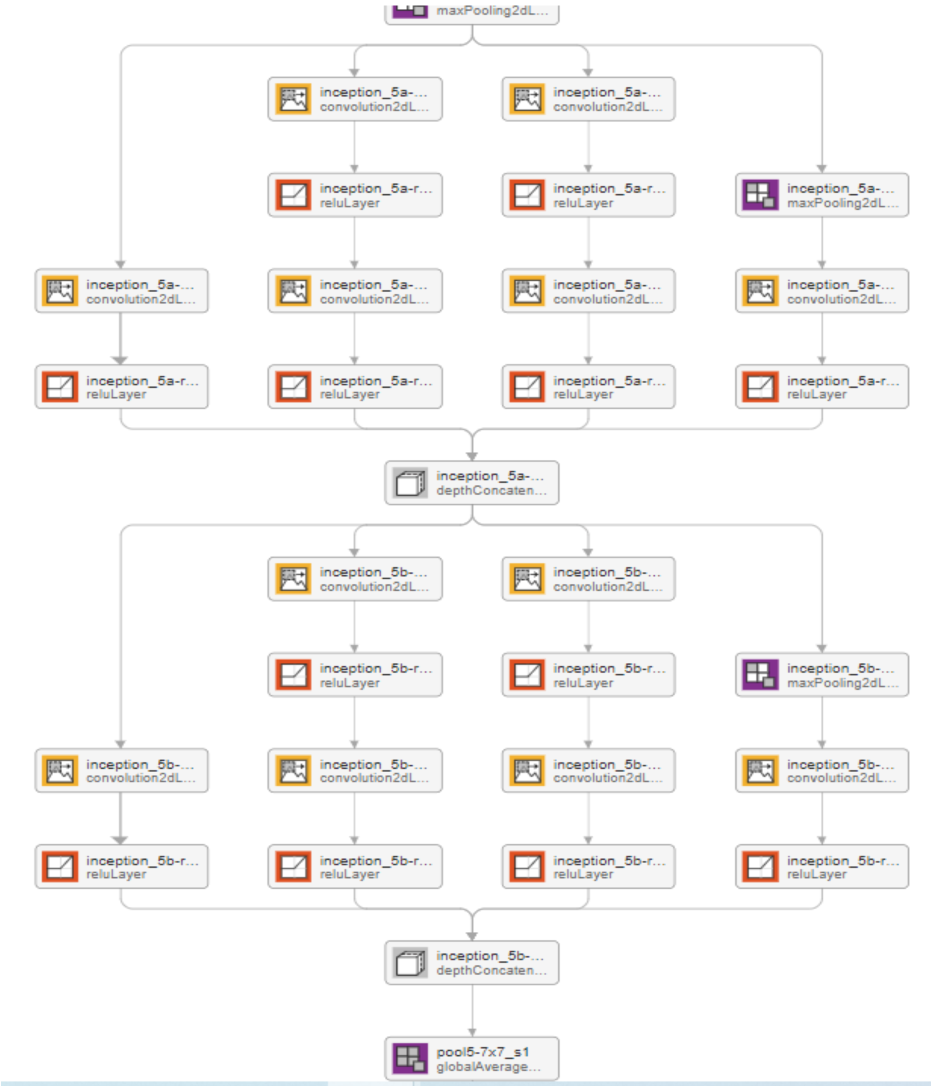

Figure 10. Flowchart of the GoogleNet Algorithm

B. Training the GoogleNet Model

In this study, a pre-existing dataset was not used. The capturing of high-resolution images for analysis and the role of electronics for potential real-time applications in agricultural environments are emphasized within this work. These contributions, combined with the unique dataset created in this study, aim to fill gaps in the literature and contribute to the development of more effective pest management strategies for agriculture. A comprehensive dataset has been created using high-resolution images to detect diseases in tomato plants. The images collected from a tomato greenhouse established in Kirkuk were captured using a Redmi Note 9 Pro AI four-camera system, which includes a 64-megapixel main camera, an 8-megapixel ultra-wide camera, a 2-megapixel depth camera, and a 5-megapixel macro camera. This dataset consists of a total of 2,921 images, encompassing tomato leaves affected by diseases such as downy mildew, early leaf burn, gray mold, and bacterial cancer. Every image has been processed for classification and labelled with the disease kind. The TensorFlow Keras library's ImageDataGenerator class was utilized to implement the data augmentation methods employed in this investigation. By diversifying our picture-based dataset, these methods aim to improve our model's capacity to identify various image variations. The following are the primary methods and parameters used for data augmentation:

- Rotation: To improve the model's recognition of photos collected from various perspectives, images are rotated within a range of +/- 10 degrees.
- Shift: Pictures are arbitrarily moved by 10% in height and 10% in width. This makes it possible for the model to function well with portions of the image that are not complete.
- Shear: To test the model's capacity to identify minute geometric changes in the images, a 10% shear rate was applied.

- Zoom: A zoom in and out of about 20% was used. This improved the model's capacity to identify items of various sizes.
- Flip: To improve the model's recognition of reflection differences, images were randomly flipped horizontally.
- Brightness Adjustment: To make sure the model could adjust to various lighting circumstances, the brightness of the photos was changed from 80% to 120%.

The model was trained using these parameters, and the rates assigned to each transformation technique were carefully selected to optimize the model's ability to react to real-world data circumstances. Regarding the quantity of augmentations, for every original image, roughly five to ten augmented images were created. Our dataset became much more diverse as a result, and the model was able to learn from a wider variety of data during training. Consequently, our model's overall performance has been greatly enhanced by the employed data augmentation strategies, making it capable of performing better in challenging picture identification tasks. Table 1 lists the parameters that were utilized to train the model along with pertinent information about them.

Table 1: Dataset Values

Parameter	value
Dataset Details	Augmented with 4000 data
Solver	SGDM
Initial Learning Rate	0.01
Validation Frequency	50 iterations
Maximum Number of Epochs	30
Mini-Batch Size	128
Execution Environment	Auto
L2 Regularization	0.0001
Gradient Clipping Method	12 norm
Gradient Clipping Value	Inf
Validation Patience	Inf
Shuffle	Every Epoch
Learning Rate Schedule	No
Learning Rate Reduction Factor	No
Learning Rate Reduction Period	No

The study involved the classification of tomato plant disease types. A total of 10,898 datasets were created from a dataset consisting of 2,921 images, which included 4 different diseases and a healthy tomato class. As seen in Table 2, 70% of the datasets were used for training and 30% for testing.

Table 2: Numbers of Data Sets for Tomato Diseases

Disease	Original Data	Augmented Data	Training Data	Test Data
Tomato Mildew Disease	1045	3016	2111	905
Tomato Early Leaf Blight	360	822	575	247
Tomato Lead Mold	651	1898	1329	569
Tomato Bacterial Cancer and Blemish	865	2580	1806	774
Normal Healthy Tomato	862	2582	1807	775
Total	3783	10898	7628	3270

The performance of the GoogleNet model is shown in the Accuracy-Loss Value Graph in Figure 11. The missing value of the model gradually decreases throughout the training process and drops to 0.015 when the training is completed. These results show that the model has successfully learned the second dataset and has the ability to generalize. Figure 11 shows that the accuracy rate is 80.42% and the training time is 1412.21 minutes.

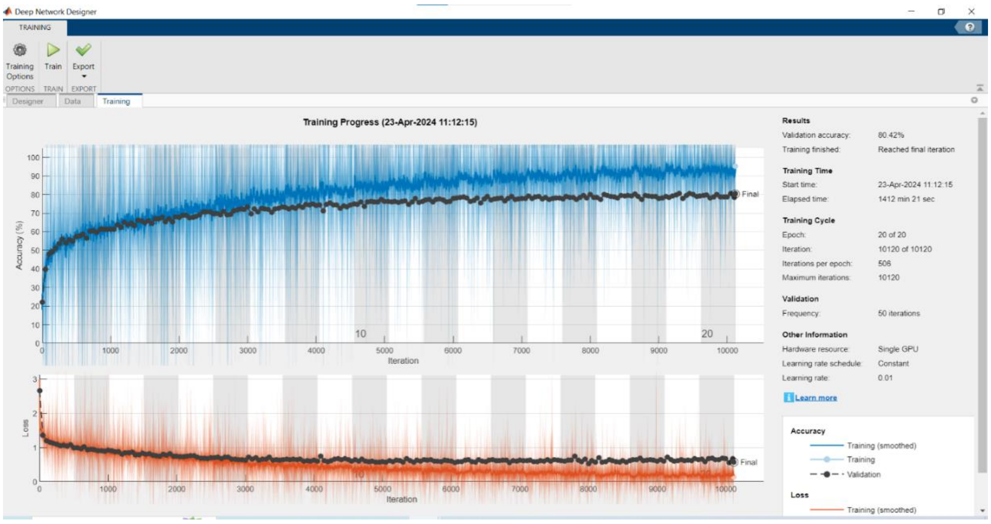

Figure 11. Training Accuracy-Loss Value Graph.

The validation error graph obtained after training the generated dataset is shown in Figure 12.

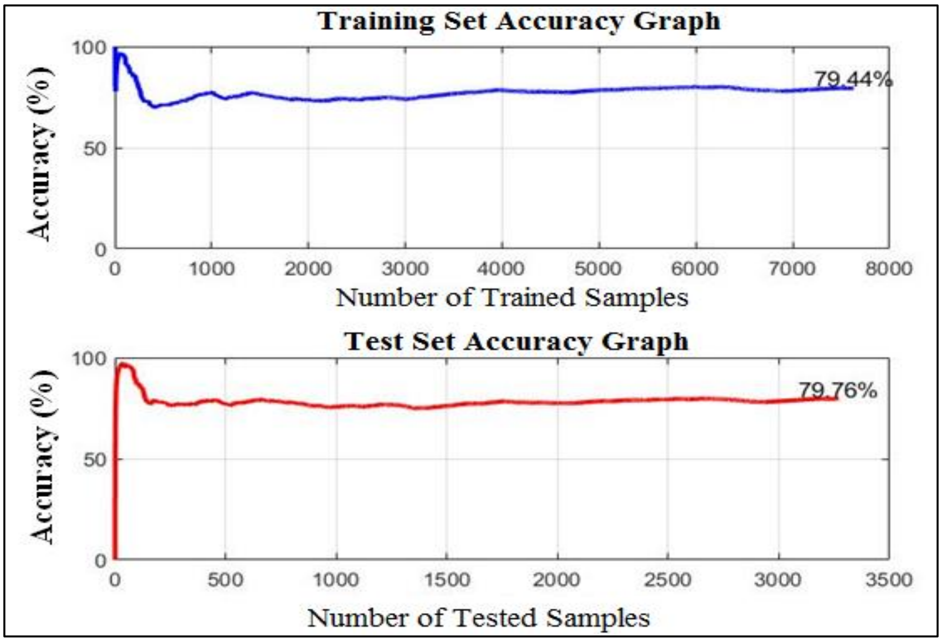

Figure 12. Training and Test Set Validation Graph.

The model's performance on the training set is evaluated by the metrics in the graph displayed in Figure 4.6. Metrics like F1 score, accuracy, and recall demonstrate how effectively the model works for each class. The model's capacity to produce accurate and consistent predictions is demonstrated by its high accuracy, recall, and F1 score. These measures give you insight into the model's overall accuracy as well as how well it performs throughout training. It is also possible to use these metrics to identify which classes the model performs better or worse in. This data offers crucial hints for enhancing and refining the model. The classified patients' training success rates are displayed in Figure 13. As can be seen in Figure 13, Tomato Mild Mildew disease was trained with 81% accuracy, Tomato Early Leaf Blight with 91%, Tomato Lead Blight with 83%, Tomato Bacterial Cancer and Spot Disease with 64%, and Healthy Tomatoes with 87% accuracy.

Figure 13. Training Data Set Classification Values.

The number of correct predictions corresponding to these rates is given in the graph in Figure 14. As seen in Figure 4.7, Tomato Mildew Disease was correctly predicted 1703 data, Tomato Early Leaf Blight 523 data, Tomato Lead Blight 1108 data, Tomato Bacterial Cancer and Spot Disease 1148 data, and Healthy Tomatoes 1578 data

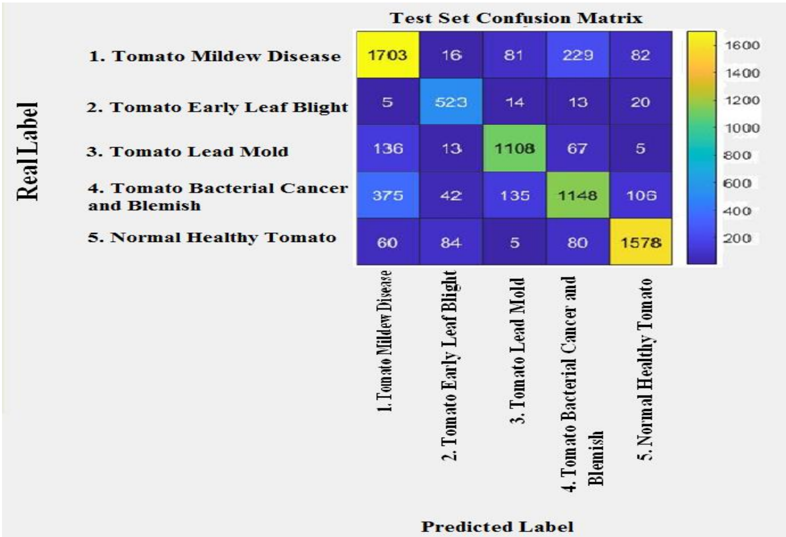

Figure 14. Training Data Comparison Matrix.

As can be seen in Figure 15, Tomato Mildew disease was correctly or incorrectly predicted in 2279 data, Tomato Early Leaf Blight in 678 data, Tomato Lead Blight in 1343 data, Tomato Bacterial Cancer and Spot Disease in 1537 data, and Healthy Tomatoes in 1791 data classifications.

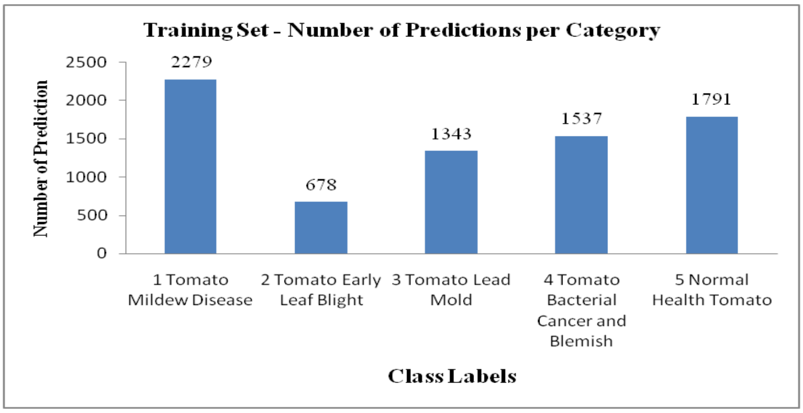

Figure 15. Number of Correct Predictions per Category in the Training Data Set.

Figure 16 shows the success rates of the classified patients in the test. As can be seen in Figure 16, Tomato Mild Mildew disease was tested with 79% accuracy, Tomato Early Leaf Blight with 90%, Tomato Leady Blight with 84%, Tomato Bacterial Cancer and Spot Disease with 66%, and Healthy Tomatoes with 88% accuracy.

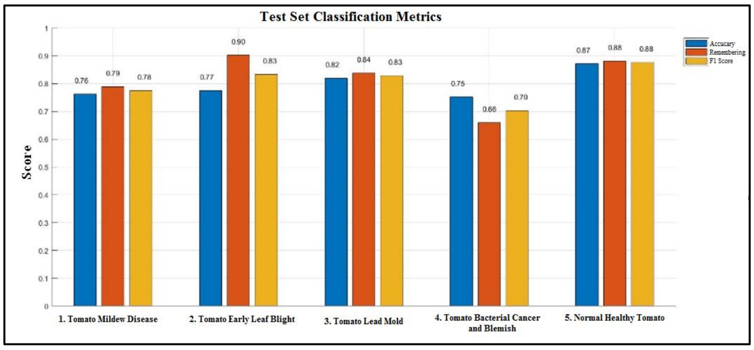

Figure 16. Test Data Set Classification Values.

As can be seen in the graph in Figure 17, Tomato Mildew disease was correctly predicted with 714 data, Tomato Early Leaf Blight with 223 data, Tomato Lead Blight with 477 data, Tomato Bacterial Cancer and Spot Disease with 511 data, and Healthy Tomatoes with 683 data.

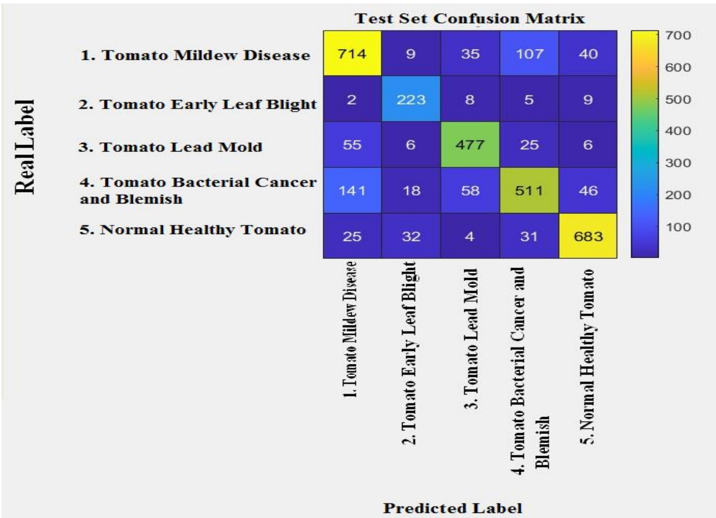

Figure 17. Test Data Comparison Matrix.

As seen in Figure 18, Tomato Mildew Disease was correctly or incorrectly predicted in 937 data, Tomato Early Leaf Blight in 288 data, Tomato Lead Blight in 582 data, Tomato Bacterial Cancer and Spot Disease in 679 data, and Healthy Tomatoes in 784 data classifications.

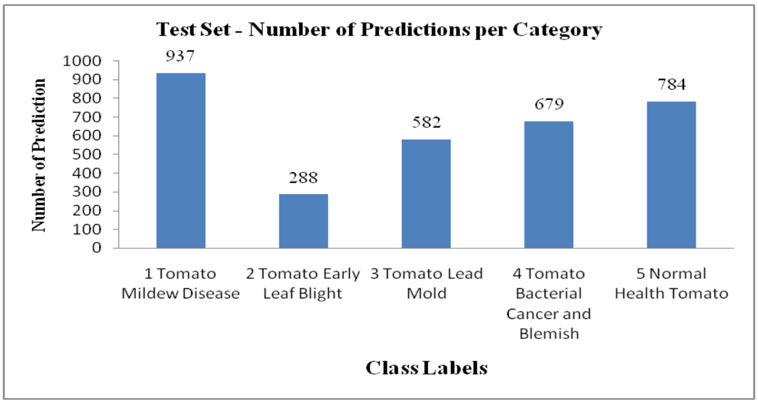

Figure 18. Number of Correct Predictions per Category in Test Data Set.

5. Conclusion

Tomato cultivation is one of the most developed sectors in Turkey. The productivity of tomatoes is of great importance as it is one of the basic foods offered for human consumption. This is why the goal is to identify and categories four prevalent tomato plant diseases. One deep learning method, the GoogleNet model, was used in this work to detect tomato plant illnesses with notable success. The accuracy for tomato mildew disease was 80.67%, early leaf blight was 90.96%, lead mould was 83.37%, bacterial cancer was 63.57%, and healthy tomatoes were 87.33%, based on the training results. The test value percentages of the experimental outcomes of training the GoogleNet model are displayed in Table 3.

Table 3: GoogleNet Model Training Values Percentage Display

	Tomato Mildew Disease	**Tomato Early Leaf Blight**	**Tomato Powdery Mildew**	**Tomato Bacterial Cancer and Spot**	**Normal Healthy Tomato**	
Tomato Mildew Disease	1703 %80.67	16 %0.76	81 %3.84	229 %10.85	82 %3.88	80.67%
Tomato Early Leaf Blight	5 %0.87	523 %90.96	14 %2.43	13 %2.26	20 %3.48	90.96%
Tomato Powdery Mildew	136 %10.23	13 %0.98	1108 %83.37	67 %5.04	5 %0.38	83.37%
Tomato Bacterial Cancer and Spot	375 %20.76	42 %2.33	135 %7.48	1148 %63.57	106 %5.87	63.57%
Normal Healthy Tomato	60 %3.32	84 %4.65	5 %0.28	80 %4.43	1578 %87.33	87.33%

Tomato mildew was 79%, early leaf blight was 90%, lead mould was 84%, bacterial cancer was 66%, and healthy tomatoes were 88%, according to the test results. These outcomes demonstrate how well the GoogleNet model can categories tomato illnesses. The test value percentages of the experimental outcomes of testing the GoogleNet model are displayed in Table 4.

Table 4: GoogleNet Model Test Values Percentage Display

	Tomato Mildew Disease	**Tomato Early Leaf Blight**	**Tomato Powdery Mildew**	**Tomato Bacterial Cancer and Spot**	**Normal Healthy Tomato**	
Tomato Mildew Disease	714 %78.90	9 %0.99	35 %3.87	107 %11.82	40 %4.42	78.90%
Tomato Early Leaf Blight	2 %0.81	223 %90.28	8 %3.24	5 %2.02	9 %3.64	90.28%
Tomato Powdery Mildew	55 %9.67	6 %1.05	477 %83.83	25 %4.39	6 %1.05	83.83%
Tomato Bacterial Cancer and Spot	141 %18.22	18 %2.33	58 %7.49	511 %66.02	46 %5.94	66.02%
Normal Healthy Tomato	25 %3.23	32 %4.13	4 %0.52	31 %4.00	683 %88.13	88.13%

These findings demonstrate the efficacy of technological solutions for early disease detection in agricultural and greenhouse settings. With a bigger and more varied dataset, future research can try to enhance the model's performance. This would not only increase the accuracy but also allow the model to function with various diseases and situations.

Table 5 summarizes the results of the experiments. This study presents the results comparing the performance of three different deep learning models (CNN, SqueezeNet and GoogleNet) for tomato disease diagnosis. As can be seen from Table 5, the GoogleNet algorithm was able to identify Tomato Mildew Disease with 79.37% accuracy, Tomato Early Leaf Blight with 88.33% accuracy, Tomato Lead Blight with 83.87% accuracy, Tomato Bacterial Cancer and Spot Disease with 62.98% accuracy, and Healthy Tomatoes with 90.28% accuracy. This shows that the GoogleNet algorithm gives better results on average for these data compared to other methods. Although the CNN algorithm has low accuracy rates in general average, it has been observed that it gives better results than all algorithms with an accuracy rate of 93.12% for Tomato Mildew disease.

In contrast, the SqueezeNet model obtained an accuracy of 81.6% for healthy tomatoes, 70.83% for tomato early leaf blight, and 84.53% for tomato mildew illness. The results of this study demonstrate that deep learning models can be applied to the creation of automated agricultural disease diagnosis systems. The fact that each model functions differently for various illness types, however, implies that model selection should be done carefully. It is anticipated that employing various model architectures, data augmentation strategies, and hybrid models may yield more fruitful outcomes in subsequent research.

Additionally, farmers may find disease detection more practical and accessible if this model is included into agricultural management systems or mobile applications. Such applications can be quite beneficial, particularly for farmers who reside in rural locations with limited access to cutting-edge technologies. As a result, farmers may identify crop illnesses promptly and take appropriate action to reduce crop losses.

This study concludes by showing the potential and usefulness of deep learning-based disease diagnosis systems in the field of agriculture. Sustainable agricultural practices can spread as a result of more efficient use of technology in agricultural activities, which can enhance production procedures and product quality. This study might be viewed as a significant step that will motivate further research in this regard.

Table 5: Experiment Results

		Disease Categories				
		Tomato Mildew Disease	Tomato Early Leaf Blight	Tomato Powdery Mildew	Tomato Bacterial Cancer and Spot	Normal Healthy Tomato
CNN Model	Training Data Accuracy Percentages	95.12%	5.04%	1.42%	0.49%	26.95%
	Test Data Accuracy Percentages	94.80%	7.29%	1.93%	0.77%	29.03%
SqueezeNet Model	Training Data Accuracy Percentages	84.22	68.52%	53.79%	18.38%	78.97%
	Test Data Accuracy Percentages	84.08%	71.65%	49.03%	19.76%	77.16%
GoogleNet Model	Training Data Accuracy Percentages	80.67%	90.96%	83.37%	63.57%	87.33%
	Test Data Accuracy Percentages	78.90%	90.28%	83.83%	66.02%	88.13%

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Şahinli MA. Comparative advantage of agriculture sector between Turkey and European Union. African Journal of Agricultural Research 2013;8:884–95.

[2] Saygili H, Sahin F, Aysan Y, Mirik M. NEW SYMPTOMS OF TOMATO SOFT ROT DISEASES IN TURKEY. Acta Horticulturae 2005:291–4. https://doi.org/10.17660/actahortic.2005.695.32.

[3] ÖZDEMİR R, DEMİRBAŞ N. Meyve ve sebze üretiminde ortaya çıkan kayıplar üzerinde etkili olan faktörler: İzmir ili örneği. Mediterranean Agricultural Sciences 2020;33:85–91. https://doi.org/10.29136/mediterranean.659011.
[4] Shin J, Chang YK, Heung B, Nguyen-Quang T, Price GW, Al-Mallahi A. A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. Computers and Electronics in Agriculture 2021;183:106042. https://doi.org/10.1016/j.compag.2021.106042.
[5] Zhang S, Wu X, You Z, Zhang L. Leaf image based cucumber disease recognition using sparse representation classification. Computers and Electronics in Agriculture 2017;134:135–41. https://doi.org/10.1016/j.compag.2017.01.014.
[6] Shah JP, Prajapati HB, Dabhi VK. A survey on detection and classification of rice plant diseases. 2016 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC) 2016:1–8. https://doi.org/10.1109/icctac.2016.7567333.
[7] Ghosh S, Singh A, Kavita, Z. Jhanjhi N, Masud M, Aljahdali S. SVM and KNN Based CNN Architectures for Plant Classification. Computers, Materials & Continua 2022;71:4257–74. https://doi.org/10.32604/cmc.2022.023414.
[8] Sibiya M, Sumbwanyambe M. A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks. AgriEngineering 2019;1:119–31. https://doi.org/10.3390/agriengineering1010009.
[9] Soylu EM, Kurt Ş, Soylu S. In vitro and in vivo antifungal activities of the essential oils of various plants against tomato grey mould disease agent Botrytis cinerea. International Journal of Food Microbiology 2010;143:183–9. https://doi.org/10.1016/j.ijfoodmicro.2010.08.015.
[10] Albayrak Ü, Gölcük A, Aktaş S. Agaricus bisporus'ta Görüntü Tabanlı Hastalık Sınıflandırması için Kapsamlı Veri Seti. Journal of Fungus 2024;15:29–42. https://doi.org/10.30708/mantar.1452976.
[11] Yasar A, Golcuk A, Sari OF. Classification of bread wheat varieties with a combination of deep learning approach. European Food Research and Technology 2023;250:181–9. https://doi.org/10.1007/s00217-023-04375-x.
[12] Golcuk A, Yasar A. Classification of bread wheat genotypes by machine learning algorithms. Journal of Food Composition and Analysis 2023;119:105253. https://doi.org/10.1016/j.jfca.2023.105253.
[13] Chaerani R, Voorrips RE. Tomato early blight (Alternaria solani): the pathogen, genetics, and breeding for resistance. Journal of General Plant Pathology 2006;72:335–47. https://doi.org/10.1007/s10327-006-0299-3.
[14] Nowicki M, Foolad MR, Nowakowska M, Kozik EU. Potato and Tomato Late Blight Caused by Phytophthora infestans: An Overview of Pathology and Resistance Breeding. Plant Disease 2012;96:4–17. https://doi.org/10.1094/pdis-05-11-0458.
[15] Çelik I, Özalp R, Çelik N, Polat I, Sülü G. Development of long pepper (Capsicum annuum L.) lines resistant to Tomato spotted wilt virus (TSWV). Acta Horticulturae 2020:37–42. https://doi.org/10.17660/actahortic.2020.1282.7.
[16] Felipe V, Romero AM, Montecchia MS, Vojnov AA, Bianco MI, Yaryura PM. Xanthomonas vesicatoria virulence factors involved in early stages of bacterial spot development in tomato. Plant Pathology 2018;67:1936–43. https://doi.org/10.1111/ppa.12905.
[17] HORUZ S, TİRENG KARUT Ş, AYSAN Y. Domates bakteriyel kanser ve solgunluk hastalığı etmeni Clavibacter michiganensis subsp. michiganensis'in tohumda aranması ve tohum uygulamalarının patojen gelişimine etkisinin belirlenmesi. Tekirdağ Ziraat Fakültesi Dergisi 2019;16:284–96. https://doi.org/10.33462/jotaf.526167.
[18] PINAR AKTEPE B. Domateste bakteriyel benek hastalığının biyolojik mücadelesinde farklı bitki aktivatörleri ve biyolojik preparatların etkisi. Mustafa Kemal Üniversitesi Tarım Bilimleri Dergisi 2021;26:355–64. https://doi.org/10.37908/mkutbd.908921.
[19] Srikanth, M., Mohan, R. J., & Naik, M. C. (2024). Neutrosophic Logic-Based Crop Yield Prediction and Risk Assessment using Least Squares Regression. Neutrosophic Systems with Applications, 23, 1-12. https://doi.org/10.61356/j.nswa.2024.23362
[20] KILIÇARSLAN S, PACAL I. Domates Yapraklarında Hastalık Tespiti İçin Transfer Öğrenme Metotlarının Kullanılması. Mühendislik Bilimleri ve Araştırmaları Dergisi 2023;5:215–22. https://doi.org/10.46387/bjesr.1273729.
[21] Montajab, S. (2023). A View through Artificial Intelligence and Its Relationships with Machine Learning and Deep Learning. Pure Mathematics for Theoretical Computer Science, 08-17. DOI: https://doi.org/10.54216/PMTCS.020101
[22] Ertem S, Özbay E. Sınıflandırma Probleminde Derin Özellik Birleştirme Yaklaşımıyla Domates Yaprağı Görüntülerinde Hastalık Tespiti. Avrupa Bilim ve Teknoloji Dergisi 2022:84–92.
[23] Metlek S, Çetiner H. Matlab Ortamında Derin Öğrenme Uygulamaları. Ankara/Turkey 2021.