e-ISSN 2320 –7876 www.ijfans.org
Vol.11,S Iss 1, 2022
Research paper © 2012 IJFANS. All Rights Reserved UGC CARE Listed (Group -I) Journal

MACHINE LEARNING ALGORITHM FOR CROP DISEASES DIAGNOSIS AND IDENTIFICATION

T. Rajasekaran[1*], Vempati Krishna[2], P. Dinshi[3], Rohit Kumar[4], Gurwinder Kaur[5], C. P. Shirley[6]

[1]Department of Computer Science and Engineering, Sri Venkateswara College of Engineering. Sriperumbudur, Chennai, Tamil Nadu, India

[2]Department of Computer Science and Engineering, TKR College of Engineering & Technology, Hyderabad, Telangana, India.

[3]Department of computer science Engineering, Koneru Lakshmaiah Educational Foundation, Guntur, Andhra Pradesh, India

[4]Department of Information Technology, Institute of Technology and Science, Mohan Nagar, Uttar Pradesh, India.

[5]Department of Food Science and Technology, I.K. Gujral Punjab Technical University, Kapurthala, Punjab, India

[6]Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences,Coimbatore, Tamil Nadu, India.

Corresponding mail: rajasekaran@svce.ac.in

ABSTRACT

Plant diseases destroy crops, costing the agricultural business money. Although pesticides have been used to boost agricultural output, their abuse is harmful to the environment. As a consequence, the capacity to diagnose illnesses and differentiate them from nutritional deficits has a significant impact on assessing whether pesticides are essential. Traditional approaches for diagnosing plant diseases in the lab entail time-consuming and difficult chemical procedures. This work proposes an automated strategy that combines machine learning (ML) and image processing techniques to detect and categorise plant illnesses. To train the algorithm on photos, the feature extraction approach is utilised. In order to choose the optimal algorithm for illness detection, the efficacy of multiple deep learning algorithms is tested using training information. The unseen images may be found in the test folder, which is meant to put the system's capacity to detect plant ailments to the test. The overall accuracy of the procedure is 95%. A vast number of photos may be utilised to train the system, resulting in faster and more accurate outputs.

Keywords: Machine learning, plant disease, Plant illness, agriculture

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal

1. Introduction

Global food security is being threatened by plant diseases. Farmers who depend on robust crops also suffer as a result of their harsh effects [1]. In developing nations like India, wherever agriculturalists with small plots of land provide the mainstream of the agricultural output, vermin and illness have been shown to cause significant productivity loss. In order to promote plant development, farmers choose pesticides and seeds based on their agricultural relevance. Leaf damage is made worse by inconsistent pesticide use and dietary restrictions [2]. Utilizing integrated pest management techniques performs better than using conventional pesticides. The prevention of crop loss from disease has been addressed in a number of ways. In a well-thought-out disease management strategy, early detection and treatment of emerging plant diseases are crucial [3].

Visual examination of the plant disease symptoms revealed a considerable lot of complexity [4]. Due to these challenges, senior agronomists often misdiagnose specific illnesses and pathological problems in intensively farmed plants, which results in incorrect findings and analyses [5]. In order to quickly identify plant leaf diseases, this study use the ML Random Forest technique. Regression, classification, and other training methods are combined in a random decision forest system. Numerous decision trees make up a random forest. In order to produce the various case of a arrangement problematic or the mean forecast in the case of a regression perfect, Random Forest combines the training results of all decision trees. In decision tree algorithms, overtraining to their data set is a major problem that random conclusion trees address. This method of supervised categorization entails the construction of several trees [6]. The strength of the forest increases with the amount of trees present. When using a random forest classifier, the accuracy of the method rises as the quantity of decision trees does. Regression and classification, two benefits of this approach, form the foundation of a machine learning system. Deep decision trees, specifically, refer to developing competence with circumstances that are random. Because result trees overfit to the training data, they have a minimum slope and a big diversion. In order to decrease variation, random forests algorithm the results of many decision trees that have been skilled on numerous topographies of a alike training information [7]. The model still performs effectively despite a little increase in inclination and a slight decrease in interpretability as a consequence. A flexible and user-friendly algorithm, Random Forest generates results rapidly. Regarding machine learning The crucial step is feature engineering, commonly referred to as feature extraction. In order for ML algorithms to utilise the data during the training phase, data is transformed into information. By avoiding data duplication, this method enables generalisation throughout the training phase [8].

This approach's main goal is to organise the most important information from the original data in a smaller geographical context. Large and repetitive input data for the ML algorithm are condensed to a smaller feature set. Feature extraction is the process of taking a list of features from input data. If a feature set has information that can identify one item from another, it is classified as exceptional. The feature set that will be used should be constrained to just those values that can reliably distinguish samples from distinct classes. These may be

 © 2012 IJFANS. All Rights Reserved

split into two groups: local features and global features. Shape, colour, and texture are the three major criteria used in the project. To obtain these properties, use the method that follows. Haralick Texture analysis is the study of the elements that are rotationally invariant in a picture. To estimate the texture, the GLCM coincidence conditions are created and added at viewpoints of 0, 45, 90, and 135 degrees. The histogram, which gauges how rough a picture of a leaf is, is the most used method for expressing colour characteristics. It displays the number of pixels in a picture for each colour. Moments of Hu These seven numbers, which are unaffected by changes in the picture, were produced using central moments. The sign of the seventh instant is altered by picture reflection, whereas the first six moments are unaffected by reflection, translation, scale, or rotation. It assesses the leaf image's form.

2. Methodology

To precisely forecast the incidence of the disease on a plant leaf, the project's approach extracts features from the leaves before using ML algorithms. We begin by using a openly obtainable information of ill and good leaves. A method is utilised to categorize the leaves of healthy and ill plants. The identification and differentiation of healthy and unhealthy plant leaves is the project's foremost objective. The eight different diseases of plants are shown in figure 1.

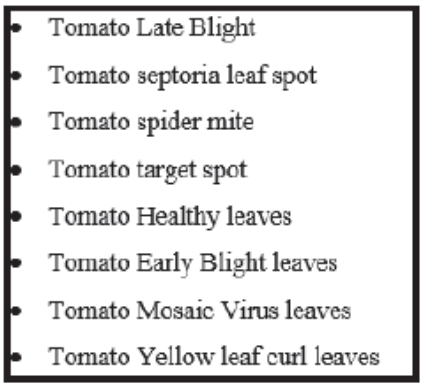

Fig. 1 Name of the different plant diseases

The algorithm is trained using eight modules of tomato leaf pictures from the Plant Village information. A training set and a testing set are created using the exact same samples. The research is alienated into two categories. In the primary step, features from each image are extracted and recorded in an array. During the additional step, deep learning method such as K Nearest Neighbors, Decision Tree, Logistic Regression, , Support Vector Machine, Random Forest are used to train the information. The deep learning technique that uses the dataset to train itself and predicts images that it

 © 2012 IJFANS. All Rights Reserved

hasn't yet seen will provide the model the highest degree of accuracy. The main purpose to evaluate how well an ML model performs on faulty data. Specifically, to evaluate the model's overall performance when used to generate forecasts based on data that was excluded from the model's training set using a restricted example. Since it is easy to use, understand, and has less propensity than other tactics, it is a well-known strategy. The flowchart in figure 2 illustrates the design of forest algorithm

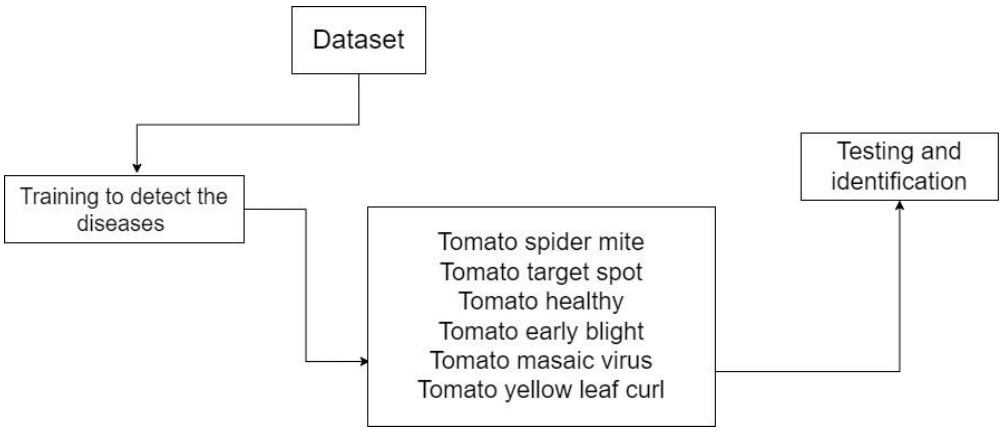

Fig. 2 Flowchart of the different diseases

3. Result and Discussion

The Python program is applied using the Spyder IDE. The anaconda set management is used in this reserch. The collection contains images of tomato leaves in good and bad health. The dataset is divided into eight classes, each with 140 photographs. The dataset's classes are as follows:

Individually binder in the training set is labelled with titles displayed above. During the training phase, the folder names (labels) are added to the labels array while the topographies from the greeneries are removed and added to the feature array. The array and the labels array are employed to train deep learning algorithms. The information is detached into 90% training sets and 10% test sets to evaluate the efficiency of the algorithm using cross-fold verification. Figure 3 represent the training folder of diseased leaves.

```
Images from tomato_early blight processed
Images from tomato_healthy processed
Images from tomato_late blight processed
Images from tomato_mosaic virus processed
Images from tomato_septoria leaf spot processed
Images from tomato_spider mite processed
Images from tomato_target spot processed
Images from tomato_yellow leaf curl processed
```

 © 2012 IJFANS. All Rights Reserved

Fig. 3 Training folder of diseased leaves

ML algorithm accuracy comparisons are performed using a bar plot. After being trained on the information set, deep learning algorithm is put to the test by being presented with a completely new image that was not utilised in the training process. The IPython Console shows the image class predicted by the ML algorithm. The tags are recorded in the labels array, and the features retrieved during training are saved in the training feature array. Figure 3 represent the split training and ML algorithm accuracy.

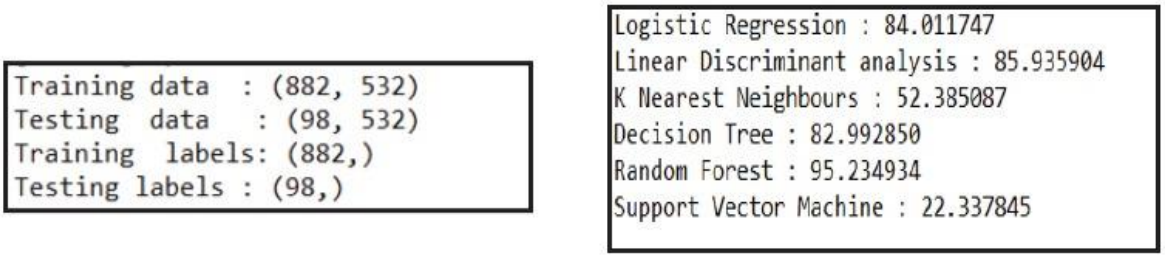

Fig. 4 Split train and accuracy of ML algorithm

The training features and labels generate 90% of the training information and 10% of the test information, respectively. This ratio, however, may be modified by altering the split ratio. Use of the k-fold The algorithm is assessed via cross validation. The accurateness of the dissimilar algorithms is shown in the graph below. According to the graph above, Random Forest had the most accuracy (95.2%), while Support Vector Machine had the lowermost accurateness (22%). ML algorithm accuracy comparisons are performed using a bar plot shown in figure 5. Figure 6 shows the various diseased and healthy tomato leaf used in this research.

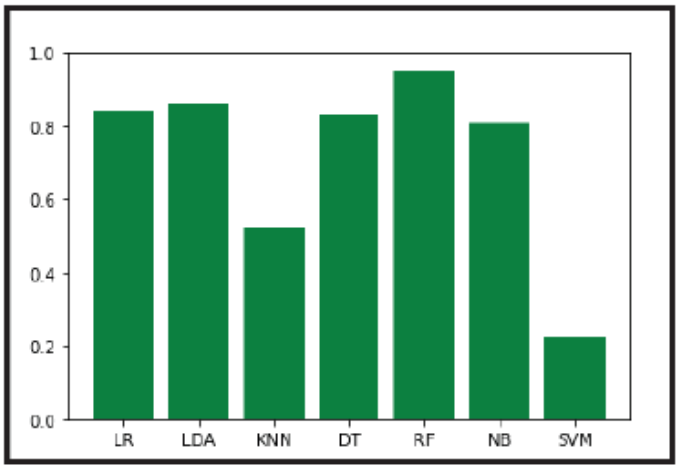

Fig. 5 Performance plot in detecting the diseased leaf.

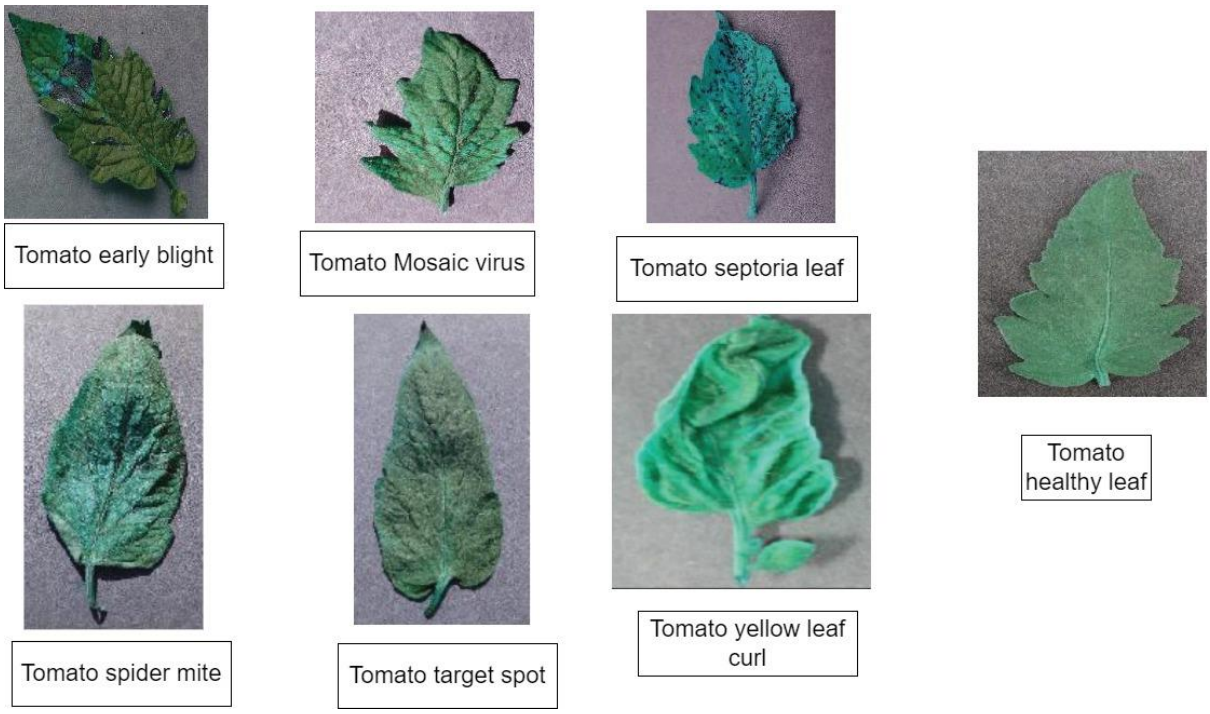

Fig. 6. Various diseased and healthy leaf

4. Conclusion

The research helps to recognises and categorises tomato diseases using random forest deep learning. To train the ML algorithm, features are taken from the pre-processed pictures. The technique finds tomato illnesses such target spot, yellow leaf curl virus, mosaic virus, septoria leaf spot, early blight, late blight, and spider mite. The results of the testing show that almost all of the test photographs provide the most accurate categorization of the various image classes. Accuracy of the system is predicted to be 95%. Since signs of leaves sicknesses, such the mosaic virus in tomato leaves, mirror many nutritional deficiencies, it may be difficult to diagnose them. ML is crucial in identifying the diseases in the leaves since they change based on the age of the plant at the time of contamination. Simply by utilising an image dataset to train the ML system, any leaf disease may be identified.

References

[1] R. Sujatha, J. M. Chatterjee, N. Z. Jhanjhi, and S. N. Brohi, "Performance of deep learning vs machine learning in plant leaf disease detection," *Microprocessors and Microsystems*, vol. 80, no. October 2020, p. 103615, 2021, doi: 10.1016/j.micpro.2020.103615.

[2] J. Parraga-Alava, K. Cusme, A. Loor, and E. Santander, "RoCoLe: A robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition," *Data in Brief*, vol. 25, 2019, doi: 10.1016/j.dib.2019.104414.

[3] G. Yashodha and D. Shalini, "An integrated approach for predicting and broadcasting tea leaf disease at early stage using IoT with machine learning - A review," *Materials*

Today: Proceedings, vol. 37, no. Part 2, pp. 484–488, 2020, doi: 10.1016/j.matpr.2020.05.458.

[4] J. A. Barriga, F. Blanco-Cipollone, E. Trigo-Córdoba, I. García-Tejero, and P. J. Clemente, "Crop-water assessment in Citrus (Citrus sinensis L.) based on continuous measurements of leaf-turgor pressure using machine learning and IoT," *Expert Systems with Applications*, vol. 209, no. June, p. 118255, 2022, doi: 10.1016/j.eswa.2022.118255.

[5] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 305–310, 2022, doi: 10.1016/j.gltp.2022.03.016.

[6] H. Pallathadka *et al.*, "Application of machine learning techniques in rice leaf disease detection," *Materials Today: Proceedings*, vol. 51, pp. 2277–2280, 2022, doi: 10.1016/j.matpr.2021.11.398.

[7] S. M. Javidan, A. Banakar, K. A. Vakilian, and Y. Ampatzidis, "Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning," *Smart Agricultural Technology*, vol. 3, no. March 2022, p. 100081, 2023, doi: 10.1016/j.atech.2022.100081.

[8] F. Jiang, Y. Lu, Y. Chen, D. Cai, and G. Li, "Image recognition of four rice leaf diseases based on deep learning and support vector machine," *Computers and Electronics in Agriculture*, vol. 179, no. October, p. 105824, 2020, doi: 10.1016/j.compag.2020.105824.