Tamap Journal of Engineering
http://www.tamap.org/

doi:10.29371/2021.3.125
Volume 2021, Article ID 125
Research Article

Plant Disease Detection by Using Adaptive Neuro-Fuzzy Inference System

Rabia Bayraktar[1*], Bülent Haznedar[2], Kadir Sercan Bayram[3] and Muhammet Fatih Hasoğlu[4]

[1] *Electronics-Computer Engineering, Institute of Natural Science, Hasan Kalyoncu University, Gaziantep, Turkey*
[2] *Department of Computer Engineering, Engineering Faculty, Gaziantep University, Gaziantep, Turkey*
[3] *Department of Computer Engineering, Engineering Faculty, Hasan Kalyoncu University, Gaziantep, Turkey*
[4] *Department of Electrical-Electronics, Engineering Faculty, Hasan Kalyoncu University, Gaziantep, Turkey*

Received: 06.07.2020 • **Accepted**: 01.09.2021 • **Published Online:** 07.09.2021

Abstract: This study aims the detection and recognition of plant diseases by using the most modern methods including Support Vector Machine (SVM), Convolution Neural Network (CNN), Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Interface System (ANFIS). In the studies, 2340 training and 585 test data were used with 3 different tomato plant leaves as Healthy, Early blight, and Yellow leaf curl virus. These methods are used in a wide spectrum of research areas. While creating the dataset, a total of 11 features were extracted from the existing image data. 91% accuracy has been achieved with the proposed ANFIS which is the best compared to the other methods with 11 features.

Keywords: ANFIS, Plant Diseases, Recognition, Detection

1. INTRODUCTION

Plant disease is an abnormal condition that affects the normal growth of the plant's stem and leaves [1]. Plant diseases are generally named according to the symptoms and can affect plants of all types and at ages. Just as in humans, 3 basic factors are required for the emergence of the disease in plants such as a disease-sensitive plant, a pathogen that can cause disease, and a disease-friendly environment. If these three factors coexist, unfortunately infectious plant diseases occur. On the other hand, if the plant shows symptoms such as poor growth or yellow and rotting leaves without the presence of any pathogens, the presence of non-infectious plant disease is highly likely. The first step in identifying plant diseases is to characterize the symptoms. Symptoms can be local and limited, such as leaf spots, or general and systematic, such as poor growth or withered leaves.

Green parts of the plant, such as leaves, may occasionally exhibit color changes or deformations. This may mean that the pathogen is within the metabolism of the plant and that the related disease is caused by a virus. Virus disease symptoms are often confused with symptoms of diseases caused by nutritional deficiencies. It is possible to mention that if the plant is suddenly ill and twisted, although the soil is the same and evenly watered, the plant suffers from a soil-borne disease.

The fact that there is a variety of plant diseases and factors requires the use of technological methods in the detection of the disease. Until now, image processing, machine learning, deep learning methods have been frequently used in

[*] Correspondence: rabia.bayraktar@std.hku.edu.tr

the recognition of both plant and human diseases. In this study, an ANFIS model has been applied for plant disease detection. It is mainly different from the previous methods with remarkable results and outperformed other methods with its performance and accuracy. In these methods, 11 features are obtained from image processing of feature extraction method. In the first stage, the diseased and healthy labels were taken by created models. Then, one more label was added to the existing labels to recognize the disease.

2. MATERIALS and METHODS

2.1. Experimental Datasets

The dataset of all plants used to recognize the disease was taken over GitHub, an open-source sharing site [14]. The data used in the modelling are formed from images with the same size and format of the Joint Photographic Experts Group (JPG). In order to use these images in the model, they must be converted into processable numerical data. The numerical dataset in this study consist of a total of 2340 data with 780 data for each label. The test data was determined as 25% of the total data, so 585 test data and 1755 train data were used for each method.

2.2. Image Processing

Image-processing can be defined as, the process of obtaining useful information by using characteristic features over an image or video. The basic logic behind the image processing is to handle the image as a function and convert it into a two-dimensional array. Performing mathematical operations on the converted two-dimensional array mean doing the same on the image, but the operations should be converted into numerical data that the computer can process. In this study, the methods used for feature extraction are explained in the following subsections.

This is the stage of preparing the data to be processed. At this stage, the Gaussian Blur filter, which is commonly-used method in image processing is incorporated in this study. The mathematical expressions of this filter are based on Gaussian Distributions and given as below formulas for two-dimensional plane and three-dimensional space [2].

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma}} \tag{1}$$

$$G(x) = \frac{1}{\sigma^2\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

2.3. Color Feature Extraction

Feature extraction is a size reduction process. It has always been a recommended method to process the distribution of color codes, namely histogram, instead of processing the entire picture. Histogram intensity extraction is an image processing method suitable for images whose color values are not evenly distributed [3].

$$P_r(r_k) = \frac{n_k}{n}, k = 0, 1, 2, ..., L-1 \tag{3}$$

$P(r_k)$, k indicates at what rate the color value is represented in the image, n_k is the number of colors are in the image which defined color as k and n is the total number of pixels. The cumulative probability function is calculated by the following expression:

$$S_k = T(r_k) = \sum_{j=0}^{k}\left(P_r(r_j)\right) = \sum_{j=0}^{k}\frac{n_j}{n}, k = 0,1,2,..,L-1 \tag{4}$$

Then reverse conversion also can be done. This process finds the color that will replace the calculated colors.

$$r_k = T^{-1}(S_k), 0 \ll S_k \ll 1 \tag{5}$$

$$T^{-1}(S_k) = (L-1) * T(r_k), k = 0,1,2,..,L-1 \tag{6}$$

2.4. Texture Analysis

Texture analysis is a widely-used method in image processing. It was developed in the 1970s for image analysis and classification [4-7]. Texture analysis is a way of defining the distribution of densities on an image. In other words, similar and different regions can be distinguished by using this processing method.

Haralick texture features, calculated by using the gray level co-occurrence matrix (GLCM), are often used because they are easy to apply. Also, the performance is quite high in this method. GLCM requires that the images used are reduced to a gray level. Therefore, GLCM can be defined as the histogram of grayscale colors created on an image with a certain threshold value. There are 14 statistics that can be calculated using GLCM. The statistics used in this study are formulated below. N_g is a square matrix $[i, j]$ in size, the number of gray levels in the image.

$$Ct = \sum_{n=0}^{N_{q-1}} n^2 \left\{\sum_{i=1}^{N_q}\sum_{j=1}^{N_q} P(i,j)\right\}, |i-j| = n \tag{7}$$

For correlation statistic, where μ_x, μ_y, σ_x and σ_y are the means and standard deviations of p_x and p_y, the partial probability density functions.

$$C = \frac{\sum_i \sum_j (ij) P(i,j) \mu_x \mu_y}{\sigma_x \sigma_y} \tag{8}$$

For inverse difference moment;

$$I = \sum_i \sum_j \frac{1}{1+(i+j)^2} P(i,j) \tag{9}$$

For entropy;

$$E = -\sum_i \sum_j p(i,j) \log\left(p(i,j)\right) \tag{10}$$

2.5. Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is a type of Artificial Neural Network (ANN) method based on a fuzzy inference system [8]. It is a system that combines the fast training algorithms of Artificial Neural Network and the advantages of Fuzzy Logic. It was developed in the early 1990s by Jang to model nonlinear functions and predicting chaotic time series [9].

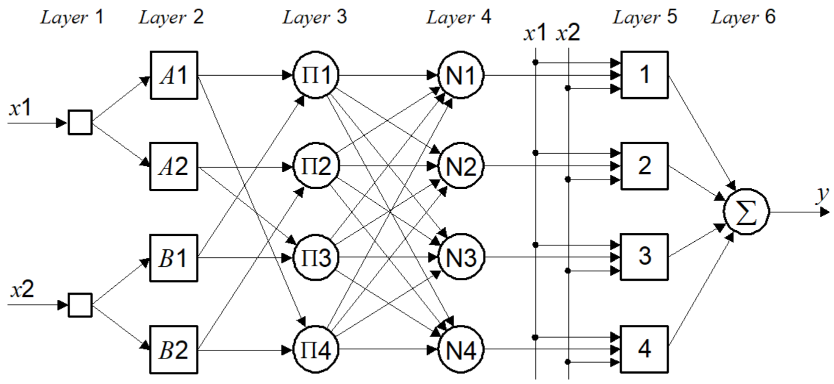

Figure 1. ANFIS Structure

Similar to Artificial Neural Network, ANFIS also consists of input and output pairs. The formula that forms the basic structure of ANFIS can be written as the following:

$$F_1 = p_1 x_1 + q_1 x_2 + r_1 \tag{11}$$

where x_1 and x_2 input values and p and q are linear output parameters. ANFIS layers can be briefly described as follows.

Layer 1: The layer where the membership function is defined for each entry. The output of this layer indicates the membership degrees. The output obtained for each input x_1 and x_2 is formulated as follows:

$$\begin{aligned} O_{i,j} &= \mu . A_i(x_1) \\ O_{i,j} &= \mu . B_{i-j}(x_2) \end{aligned} \tag{12}$$

Layer 2: Also called the rule layer. Precision degrees of the rules are calculated in this layer.

$$O_{2,i} = w_i = \mu A_i(x_1) . \mu B_i(x_2) \tag{13}$$

Layer 3: In this layer, the normalization process of the rules is performed, therefore it is also called normalization layer.

$$O_{3,i} = w_i = \frac{\overline{w_i}}{w_1 + w_2} \tag{14}$$

Layer 4: In this layer, each normalized rule is multiplied by its output function. Therefore it is also known as defuzzification layer.

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}\left(p_i x_1 + q_i x_2 + r_i\right) \tag{15}$$

Layer 5: It is the layer where the output of ANFIS is obtained by adding the values from Layer 4.

$$O_{5,i} = f = \sum_{i=1}^{n} w_i f_i = \frac{\sum_{i=1}^{n} w_i f_i}{\sum_{i=1}^{n} w_i} \tag{16}$$

3. EXPERIMENTAL STUDY

This study aims to use new-generation learning algorithms and methods for the detection of plant leaves with disease. The methods chosen for determining the disease have proven their effectiveness in many areas and have shown high performance. In this study, we will evaluate which method yields more accurate results in a problem with high complexity. To compare the results obtained according to the complexity of the problem, two different studies were conducted by using two and three labels. Performance metrics to compare the results of different methods and their corresponding mathematical formulas are presented in Table 1.

Table 1. Performance metrics and their mathematical expressions.

Metrics	Formula
Accuracy	$\frac{Tp+Tn}{Tp+Tn+Fp+Fn}$
Precision	$\frac{Tp}{Tp+Fp}$
Recall	$\frac{Tp}{Tp+Fn}$
F-score	$\frac{2*recall*precision}{recall+precision}$
Rmse	$\sqrt{\frac{\sum_{i=1}^{N}\left(x_i-y_i\right)^2}{N}}$
Mse	$\left(Rmse\right)^2$

3.1. Detection of Two-labelled Plant Disease

The labels used in this study are only diseased or healthy labels. One output has been obtained from the learning models based on given 11 features. The datasets used in the models are listed in Table 2.

Table 2. Dataset and descriptions

Dataset	Label	Leaf type description	Number of Data
		Early-Blight Disease	*780*
Diseased	*0*	*Yellow curl virus*	*780*
Healthy	*1*	*Healthy Tomato plant leaves*	*780*

As can be noticed from the table, a single label was used for two different diseases. In other words, there is no item distinguishing the disease here. Results obtained from each method have been evaluated according to the metrics given in Table 1. The accuracies are used as comparison for the performance rates of the methods. The results obtained from the methods are as follows.

Table 3. Accuracy values of all methods.

	% Accuracy	
Method	Train	Test
SVM	88.94%	88.05%
CNN	99.39%	97.85%
ANN	98.81%	97.14%
ANFIS	99.72%	99.11%

The proposed ANFIS method has yielded the best accurate results among all other methods. The detailed results of the SVM algorithm, which has the farthest results to ANFIS, are as presented in Table 4.

Table 4. Classification report of SVM classifier.

	Precision	**Recall**	**F1-score**	**Support**
0	0.44	0.55	0.49	238
1	0.59	0.48	0.53	323
Macro avg	0.51	0.51	0.51	561
Weighted avg	0.52	0.51	0.51	561

Table 5. Confusion matrix of SVM with linear kernel type.

		Actual Values	
		Positive	*Negative*
	Positive	130	108
Predicted Values	*Negative*	168	155

The basis for classification problem metrics is the balance between accuracy and precision. The ROC curve is used to express this balance. ROC can also be expressed as the fraction of true positives to false positives. The ROC curve of the SVM classifier is shown in Figure 2. CNN and ANN are both based on artificial neural network, so it is expected that both methods should give close values to each other. Detailed results of two different methods based on the same logic are presented in Table 6.

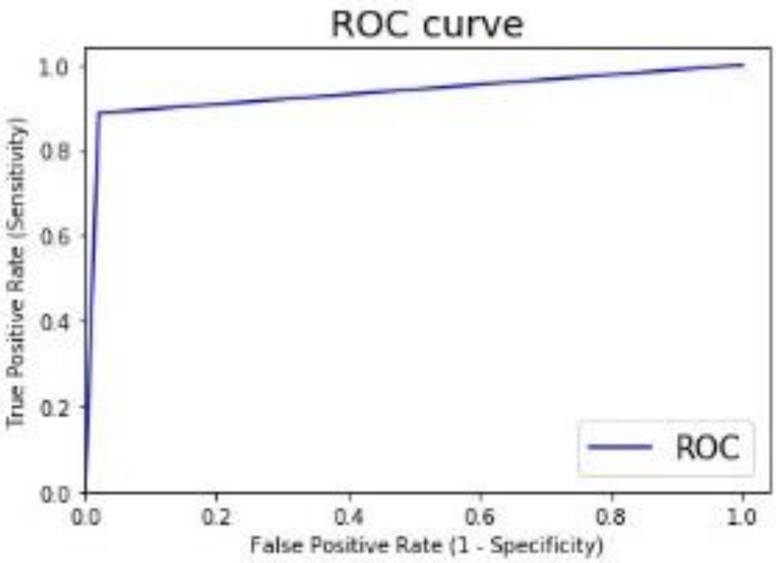

Figure 2. ROC curve of SVM classifier

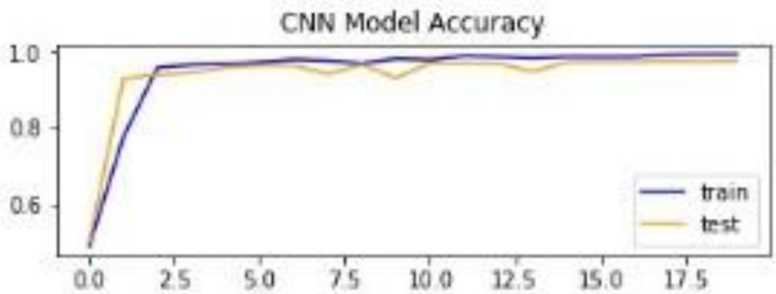

Figure 3. CNN model accuracy

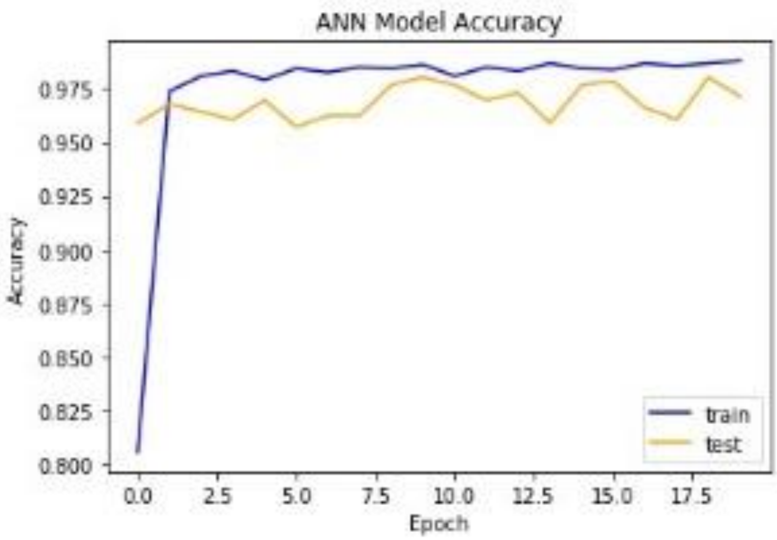

Figure 4. ANN model accuracy

Table 6. ANN and CNN model loss

Method	Epoch	Model Loss	
		Train	*Test*
ANN	20	0.0444	0.0982
CNN	20	0.0223	0.0542

3.2. Detection of Three-labelled Plant Disease

In this part of the study, one more label was added to the same dataset. With the added label, the complexity of the problem has been increased and proposed ANFIS method has been found to give better results compared to other methods for difficult problems.

Table 7. Three-labeled dataset

Dataset	Label	Leaf type description	Number of Data
Diseased	*0*	*Early-Blight Disease*	*780*
	2	*Yellow curl virus*	*780*
Healthy	*1*	*Healthy Tomato plant leaves*	*780*

After processing the three-labeled dataset, the accuracy of results from all methods decreased as expected. Significant decreases were achieved in methods that yielded over 98% with two labels. The results for this part of the study are as follows:

Table 8. Accuracy values of all methods with three-labeled.

Method	% Accuracy	
	Train	Test
SVM	70.25%	65.47%
CNN	44.55%	38.92%
ANN	63.70%	41.02%
ANFIS	86.49%	87.18%

MSE is a metric used to express how close the estimated response value for a particular observation to the actual response value of that observation. It is frequently used in machine learning algorithms and deep learning methods to evaluate model performance. This value is interpreted as the closer the estimated value is to the actual value, the smaller the MSE value. In this study, the values of MSE and RMSE values for applied methods are presented in Table 9.

Table 9. MSE and RMSE values.

Method	METRICS	
	MSE	*RMSE*
SVM	0.9350	0.9669
CNN	0.5128	0.7161
ANN	0.7188	0.8478
ANFIS	0.1437	0.3790

The RMSE values obtained using ANFIS are shown in the Figure 5.

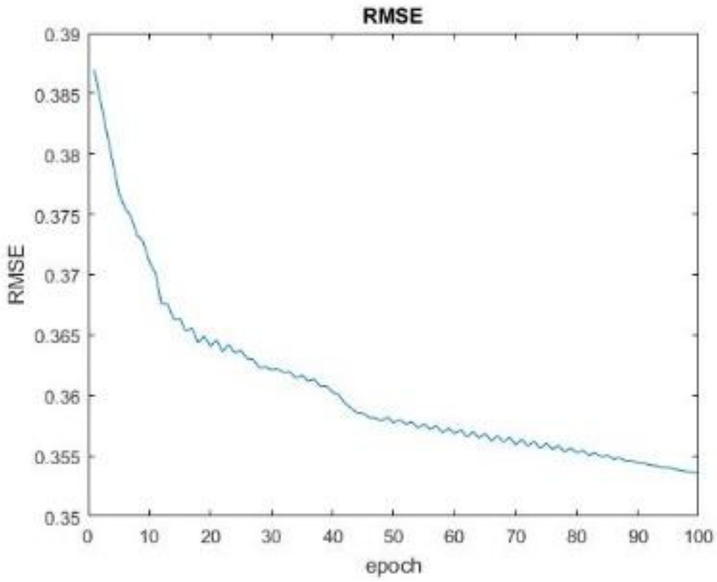

Figure 5. RMSE value of ANFIS

4. CONCLUSIONS

It is the ability to estimate the closest to real values and this is what we expect from artificial intelligence algorithms. In other words, the higher the accuracy rate in artificial intelligence models, the better the model. In addition, it is also important to point that the techniques used in artificial intelligence methods differ according to the problem types and difficulties.

In this study, widely used machine learning, deep learning, and soft computing methods have been applied on the same dataset and the results are evaluated and discussed. Among the methods used for the problem of identification and categorization of plant diseases, the proposed ANFIS method has outperformed other methods in the models we created using both 2 labels and 3 labels. The models, which can achieve results close to ANFIS for two-labeled dataset, lagged far behind ANFIS as the difficulty of the problem increased for three-labeled dataset.

As stated in many previous studies [7, 14], ANFIS provides a high level of performance in solving difficult problems. The results obtained in this study support many previous studies with 99.11% test accuracy for 2 labels and 87.18% test accuracy for 3 labels. Besides the accuracy values, RMSE values, from ANFIS method yielded the lowest values (best) for both labels.

REFERENCES

[1] Sabrol, H., Kumar, S., "Recent Studies of Image and Soft Computing Techniques for Plant Disease Recognition and Classification," *International Journal of Computer Applications,* vol. 126, no. 1, September, 2015.

[2] Gedraite, E., Hadad, M., "Investigation on the Effect of a Gaussian Blur in Image Filtering and Segmentation," *53rd International Symposium ELMAR*, 14-16, September, 2011.

[3] Singh, V., Misra, A.K., "Detection of plant leaf diseases using image segmentation and soft computing techniques,"*Information processing in Agriculture,* 41-49, 2017.

[4] Porebski, A., Vandenbroucke, N., Macaire, L., "Haralick feature extraction from LBP images for color texture classification," *IEEE Tools & Applications*, 2008.

[5] Durga, N. K.,Anuradha, G., "Plant Disease Identification Using SVM and ANN Algorithms," *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878, Vol. 7, Issue: 5S4, February , 2019.

[6] Militante, S.V., Gerardo, B.D., Dionisio, N.V., "Plant Leaf Detection and Disease Recognition using Deep Learning,'' *IEEE Eurasia Conference on IOT;Communication and Engineering,* ISBN: 978-1-7281-2501-5, 2019.

[7] Sladojevic, S., Arsenovic, M., Anderla, A., "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification,'' *Hindawi Publishing Corporation Computational Intelligence and Neuroscience,* vol. 2016, Article ID: 3289801, May 2016.

[8] Ozcalik, H., Uygur, A., "Efficient modeling of dynamic systems by using adaptive network based fuzzy inference system," *KSU Journal of Science and Engineering*, vol. 6, no. 1, pp. 36-46, 2003.

[9] Avci, E., Akpolat, Z.H., "Speed control of DC motor by using adaptive network based fuzzy inference system," *in ELECO'2002 Electrical-Electronics and Computer Engineering Symposium*, pp. 193-196, Bursa, Turkey, 2002.

[10] Ozgan, E., Kap, T., Beycioglu, A., Emiroglu, M., "The prediction of Marshall Stability of asphalt concrete by using adaptive neuro fuzzy inference system," *in International Advanced Technologies Symposium*, Karabuk, Turkey, 2009.

[11] Haznedar, B., "Determine the presence of genetic anomaly in male infertile patients by using artificial intelligence techniques," M.S.c. Thesis, Erciyes University, Kayseri, Turkey, 2010.

[12] Khirade, S.D., Patil, A.B., "Plant Disease Detection Using Image Processing," *IEEE, International Conference on Computing Communication Control and Automation,* 2015.

[13] Sabrol, H., Satish, K., "Tomato Plant Disease Classification in Digital Images using Classification Tree," *International Conference on Communication and Signal Processing*, India, April 6-8, 2016.

[14] Paul, S., "PlantVillage Dataset," http://https://github.com/spMohanty/PlantVillage-Dataset, May 20, 2020.