

**Research Article****Early Diagnosis of Tomato Diseases Using Machine Learning to Reduce Pesticide Application****Nasir Cetin** <sup>a,\*</sup> , **Ziya Ozcelik** <sup>b,</sup> , **Yavuz Selim Taspinar** <sup>c,</sup> <sup>a</sup> *Integra Hydraulic, Konya, Türkiye.*<sup>b</sup> *Selçuk University, Faculty of Technology, Department of Mechatronic Engineering, Konya, Türkiye.*<sup>c</sup> *Selçuk University, Faculty of Technology, Department of Mechatronic Engineering, Konya, Türkiye.*

## ARTICLE INFO

*Article history:*

Received 26 December 2025

Accepted 30 March 2026

*Keywords:*Classification,  
Machine learning,  
Organic farming,  
Pesticides,  
Tomato diseases

## ABSTRACT

Pesticides, which are widely used today to increase yield and prevent diseases in agricultural products, have negative effects on human health and the environment. Due to these effects, efforts to restrict pesticide use gained momentum with environmental protection movements starting in the 1970s. Strict controls and licensing processes have been introduced for pesticide production to prevent their unconscious use. The aim of this study is to classify diseases observed in tomato plants, thereby preparing a scientific groundwork for reducing pesticide use. For this purpose, a dataset comprising a total of 13,883 records of tomato plants belonging to six different classes was utilized. For the classification of the data, the following Machine Learning methods were employed: Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Logistic Regression (LR) algorithms. Model performance was evaluated using metrics such as accuracy, precision, recall, and F1 Score, along with the confusion matrix and ROC analysis. The classification success rates obtained from model training and testing were determined as: ANN 100.0%, KNN 99.8%, SVM 99.9%, and LR 99.9%. The results indicate that all models achieved high and very similar success rates. In conclusion, it is expected that all classification models used can be successfully utilized for the reliable diagnosis of tomato plant diseases, and this capability is projected to make significant contributions to organic farming applications.

This is an open access article under the CC BY-SA 4.0 license.  
(<https://creativecommons.org/licenses/by-sa/4.0/>)

**1. Introduction**

Pesticides are chemical substances or mixtures of substances used to control or prevent damage from insects, weeds, microorganisms and pests that damage agricultural products or animal food products during production, harvesting or display [1]. Pesticides are not applied directly to plants or soil. Because they are toxic substances, they are used by mixing them with some auxiliary substances (solid, liquid) in a way that minimizes harm to human and environmental health against pests. The pesticide in this mixture is called the "active substance" or "inactive ingredient". Not every toxic substance can be used and named as a pesticide. For a toxic

substance to be considered a pesticide, it must be biologically active, effective, sufficiently stable, safe for users, third parties and consumers, safe for livestock, not harmful to wildlife and beneficial organisms, acceptable for the environment and have characteristics that will not cause problems in trade. Many studies are being conducted on this subject in the literature. Some of those related to this study are given in order [1].

In one study, Tiryaki et al. conducted the necessary research by investigating the benefits and harms of pesticides using data on average pesticide consumption per unit area in some European Union (EU) countries. Their contribution to the literature was to compare pesticide use in Turkey, EU countries, and the world, explaining the

\* Corresponding author. E-mail address: [nasr.cetin@gmail.com](mailto:nasr.cetin@gmail.com)  
DOI: 10.58190/ijamec.2026.161

advantages and disadvantages of pesticide use and the behavior of pesticides in the agricultural ecosystem. They also addressed issues such as the safe and effective application of pesticides, possible residues in products, and factors affecting these residues. In addition, they compared the situation of countries that received residue alerts published in the EU rapid alert system [2].

Özcan and Tongur analyzed the toxic effects of insecticides and herbicides using the *Lepidium sativum* toxicity test method, determined that herbicides were more toxic than insecticides, and found that the harm of some banned pesticides reached high levels [3].

In their survey study with rose oil producers in Isparta, Arıcı et al. observed that a significant portion of the producers did not adequately consider the environmental risks when applying pesticides, but a large portion of them complied with the waiting periods for the pesticides [4].

In one of their studies, Yüzbaşıoğlu and Topkaya conducted a study on pesticide use with 95 plant producers randomly selected from villages in the central district of Tokat province. As a result of the interviews, it was observed that the producers were within the working population and had not received any education beyond compulsory education. It was determined that the producers were influenced by pesticide dealers in their pesticide use. It can be said that they did not take any protective measures during agricultural spraying. It was observed that the producers, who thought that pesticide use did not harm the product, spent very high prices financially. It was observed that they were unaware of the environmental damage caused by the disposal of the waste from the pesticides they used. In conclusion, it can be said that the producers have very little knowledge about pesticide use and its aftermath. As a contribution to the literature, they determined that there is a need for relevant institutions and organizations to organize seminars on pesticide application and post-application information [5].

Polat stated that global pesticide use has reached 3.5 billion kg annually in the last 20 years, and in Türkiye, use increased by 11% compared to the previous year, reaching 60 thousand tons. The study emphasized that chemical control alone is not sustainable due to risks such as environmental toxicity and resistance development. In conclusion, the author argues that pesticides are necessary but should be applied in a controlled manner, integrated with alternative control methods, in order to prevent ecological damage [6].

In a study, Özeran and Taşçı examined pesticide usage amounts in Turkey based on provinces, regions, and pesticide groups, using the most recent data available for 2020. Based on this data, the province with the highest pesticide usage in Turkey is Antalya, while the province with the lowest usage is Ardahan. Pesticides are used most in the Mediterranean Region due to the diversity of crops grown there, while the region with the lowest use is the

Black Sea Region. The most widely used pesticide groups in Turkey in 2020 were fungicides (38.4%), herbicides (27.4%), and insecticides (23.0%), respectively. When examining the top three provinces with the most intensive pesticide use, it was found that the most widely used pesticide group was fungicides in Antalya (38.3%) and Manisa (81.0%), and insecticides in Adana (65.2%). Chemical control is the most preferred agricultural control method in Turkey for combating diseases, pests, and weeds to ensure sustainable production. Although pesticide use seems inevitable, it has been stated that raising producer awareness and promoting integrated control would have positive effects on human and environmental health by reducing usage [7].

In a study, Yılmaz et al. aimed to analyze changes in pesticide use by utilizing data from the United Nations Food and Agriculture Organization (FAO) and data on pesticides and pesticide groups used worldwide, in EU countries, and in Turkey between 2000 and 2021, as found in articles and reports on the subject. According to the findings of the study, as of 2021, the global average pesticide use was 1.07 kg ha<sup>-1</sup>, the EU average was 3.20 kg ha<sup>-1</sup>, and the Turkish average was 2.26 kg ha<sup>-1</sup>. Accordingly, the amount of pesticides used per hectare in Turkey is lower than the EU average but higher than the global average. While herbicides have the highest share of pesticide use per unit area worldwide (21.92%), fungicides and bactericides have the highest share in the EU (43.18%) and Turkey (36.06%). While the European Union has implemented comprehensive legal regulations to reduce pesticide use and ensure environmental sustainability, regulations on pesticide use in Turkey remain limited. Despite the inadequacy of legal regulations on pesticide use in Turkey, it can be said that there have been significant efforts in recent years to promote sustainable agricultural practices. According to the literature, Turkey should develop more specific legal regulations, policies, and strategies that reduce the risks associated with pesticide use, taking into account the EU harmonization process, and implement the EU's structurally focused and comprehensive legal regulations. They have contributed by suggesting the implementation of more effective agricultural environmental policy measures, such as taxation and subsidies, to develop alternative disease and pest control methods and ensure farmer adaptation for sustainable agricultural production in Turkey [8].

In a study, Ayyıldız utilized FMOLS-DOLS models to determine the long-term effects of pesticides using chemical pesticide, environmental, and economic indicator data from 1990 to 2021. According to the findings, fungicide and herbicide use had a significant effect on the crop production index and agricultural greenhouse gas emissions, while insecticide use was found to be statistically insignificant. The results suggest that pesticide use contributes significantly to productivity but

causes increasing environmental damage over time. In the literature, the author considers it important to establish legal regulations aimed at reducing pesticide use and ensuring its correct use, and to activate publication activities. He has contributed by stating that priority should be given to R&D studies for the widespread use and improved effectiveness of biopesticides as an alternative to chemical pesticide use [9].

A study by Tıraş found that as of 2020, the top six countries with the highest agricultural production globally (China, India, Brazil, the United States, Indonesia, and Russia) using panel data from 1995 to 2019, estimated the effect of fertilizer and pesticide use on agricultural Gross Domestic Product (GDP) using panel econometric methods. It stated that the effect of fertilizer use on agricultural GDP was significant, but the effect of pesticide use was statistically insignificant [10].

Doğan and Karpuzcu simulated the effects of structural (filter strips) and non-structural (use reduction) Best Management Practices (BMPs) on pesticide load in a water basin in Istanbul using a SWAT model. In the study, it was predicted that 25-meter-wide filter strips provided 95% retention and that the basin would be cleaned in 3-4 years if pesticide use were completely eliminated. The authors emphasized that mathematical models play a critical role in decision support processes, but the model sensitivity should be increased with additional parameters (Koc, land slope, etc.) [11].

In a study, Aydoğan and Baran used data obtained from surveys of 94 tobacco farmers in the Çelikhan district of Adıyaman province and determined that 44.0% of farmers had a low level of awareness of agricultural pest control, 45.0% had a medium level, and 11.0% had a high level. Furthermore, differences were observed in the awareness levels of farmers before, during, and after application, and it was found that farmers adhered more strictly to the rules before agricultural pest control. The results were used in the field of agricultural activities. Literature In the agricultural education and extension activities planned in the region, supporting educational content that conveys the negative effects of agricultural pesticides to farmers with visuals and strictly monitoring prescription-based pesticide use may help solve the problem. On the other hand, the direct consumption of Çelikhan tobacco leaves has contributed to the emergence of the residue problem and the need for further studies on pesticide residues in tobacco and their effects on human health [12].

In a study, Hayran et al. used data collected from 68 randomly selected farmers to evaluate pesticide use, analyze farmers' risk perception regarding pesticide use, and identify its determinants using descriptive statistics and logistic regression analysis methods. Based on these methods, it was determined that farmers' risk perceptions were at a moderate level and that these perceptions were influenced by variables such as education, experience,

internet use, non-agricultural income, and greenhouse area. This is one of the limited studies in the literature examining greenhouse farmers' risk perceptions regarding pesticide use, and the authors have contributed to this field [13].

In one study, Doğan and Karaaslan examined the production practices of grain producers benefiting from the ÇATAK program in the Seyfe Gölü region of Kırşehir province from an environmental perspective. They found that producers relied on their experience in the use of chemical fertilizers and pesticides and focused on cost rather than environmental risk, recommending that local actors keep their information activities dynamic [14].

In a study, Tözün and Akar conducted a systematic review of articles published in the national literature in Turkey and found a total of 35 records that met the inclusion criteria using search data from Izmir Katip Çelebi University Electronic Library resources and Google Scholar. It was determined that 16 (45.7%) of these contained pesticide residues and exceeded the Maximum Residue Levels (MRL); 10 (28.6%) contained pesticide residues but did not exceed the MRL values. Although the food samples studied varied, pesticide residues were studied extensively and found in grapes, strawberries, citrus fruits, and some commonly consumed vegetables. Milk and honey are notable among the foods in which no pesticide residues were found. They contributed to the literature with recommendations such as reviewing the legislation on the subject and introducing deterrent penalties, as well as increasing educational efforts on the subject for public and occupational health [15].

In a study, Gedikli et al. 2011 used a simple random sampling method in the districts of Alaçam, Bafra, and Terme in the province of Samsun during the 2011 production year to survey 99 farmers in 15 villages regarding their environmental awareness in the use of agricultural pesticides in rice, corn, and wheat production. In addition to these data, data from local and foreign literature and state institutes and organizations were used. According to the research results, although agricultural institutions are not at a sufficient level, they are visited by farmers. Farmers consult with pesticide dealers and agricultural engineers in the selection of chemical pesticides (48.8%) and dosage adjustment (33.34%), and they obtain pesticides when they start to see damage (70.70%). In addition, 67.67% of producers dispose of pesticide packaging by burning it after spraying, 26.26% bury it in the ground, and 6.06% dispose of it in the trash by bagging it. The literature on harmful species (disease, pest, and weed) and their environmental relationships, and ensuring that the population of harmful species is kept below the level of economic damage. Thus, by researching and applying alternative methods to chemical control, the approach of minimizing pesticide use will be adopted, thereby contributing to the significant protection of both

the natural environment and human health [16].

The primary objective of this study is to diagnose diseases and weeds affecting tomato plants with high accuracy and at an early stage using deep learning architectures and machine learning algorithms as an alternative to traditional methods. The rapid and accurate detection of plant diseases aims to reduce environmental impacts and maintain production efficiency by enabling the application of pesticides only when necessary and in the correct amounts.

The contributions of this study to the literature and practice can be summarized as follows:

- **Early Diagnosis with Deep Learning:** Using the SqueezeNet deep learning architecture, early diagnosis was made possible through visual data containing healthy and diseased states of tomato plants.

- **Multi-Classification Comparison:** Classical machine learning algorithms such as Artificial Neural Networks (ANN), K-Nearest Neighbor (KNN), Support Vector Machines (SVM), and Logistic Regression (LR) were systematically applied on deep features obtained from SqueezeNet, and their performances were compared. This approach demonstrates the potential for integrating transfer learning with classical ML.

- **Application on a Large Dataset:** The study evaluated the performance of multi-class detection using a six-class (diseased and healthy) tomato leaf dataset consisting of 13,883 different photographs.

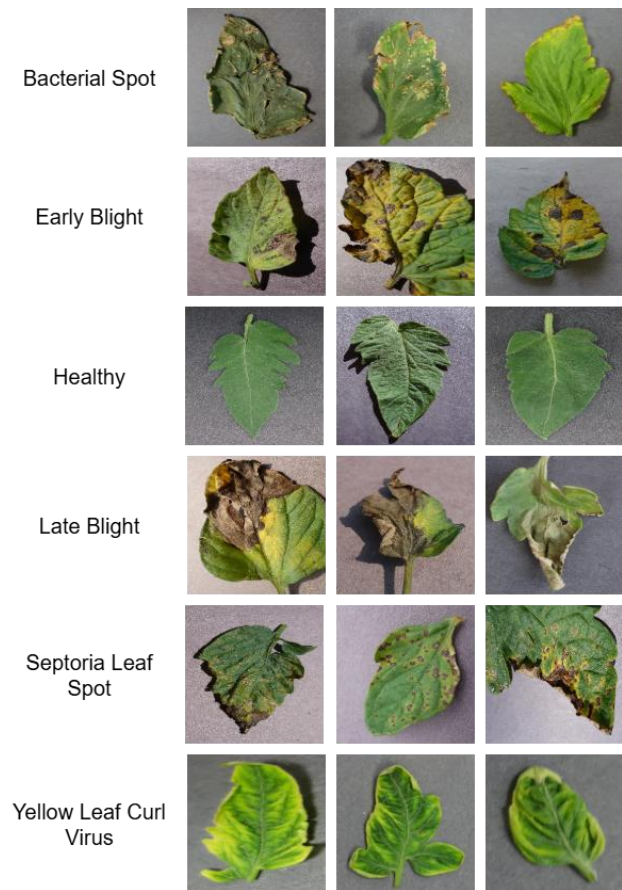
- **Weed Segregation:** By detecting healthy plants, the study indirectly demonstrates the potential for segregating harmful weeds growing alongside them, paving the way for precision farming applications.

The remainder of the study is organized as follows: the second section, the material and methods section, presents the dataset used in the study, the methods employed, and the performance metrics. The third section presents the experimental results obtained from the study and the discussions. The fourth section presents the results, contributions, limitations, and areas of application of the study.

## 2. Material and Methods

### 2.1. Crop Disease Detection Dataset (CDDD)

The dataset used in the study contains a total of 13,883 images across 6 classes. There are 2,127 images in the Bacterial Spot class, 2,400 in the Early Blight class, 2,407 in the Healthy class, 2,314 in the Late Blight class, 2,184 in the Septoria Leaf Spot class, and 2,451 in the Yellow Leaf Curl Virus class. Each image is 256x256 pixels in size. Figure 1 shows sample images from the dataset [17].



**Figure 1.** Sample images by class in CDDD

Bacterial Spot disease begins as water-soaked spots; the spots increase in number and may merge to form necrotic areas on the leaves, giving them a burnt appearance. Leaf drop can cause severe defoliation and make the fruit susceptible to sunscald. Mature spots have an oily appearance and may appear transparent when held up to the light. The centers of the spots dry out and fall off the leaf; burnt leaves usually remain attached to the plant and give it a burnt appearance; fruit infections start as slightly raised blisters; spots may have a slight halo that disappears over time; lesions on the fruit may have a raised edge and a sunken center, giving the fruit a pitted appearance [18].

Early Blight disease symptoms begin as oval-shaped lesions with a yellow chlorotic zone along the lesions; concentric leaf remnants may be seen on fused leaves; leaf structure between veins is provided; severe blight can cause complete collapse of the leaves; the disease severely burns the leaves and leads to loss of yield; tomato stems can become infected with the fungus and lead to Alternaria stem canker; the first signs of stem cankers are changes of dark brown areas on the stem; stem cankers can grow to encompass the entire stem and cause the entire main center; brown lines may be present on top of the canker and below the veins; fruit symptoms include small black v-shaped lesions on the shoulders of the fruit (the disease is also called black shoulder); lesions are also resistant as dark spots with a concentric ring pattern on the fruits; fruit pieces may be visible in the field or may develop during

the presentation of the fruit to the market; lesions may have a velvety appearance due to sporulation of the fungus [18].

Late Blight disease affects all above-ground parts of the tomato plant; the first signs of the disease appear as waterlogged green to black areas on the leaves and rapidly turn into brown lesions; in wet weather, fluffy white fungal growth may be seen in infected areas and on the undersides of the leaves; as the disease progresses, the leaves shrivel and turn brown and the whole plant may die; fruit lesions begin as irregularly shaped waterlogged areas and turn into oily spots; the entire fruit may become infected and a white fuzzy growth may be seen in wet weather [18].

Symptoms of Septoria Leaf Spot disease can appear at any stage of tomato development and may begin as small, water-soaked spots or circular gray-white spots on the underside of older leaves; the spots have a gray center and a dark edge and may sink inwards; fungal fruiting bodies may be visible as small black dots in the center of the spot; spots may also be seen on stems, fruit calyxes and flowers [18].

In Yellow Leaf Curl Virus disease symptoms, infected leaves shrink, curl upwards, appear wrinkled, and yellowing is seen in the veins and leaf edges. Internodes shorten and the whole plant appears stunted and bushy. The whole plant stands upright and only upright growth is observed. Flowers may not develop and may fall off [18].

## 2.2. Convolutional Neural Network (CNN)

Deep learning is a subfield of artificial intelligence. It involves developing artificial neural networks to perform mathematical and statistical calculations and process data related to a problem. The processed data is then transformed into predictions or results at the final layer of the neural network. The success of the prediction is measured by evaluating these results using various metrics. Training artificial neural networks is also used as model training. The model training process varies depending on the type of problem. In this article, the focus will be on image-based problems, along with the field we call computer vision. Algorithms and architectures consisting of artificial neural networks used in solving image-based problems are referred to as Convolutional Neural Networks in the literature. Convolutional neural networks expect an image as input. The image given to the neural network is processed through layers according to the architecture of that neural network. Since the expected output of the neural network is a group of diseases, it can be said that classification is done for different diseases and that this problem is a classification problem [19].

## 2.3. SqueezeNet Architecture

Another popular architecture used in the CNN model is the SqueezeNet architecture. The SqueezeNet architecture was introduced by Iandola et al. in 2016. The goal of this architecture is to create a neural network with fewer

parameters, and the architecture achieves AlexNet-level accuracy with 50 times fewer parameters. The advantage of the SqueezeNet architecture is that the workload on the neural network is reduced thanks to more efficient distributed layers, allowing it to run faster.

This architecture was proposed in 2016 by researchers at DeepScale, the University of California, Berkeley, and Stanford University. Its advantages are:

It requires less communication between servers during distributed training.

Requires less bandwidth to export a new model from the cloud.

Is more suitable for deployment on specialized hardware with limited memory. Large activation maps provide higher classification accuracy when given the same number of parameters.

The authors summarize three main strategies for reducing parameter size while maximizing accuracy:

Strategy 1: Make the network smaller by replacing 3x3 filters with 1x1 filters. A 1x1 filter has 9 times fewer parameters than a 3x3 filter.

Strategy 2: Reduce the number of inputs for the remaining 3x3 filters. Fewer inputs to convolutional layers result in fewer parameters.

Strategy 3: Sampling towards the end of the network to ensure the folding layers have large activation maps. This maximizes the use of fewer parameters and achieves the highest accuracy. Delaying subsampling towards the end of the network creates larger activation/feature maps [20].

## 2.4. Artificial Neural Network (ANN)

The ANN is one of the machine learning methods inspired by biological neural systems. ANNs consist of artificial neurons arranged in layers, and each neuron weights input data and converts it into output through activation functions. This structure allows learning complex nonlinear relationships and achieves high performance in many problems such as classification, regression, and pattern recognition [21].

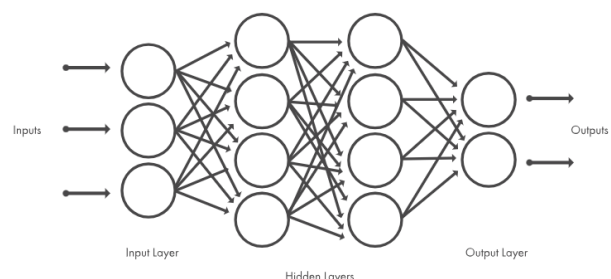


Figure 2. Typical artificial neural network architecture.

## 2.5. K-Nearest Neighbour (KNN)

The KNN algorithm is a non-parametric, supervised learning classification that makes classifications or predictions about the grouping of individual data points using proximity. It is one of the most popular and simplest

classification and regression algorithms used in machine learning today. While the KNN algorithm can be used in both regression and classification problems, it is most commonly used as a classification algorithm and can assume that similar points are likely to be found close together. Regression problems use a similar concept to the classification problem, but in this case, the average of the  $k$  nearest neighbors is taken to make a prediction about a classification. The key difference here is that classification is used for discrete values, while regression is used for continuous values. However, to perform a classification, the distance must first be defined. These distances can be measured in two ways.

**Euclidean distance (p=2):** This is the most commonly used distance measure and is limited to real-valued vectors. Using Formula 1, a straight line is measured between the inquiry point and the other point being measured.

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (1)$$

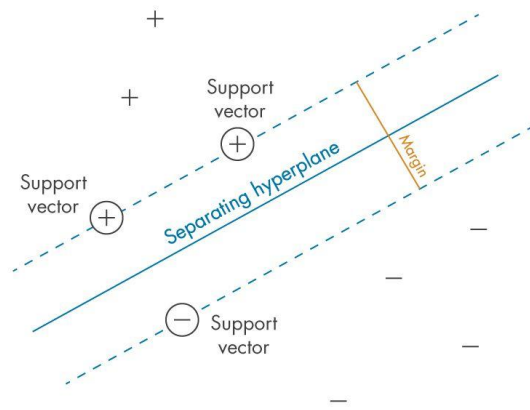
**Manhattan distance (p=1):** This is a popular distance measurement that measures the absolute value between two points. Often visualized as a grid, it's also called taxi distance or city block distance, and Formula 2 shows how to get from one address to another via city streets.

$$d(x, y) = \sum_{i=1}^m |x_i - y_i| \quad (2)$$

K-Nearest Neighbor (KNN) applications determine the age (or health) of a new, variable-status tomato plant specimen by calculating the similarity of the feature to labeled specimens shown in your dataset. The algorithm plots the features of the new leaf under examination into the dataspace and identifies "K" neighboring leaves that are most similar to it using persistent distance measurements; it then assigns the new specimen to the class of condition (healthy, early blight, septoria leaf spot, etc.) that is most prevalent among these neighbors. Instead of training a complex weighting model in the background, this method, which directly references probability indicators, physical or pixel-based variation options in your dataset, provides data-driven and direct examples in the diagnosis of tomato diseases.

## 2.6. Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm frequently used for classification and regression problems in applications such as signal processing, natural language processing (NLP), speech, and image recognition. The goal of the SVM algorithm is to find a hyperplane that best separates data points of one class from data points of another class. This hyperplane can be a line for a 2-dimensional space or a plane for an  $n$ -dimensional space, where  $n$  is the number of features for each observation in the dataset. There can be more than one hyperplane separating classes in the data. The optimal hyperplane derived by the SVM algorithm is the one that maximizes the margin between the two classes.



**Figure 3.** An SVM algorithm to maximize the width of the margin between two classes, represented by plus and minus signs and separated by a hyperplane.

SVM aims to find the ideal decision surface (hyperplane) that defines the boundary between classes by plotting the features obtained from leaves (color change, texture distortions, spot size, etc.) in a multidimensional space. Especially in cases where disease symptoms are complex and cannot be linearly separated by a simple line, it uses "kernel functions" to move the data to higher dimensions, thus mathematically drawing the most precise distinction between early blight, leaf spot, or healthy states; this allows the algorithm to achieve high accuracy rates and strong generalization capabilities in the early diagnosis of diseases when a new tomato leaf image is presented to it.

## 2.7. Logistic Regression (LR)

Logistic regression learns by providing the system with the necessary class information corresponding to the input values related to the problem. The model obtained as a result of this process can make meaningful predictions for any input value that it has not encountered before and for which it does not have class information, thanks to its generalization ability. When modeling, it aims to find the most suitable function in order to reach the closest possible result. Because the error returns to the system as learning, the process is much faster than unsupervised learning [22]. Logistic regression is a method that has the logic of creating a relationship between the dependent variable and the independent variable over a small number of variables. The most important feature that distinguishes logistic regression from linear regression is that modeling can be achieved without the need for criteria such as the normal distribution, constant variance, and linearity of the variable [22].

## 2.8. Confusion Matrix

A confusion matrix is a visualization method for classifier algorithm results. More specifically, it is a table that divides the number of actual instances of a given class by the number of predicted instances of that class. Confusion matrices are one of several evaluation metrics

that measure the performance of a classification model. They can be used to calculate many other model performance metrics such as accuracy and recall. Confusion matrices can be used with any classifier algorithm, such as logistic regression models, decision trees, etc.

In a confusion matrix, the columns represent the predicted values of a given class, and the rows represent the actual values of a given class, or vice versa. This grid structure is a useful tool for visualizing model classification accuracy by showing the number of correct and incorrect predictions side-by-side for all classes.

		PREDICTED	
		Tomato - Bacterial Spot	Tomato - Healthy
ACTUAL	Tomato - Bacterial Spot	TP	FP
	Tomato - Healthy	FN	TN

Figure 5. An example of a Confusion Matrix.

## 2.9. Performance Metrics

### 2.9.1. Accuracy

It shows the percentage of all predictions made by the model that are correct. It is the simplest and most common metric.

True Positive (TP) is the case where samples that are actually positive are correctly predicted as positive. True Negative (TN) is the case where samples that are actually negative are correctly predicted as negative. False Positive (FP) is the case where samples that are actually negative are incorrectly predicted as positive. False Negative (FN) indicates the case where samples that are actually positive are incorrectly predicted as negative.

The accuracy metric evaluates the probability of the model making the correct prediction by considering these four cases. However, accuracy alone may be insufficient in situations such as unbalanced datasets or cost-focused problems. Therefore, it is more appropriate to use it in conjunction with other metrics to evaluate the model's performance in more detail.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

All models demonstrated exceptional overall performance, achieving an accuracy rate above 0.99. In particular, the Artificial Neural Network (ANN) model achieved an accuracy of 1.000 (close to 100%), proving its near-perfect prediction of the leaf condition in the tested dataset. These high accuracy rates clearly demonstrate that the developed system can be used as a reliable early

detection tool in greenhouses or fields.

### 2.9.2. F1 Score

The F1 score is the harmonic mean of the Precision and Recall metrics. It provides a better measure than Accuracy, especially in unbalanced datasets, because it takes both metrics into account in a balanced way. This is shown in Formula 4.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

The F1 score is proof that the model does not overweight one metric while neglecting the other. The near-perfect F1 scores of 0.996 for ANN and 0.994 for SVM confirm that the models successfully detect diseased leaves with over 99% accuracy and avoid false alarms for healthy leaves. The ANN and SVM models provided the most optimal balance between the two main objectives of the study: "early detection" and "chemical minimization".

### 2.9.3. Precision

Sensitivity indicates how many of the cases that the model labels as positive are actually positive. This is quite important in situations where the cost of false positives (FP) is high. It is shown in Formula 5.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

If a model has low sensitivity, it can cause unnecessary chemical spraying by labeling healthy plants as "sick" (False Positive). Examining the table, the lowest sensitivity is found in the KNN algorithm with 0.952; meaning that approximately 4.8% of the "diseased" alarms given by the KNN model are actually for healthy leaves. This can lead to unnecessary spraying. In contrast, the ANN (0.986) and SVM (0.979) models minimize the "false alarm" rate with very high sensitivity values. In this way, chemical pesticide use can be significantly reduced by targeting only truly diseased tomato plants, thus achieving the goal of environmentally friendly agriculture (sustainability).

### 2.9.4. Recall

It shows the percentage of actually positive cases that were correctly identified by the model. It can be used in situations where the cost of false negatives (FN) is high. This is shown in Formula 6.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

The greatest danger in agricultural diseases is that a diseased plant is overlooked and mistaken for a "healthy" plant (False Negative), leading to the spread of the disease throughout the greenhouse. Looking at the sensitivity values of the models, ANN (0.993) yielded the most successful result. Interestingly, the KNN model, which is weakest in precision, is quite successful in sensitivity (0.990); that is, it works well in not missing diseases, but in doing so it can be a little too "cautious" and mistake healthy plants for diseased

ones. These high recall values obtained to prevent the spread of disease demonstrate that the system has a very strong early diagnosis capability.

### 3. Results And Discussion

This study aimed to classify different tomato plants by detecting disease symptoms using image processing and machine learning methods, employing ANN, KNN, SVM, and LR techniques with the Crop Disease Detection dataset. A computer with a Core i5-7200U 2.50 GHz processor, Intel(R) HD Graphics 620 graphics card, and 8 GB of RAM was used. The Python programming language was employed. For the ANN method, the following parameters were used: 100 hidden layers, ReLu activation function, regularization 0.0001, and number of iterations 200. For the KNN method, the following parameters were used: number of neighbors 5, metric euclidean, and weight uniform. For the SVM method, the following parameters were used: cost 1, regression loss epsilon 0.1, kernel RBF, and iteration 100. For the LR method, the regularization type Ridge L2 parameter was used. Training and testing were performed using cross-validation, with a k value of 10. The processes performed in this study are shown in Figure 6.

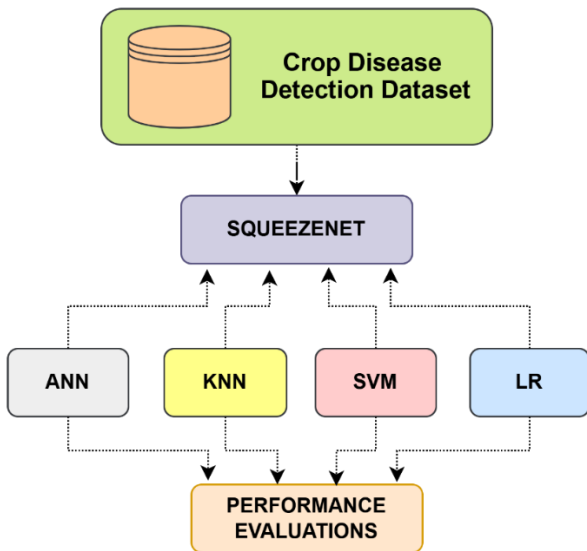


Figure 6. Flowchart of the steps performed in the study.

As an introduction to machine learning methods, the SqueezeNet model was used to extract approximately 6 features from each image, and training was performed. The confusion matrices obtained from the machine learning methods are given below. The confusion matrix obtained from the ANN method is

shown in Figure 7.

		PREDICTED					
		Tomato - Bacterial Spot	Tomato - Early Blight	Tomato - Healthy	Tomato - Late Blight	Tomato - Septoria Leaf Spot	Tomato - Yellow Leaf Curl Virus
ACTUAL	Tomato - Bacterial Spot	2068	18	0	4	18	19
	Tomato - Early Blight	31	2229	15	60	46	19
	Tomato - Healthy	0	6	2390	4	7	0
	Tomato - Late Blight	4	68	11	2199	26	6
	Tomato - Septoria Leaf Spot	18	50	8	19	2089	0
	Tomato - Yellow Leaf Curl Virus	19	20	1	4	1	2406

Figure 7. Confusion Matrix of the ANN Model

According to Figure 7, the ANN method correctly classified 2068 data points belonging to the Tomato-Bacterial Spot category. It classified 18 data points belonging to the Tomato-Bacterial Spot category as Tomato-Early Blight. It classified 4 data points belonging to the Tomato-Bacterial Spot category as Tomato-Late Blight. It classified 18 data points belonging to the Tomato-Bacterial Spot category as Tomato-Septoria Leaf Spot. It classified 19 data points belonging to the Tomato-Bacterial Spot category as Tomato-Yellow Leaf Curl Virus. In total, it incorrectly classified 59 data points belonging to the Tomato-Bacterial Spot category.

The study correctly classified 2229 data points belonging to the "Tomato - Early Blight" category. It classified 31 data points belonging to the "Tomato - Bacterial Spot" category. It classified 15 data points belonging to the "Tomato - Early Blight" category as "Tomato - Healthy". It classified 60 data points belonging to the "Tomato - Early Blight" category as "Tomato - Late Blight". It classified 46 data points belonging to the "Tomato - Early Blight" category as "Tomato - Septoria Leaf Spot". It classified 19 data points belonging to the "Tomato - Early Blight" category as "Tomato - Yellow Leaf Curl Virus". In total, it incorrectly classified 171 data points belonging to the "Tomato - Early Blight" category.

The study correctly classified 2390 data points belonging to the "Healthy Tomato" category. It classified 6 data points belonging to the "Healthy Tomato" category as "Early Blight," 4 data points as "Late Blight," and 7 data points as "Septoria Leaf Spot." In total, it incorrectly classified 17 data points belonging to the "Healthy Tomato" category.

The study correctly classified 2199 data points belonging to the "Tomato - Late Blight" category. It classified 4 data points belonging to the "Tomato - Late Blight" category as "Tomato - Bacterial Spot". It classified 68 data points belonging to the "Tomato - Late Blight" category as "Tomato - Early Blight". It classified 11 data points belonging to the "Tomato - Late Blight" category as "Tomato - Healthy". It classified 26 data points belonging to the "Tomato - Late Blight" category as "Tomato -

Septoria Leaf Spot". It classified 6 data points belonging to the "Tomato - Late Blight" category as "Tomato - Yellow Leaf Curl Virus". In total, it incorrectly classified 115 data points belonging to the "Tomato - Early Blight" category.

The study correctly classified 2089 data points belonging to the Tomato-Septoria Leaf Spot category. It classified 18 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Bacterial Spot. It classified 50 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Early Blight. It classified 8 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Healthy. It classified 19 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Late Blight. In total, it incorrectly classified 95 data points belonging to the Tomato-Septoria Leaf Spot category.

The study correctly classified 2406 data points belonging to the Tomato-Yellow Leaf Curl Virus class. It classified 19 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Bacterial Spot. It classified 20 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Early Blight. It classified 1 data point belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Healthy. It classified 4 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Late Blight. It classified 1 data point belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Septoria Leaf Spot. In total, it incorrectly classified 45 data points belonging to the Tomato-Yellow Leaf Curl Virus class.

belonging to the "Tomato - Early Blight" category as "Tomato - Healthy". It classified 68 data points belonging to the "Tomato - Early Blight" category as "Tomato - Late Blight". It classified 100 data points belonging to the "Tomato - Early Blight" category as "Tomato - Septoria Leaf Spot". It classified 23 data points belonging to the "Tomato - Early Blight" category as "Tomato - Yellow Leaf Curl Virus". In total, it incorrectly classified 571 data points belonging to the "Tomato - Early Blight" category.

The study correctly classified 2382 data points belonging to the "Healthy Tomato" category. It classified 5 data points from the "Healthy Tomato" category as "Bacterial Spot". It classified 12 data points from the "Healthy Tomato" category as "Early Blight". It classified 7 data points from the "Healthy Tomato" category as "Septoria Leaf Spot". It classified 1 data point from the "Healthy Tomato" category as "Yellow Leaf Curl Virus". In total, it incorrectly classified 25 data points from the "Healthy Tomato" category.

The study correctly classified 1876 data points belonging to the "Tomato - Late Blight" category. It classified 125 data points belonging to the "Tomato - Bacterial Spot" category. It classified 159 data points belonging to the "Tomato - Late Blight" category as "Tomato - Early Blight". It classified 63 data points belonging to the "Tomato - Healthy" category. It classified 59 data points belonging to the "Tomato - Late Blight" category as "Tomato - Septoria Leaf Spot". It classified 32 data points belonging to the "Tomato - Late Blight" category as "Tomato - Yellow Leaf Curl Virus". In total, it incorrectly classified 438 data points belonging to the "Tomato - Early Blight" category.

The study correctly classified 1705 data points belonging to the Tomato-Septoria Leaf Spot category. It classified 221 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Bacterial Spot. It classified 104 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Early Blight. It classified 76 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Healthy. It classified 60 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Late Blight. It classified 18 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Yellow Leaf Curl Virus. In total, it incorrectly classified 479 data points belonging to the Tomato-Septoria Leaf Spot category.

The study correctly classified 2302 data points belonging to the Tomato-Yellow Leaf Curl Virus class. It classified 98 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Bacterial Spot. It classified 21 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Early Blight. It classified 12 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Healthy. It classified 9 data points belonging to the Tomato-Yellow Leaf Curl Virus class as

		PREDICTED					
		Tomato - Bacterial Spot	Tomato - Early Blight	Tomato - Healthy	Tomato - Late Blight	Tomato - Septeria Leaf Spot	Tomato - Yellow Leaf Curl Virus
ACTUAL	Tomato - Bacterial Spot	2052	15	8	11	11	30
	Tomato - Early Blight	323	1829	57	68	100	23
	Tomato - Healthy	5	12	2382	0	7	1
	Tomato - Late Blight	125	159	63	1876	59	32
	Tomato - Septeria Leaf Spot	221	104	76	60	1705	18
	Tomato - Yellow Leaf Curl Virus	98	21	12	9	9	2302

Figure 8. Confusion Matrix of the KNN Model

According to Figure 8, the KNN method correctly classified 2052 data points belonging to the Tomato-Bacterial Spot category. It classified 15 data points as Tomato-Early Blight, 8 as Tomato-Healthy, 11 as Tomato-Late Blight, 11 as Tomato-Septoria Leaf Spot, and 30 as Tomato-Yellow Leaf Curl Virus. In total, it incorrectly classified 75 data points belonging to the Tomato-Bacterial Spot category.

The study correctly classified 1829 data points belonging to the "Tomato - Early Blight" category. It classified 323 data points belonging to the "Tomato - Bacterial Spot" category. It classified 57 data points

Tomato-Late Blight. It classified 9 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Septoria Leaf Spot. In total, it incorrectly classified 149 data points belonging to the Tomato-Yellow Leaf Curl Virus class.

		PREDICTED					
		Tomato - Bacterial Spot	Tomato - Early Blight	Tomato - Healthy	Tomato - Late Blight	Tomato - Septeria Leaf Spot	Tomato - Yellow Leaf Curl Virus
ACTUAL	Tomato - Bacterial Spot	1979	44	0	7	66	31
	Tomato - Early Blight	90	1958	24	175	128	25
	Tomato - Healthy	0	8	2379	5	15	0
	Tomato - Late Blight	40	286	13	1876	95	4
	Tomato - Septeria Leaf Spot	77	189	14	42	1659	3
	Tomato - Yellow Leaf Curl Virus	34	32	0	5	2	2378

Figure 9. Confusion Matrix of the SVM Model

According to Figure 9, the SVM method correctly classified 1979 data points belonging to the Tomato-Bacterial Spot category. It classified 44 data points belonging to the Tomato-Bacterial Spot category as Tomato-Early Blight. It classified 7 data points belonging to the Tomato-Bacterial Spot category as Tomato-Late Blight. It classified 66 data points belonging to the Tomato-Bacterial Spot category as Tomato-Septoria Leaf Spot. It classified 31 data points belonging to the Tomato-Bacterial Spot category as Tomato-Yellow Leaf Curl Virus. In total, it incorrectly classified 148 data points belonging to the Tomato-Bacterial Spot category.

The study correctly classified 1958 data points belonging to the "Tomato - Early Blight" category. It classified 90 data points belonging to the "Tomato - Bacterial Spot" category. It classified 24 data points belonging to the "Tomato - Early Blight" category as "Tomato - Healthy". It classified 175 data points belonging to the "Tomato - Early Blight" category as "Tomato - Late Blight". It classified 128 data points belonging to the "Tomato - Early Blight" category as "Tomato - Septoria Leaf Spot". It classified 25 data points belonging to the "Tomato - Early Blight" category as "Tomato - Yellow Leaf Curl Virus". In total, it incorrectly classified 442 data points belonging to the "Tomato - Early Blight" category.

The study correctly classified 2379 data points belonging to the "Healthy Tomato" category. It classified 8 data points belonging to the "Healthy Tomato" category as "Early Blight," 5 data points as "Late Blight," and 15 data points as "Septoria Leaf Spot." In total, it incorrectly classified 28 data points belonging to the "Healthy Tomato" category.

The study correctly classified 1876 data points belonging to the "Tomato - Late Blight" category. It classified 40 data points belonging to the "Tomato - Bacterial Spot" category. It classified 286 data points

belonging to the "Tomato - Late Blight" category as "Tomato - Early Blight". It classified 13 data points belonging to the "Tomato - Healthy" category. It classified 95 data points belonging to the "Tomato - Late Blight" category as "Tomato - Septoria Leaf Spot". It classified 4 data points belonging to the "Tomato - Late Blight" category as "Tomato - Yellow Leaf Curl Virus". In total, it incorrectly classified 438 data points belonging to the "Tomato - Early Blight" category.

The study correctly classified 1859 data points belonging to the Tomato-Septoria Leaf Spot category. It classified 77 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Bacterial Spot. It classified 189 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Early Blight. It classified 14 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Healthy. It classified 42 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Late Blight. It classified 3 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Yellow Leaf Curl Virus. In total, it incorrectly classified 325 data points belonging to the Tomato-Septoria Leaf Spot category.

The study correctly classified 2378 data points belonging to the Tomato-Yellow Leaf Curl Virus class. It classified 34 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Bacterial Spot. It classified 32 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Early Blight. It classified 5 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Late Blight. It classified 2 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Septoria Leaf Spot. In total, it incorrectly classified 73 data points belonging to the Tomato-Yellow Leaf Curl Virus class.

		PREDICTED					
		Tomato - Bacterial Spot	Tomato - Early Blight	Tomato - Healthy	Tomato - Late Blight	Tomato - Septeria Leaf Spot	Tomato - Yellow Leaf Curl Virus
ACTUAL	Tomato - Bacterial Spot	2017	35	1	16	28	30
	Tomato - Early Blight	55	2103	17	142	64	19
	Tomato - Healthy	0	16	2352	20	19	0
	Tomato - Late Blight	14	144	17	2070	52	17
	Tomato - Septeria Leaf Spot	27	78	16	44	2017	2
	Tomato - Yellow Leaf Curl Virus	30	22	1	20	3	2375

Figure 10. Confusion Matrix for the LR model.

According to Figure 10, the LR method correctly classified 2017 data points belonging to the Tomato-Bacterial Spot category. It classified 35 data points as Tomato-Early Blight. It classified 1 data point as Tomato-Healthy. It classified 16 data points as Tomato-Late Blight. It classified 28 data points as Tomato-Septoria Leaf Spot. It classified 30 data points as Tomato-Yellow Leaf Curl

Virus. In total, it incorrectly classified 110 data points belonging to the Tomato-Bacterial Spot category.

The study correctly classified 2103 data points belonging to the "Tomato - Early Blight" category. It classified 55 data points belonging to the "Tomato - Bacterial Spot" category. It classified 17 data points belonging to the "Tomato - Early Blight" category as "Tomato - Healthy". It classified 142 data points belonging to the "Tomato - Early Blight" category as "Tomato - Late Blight". It classified 64 data points belonging to the "Tomato - Early Blight" category as "Tomato - Septoria Leaf Spot". It classified 19 data points belonging to the "Tomato - Early Blight" category as "Tomato - Yellow Leaf Curl Virus". In total, it incorrectly classified 155 data points belonging to the "Tomato - Early Blight" category.

The study correctly classified 2352 data points belonging to the "Healthy Tomato" category. It classified 16 data points belonging to the "Healthy Tomato" category as "Early Blight," 20 data points as "Late Blight," and 19 data points as "Septoria Leaf Spot." In total, it incorrectly classified 55 data points belonging to the "Healthy Tomato" category.

The study correctly classified 2070 data points belonging to the "Tomato - Late Blight" category. It classified 14 data points belonging to the "Tomato - Bacterial Spot" category. It classified 144 data points belonging to the "Tomato - Late Blight" category as "Tomato - Early Blight". It classified 17 data points belonging to the "Tomato - Healthy" category. It classified 52 data points belonging to the "Tomato - Late Blight" category as "Tomato - Septoria Leaf Spot". It classified 17 data points belonging to the "Tomato - Late Blight" category as "Tomato - Yellow Leaf Curl Virus". In total, it incorrectly classified 244 data points belonging to the "Tomato - Early Blight" category.

The study correctly classified 2017 data points belonging to the Tomato-Septoria Leaf Spot category. It classified 27 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Bacterial Spot. It classified 78 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Early Blight. It classified 16 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Healthy. It classified 44 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Late Blight. It classified 2 data points belonging to the Tomato-Septoria Leaf Spot category as Tomato-Yellow Leaf Curl Virus. In total, it incorrectly classified 167 data points belonging to the Tomato-Septoria Leaf Spot category.

The study correctly classified 2375 data points belonging to the Tomato-Yellow Leaf Curl Virus class. It classified 30 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Bacterial Spot. It classified 22 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Early Blight. It classified

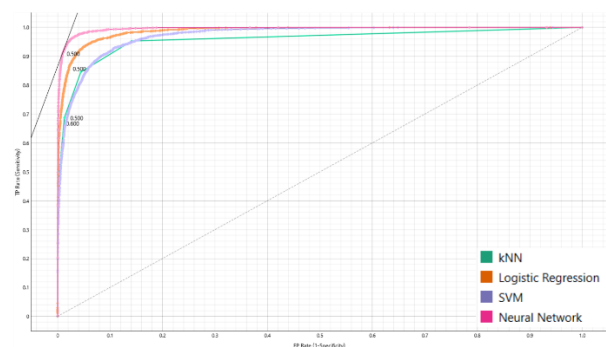
1 data point belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Healthy. It classified 20 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Late Blight. It classified 3 data points belonging to the Tomato-Yellow Leaf Curl Virus class as Tomato-Septoria Leaf Spot. In total, it incorrectly classified 76 data points belonging to the Tomato-Yellow Leaf Curl Virus class.

Performance metrics for each method were calculated using data from complex matrices obtained from the methods. These metrics are shown in Table 1.

**Table 1.** Performance Metrics for All Models

	Accuracy	F1 Score	Precision	Recall
KNN	0,998	0,983	0,952	0,990
LR	0,999	0,992	0,978	0,977
SVM	0,999	0,994	0,979	0,988
ANN	1,000	0,996	0,986	0,993

According to Table 1, the most successful model is ANN with a score of 1.000. The model with the lowest classification success is the KNN model with a score of 0.998. The classification success of the LR model is 0.999, and the classification success of the SVM model is also 0.999. When other performance metrics are examined, it is seen that the F1 Score, Precision, and Recall metric values also show parallelism with the classification success of the models. When the classification successes and F1 Score, precision, and recall metrics are examined, it can be said that all models show a success rate of 99.8% and above. Figure 11 shows the ROC analysis curve for all models of the tomato image data titled "Tomato - Bacterial Spot".



**Figure 11.** Tomato - ROC analysis curve of early blight image data.

Figure 11 shows a parallel between the classification success rates and the learning levels of the models in the Tomato Early Blight data. ANN is the best-learning model, while KNN is the least-learning model. Therefore, it can be said that the classification success of the models is determined by their learning rates. After classifying the

data, the main goal is to improve the health of tomato plants by reducing the amount of pesticides applied and implementing organic farming methods on tomato plants.

After the classification of tomato plant data is completed, certain tests must be performed and care must be taken in cultivating these plants to prevent these diseases from occurring. As a result, the bacteria that cause the first disease, Bacterial Spot, live on plant debris. The onset of the disease is facilitated by hot and rainy weather. The symptoms of Bacterial Spot are very similar to other tomato diseases, but only bacterial spots cause bacterial ooze from a cut leaf. The disease spreads through infected seeds, wind-driven rain, diseased seedlings, or infested soil. Bacteria can enter the plant through natural openings in the leaves or through openings caused by damage to the leaves. To prevent Bacterial Spot disease, it is necessary to use only certified seeds and healthy seedlings and to clean all crop residues from the planting area. Additionally, watering the plant at the base instead of using sprinklers and practicing crop rotation are measures taken to prevent tomato plants from contracting Bacterial Spot disease [18].

Another disease, Early Blight, can spread rapidly after plants bear fruit; airborne spore movement and contact with infected soil can cause the disease to spread. Applying appropriate fungicides at the first sign of the disease, eradicating certain wild-growing nightshade plants (tomatoes, potatoes, etc.), and implementing crop rotation can be effective in preventing this disease.

Late Blight disease, which is a distinct disease, can destroy all tomato cultivation areas where it occurs. Therefore, resistant tomato varieties should be planted to prevent Late Blight disease. If disease symptoms are present, or if rainy conditions are likely or overhead irrigation is used, appropriate fungicides should be applied in specific quantities and in a knowledgeable manner [18].

Septoria Leaf Spot disease can spread through water splashing. The fungus that causes this disease can survive the winter in plant debris. To prevent this disease, ensure that all tomato crop debris is cleaned up and destroyed or thoroughly mixed into the soil in the fall. Planting only disease-free plants, avoiding overhead irrigation, and staking plants to increase air circulation between leaves are measures taken to prevent the disease. As a last resort, applying the appropriate amount of fungicide prevents the disease.

The Yellow Leaf Curl Virus (YLCV) is transmitted by whiteflies and can cause a 100% yield loss if it infects plants during the early stages of the crop. The virus can also infect other hosts such as beans, ornamental plants, and various weed species. To prevent this disease, resistant tomato varieties can be grown. It is necessary to plant disease- and whitefly-free seedlings, remove and burn infected plants, and clear the field of weeds. Measures such as using yellow sticky traps to monitor and control whiteflies can be taken. If the insect infestation is severe,

spraying appropriate insecticides is one of the measures taken to prevent the disease.

It is crucial to practice organic farming consciously to limit the use of pesticides and avoid overuse. This study sought to determine what kind of disease the tomato plant data used had on tomatoes. To this end, different tomato plant data were examined and classified using specific methods. The method that applied the classification most accurately was ANN with 100%. The lowest classification success was KNN with 99.8%.

After identifying the disease with these data, the study explains what can be done to prevent it. The aim is to reduce pesticide use or limit it within certain boundaries. However, with more data, all tomato diseases can be identified, and thus pesticide use can be gradually reduced, promoting the spread of organic farming and ensuring the health of living beings and a cleaner environment.

The classifications recommended in the study are used in greenhouses and gardens where tomato plants are grown, facilitating the identification of tomato plant diseases and helping to take appropriate measures. This enables the cultivation of healthier tomato plants by preventing the indiscriminate use of pesticides.

#### 4. Conclusions

It is crucial to practice organic farming consciously to limit the use of pesticides and avoid overuse. This study sought to determine what kind of disease the tomato plant data used had on tomatoes. To this end, different tomato plant data were examined and classified using specific methods. The method that applied the classification most accurately was ANN with 100%. The lowest classification success was KNN with 99.8%.

After identifying the disease with these data, the study explains what can be done to prevent it. The aim is to reduce pesticide use or limit it within certain boundaries. However, with more data, all tomato diseases can be identified, and thus pesticide use can be gradually reduced, promoting the spread of organic farming and ensuring the health of living beings and a cleaner environment.

The classifications recommended in the study are used in greenhouses and gardens where tomato plants are grown, facilitating the identification of tomato plant diseases and helping to take appropriate measures. This enables the cultivation of healthier tomato plants by preventing the indiscriminate use of pesticides.

#### Declaration of Ethical Standards

The article does not contain any studies with human or animal subjects.

#### Credit Authorship Contribution Statement

Authors individually were responsible for the ideation, modeling, analysis, and writing of this article.

## Declaration of Competing Interest

Author claims that there are no conflicts of interest.

## Funding / Acknowledgements

No funding was received from any organization for this research.

## Data Availability

<https://www.kaggle.com/datasets/snikhilrao/crop-disease-detection-dataset>

## References

- [1] A. Akdoğan, Ü. Divrikli, and L. Elçi, "Importance of pesticides and their effects on ecosystem," *Akademik Gıda*, vol. 10, no. 1, pp. 125–132, Mar. 2012. [Online]. Available: <https://izlik.org/JA26ND97UJ>
- [2] O. Tiryaki, R. Canhıal, and S. Horuz, "The use of pesticides and their risks," *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi*, vol. 26, no. 2, pp. 154–169, Apr. 2010. [Online]. Available: <https://izlik.org/JA67XU23RC>
- [3] Z. Özcan and S. Tongur, "Pestisitlerin toksisitesinin *Lepidium sativum* test moduyla çevre ve insan sağlığı açısından değerlendirilmesi," *Ulusal Çevre Bilimleri Araştırma Dergisi*, vol. 2, no. 4, pp. 144–150, 2019.
- [4] Ş. E. Arıcı et al., "Isparta/Keçiborlu ilçesinde yağ gülü yetiştiriciliği, pestisit kullanımı hakkında üretici sorunları ve uygulama esasları," *Turkish Journal of Science and Engineering*, vol. 4, no. 1, pp. 15–20, 2022.
- [5] R. Yüzbaşıoğlu and Ş. Topkaya, "Üreticilerin tarım ilacı uygulamasındaki bilgisi ve ilaçlamada çevre duyarlılığı: Tokat ili merkez ilçe örneği," *Gaziosmanpaşa Bilimsel Araştırma Dergisi*, vol. 11, no. 3, pp. 306–316, 2022.
- [6] Ö. Ü. A. Polat, "Tarımda pestisitler: Dünya'da ve Türkiye'de kullanımları," 2022.
- [7] B. Özercan and R. Taşçı, "Türkiye'de pestisit kullanımının iller, bölgeler ve pestisit grupları açısından incelenmesi," *Ziraat Mühendisliği*, no. 375, pp. 75–88, 2022.
- [8] H. Yılmaz, A. Düzenli, and M. M. Dağ, "Dünya, Avrupa Birliği ülkeleri ve Türkiye'de pestisit kullanımı ve yasal düzenlemeler," *Türkiye Tarımsal Araştırmalar Dergisi*, vol. 11, no. 3, pp. 315–330, 2024.
- [9] M. Ayyıldız, "Türkiye'de kimyasal pestisit kullanımının ekonomi ve çevre yönüyle değerlendirilmesi," *COMU Journal of Agriculture Faculty*, vol. 10, no. 2, pp. 267–274, 2022.
- [10] H. H. Tıraş, "Tarım üretiminde önde gelen ülkelerde tarımsal gübre ve ilaç kullanımının tarımsal GSYİH'ye etkisi: Panel veri analizi," *Business & Management Studies: An International Journal*, vol. 12, no. 1, pp. 132–147, 2024.
- [11] F. N. Doğan and M. E. Karpuzcu, "En iyi yönetim uygulamalarının pestisit kirliliğinin kontrolünde kullanımı," *Çevre İklim ve Sürdürülebilirlik*, vol. 24, no. 1, pp. 25–32, 2023.
- [12] M. Aydoğan and M. Baran, "Tütün yetiştiriciliğinde tarım ilaçları kullanımını ve üreticilerin zirai mücadele bilinç düzeylerinin belirlenmesi: Çelikhana tütünü örneği, Türkiye," *Türkiye Tarımsal Araştırmalar Dergisi*, vol. 10, no. 2, pp. 180–190, 2023.
- [13] S. Hayran, A. D. Çelik, and A. Gül, "Türkiye'de sera çiftçilerinin pestisit kullanımı ve risk algılamaları: Mersin ili örneği," *Tarım Ekonomisi Dergisi*, vol. 28, no. 1, pp. 31–39, 2022.
- [14] H. G. Doğan and H. Karaaşlan, "Türkiye'de tarım sektörü kaynaklı çevre tahribatını önleme politikaları kapsamında destekleme alan üreticilerin incelenmesi (Seyfe Gölü yöresi hububat üreticileri örneği)," *Turkish Journal of Agriculture-Food Science and Technology*, vol. 10, no. 11, pp. 2213–2223, 2022.
- [15] M. Tözün and G. Akar, "Türkiye'de gıda numunelerinde pestisit kalıntıları üzerine 2010 yılı sonrası ulusal literatürün incelenmesi," *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Halk Sağlığı Dergisi*, vol. 7, no. 1, pp. 177–191, 2022.
- [16] O. Gedikli, A. Uzundumlu, and G. Tozlu, "Çeltik, mısır ve buğday üretiminde tarımsal ilaç kullanımının çevresel duyarlılık yönünden incelenmesi: Samsun ili örneği," *TÜBAV Bilim Dergisi*, vol. 8, no. 2, pp. 19–26, 2015.
- [17] D. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, 2015.
- [18] F. A. Esen and A. Onan, "Derin öğrenme yöntemleri ile bitki yaprakları üzerindeki hastalıkların sınıflandırılması," *Avrupa Bilim ve Teknoloji Dergisi*, no. 40, pp. 151–155, 2022.
- [19] S. Gökalkp and İ. Aydın, "Farklı derin sinir ağı modellerinin duyu tanımadaki performanslarının karşılaştırılması," *Muş Alparslan Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, vol. 2, no. 1, pp. 35–43, 2021.
- [20] Y. S. Taspinar, "Machine learning-driven e-nose-based diabetes detection: Sensor selection and feature reduction study," *Sensors*, vol. 25, no. 21, p. 6607, 2025.
- [21] B. Bahçıvan and A. Yılmaz, "Demografik özelliklerin çevrimiçi market alışverişini kullanımına etkisinin makine öğrenmesi yöntemleri ile tahmini," *Konya Journal of Engineering Sciences*, vol. 11, no. 4, pp. 942–957, 2023.