Leaf Disease Detection Using Machine Learning

3 authors:

Utkarsha Fulari
Vidya Pratishthans Kamalnayan Bajaj Institute of Engineering and Technology, Bara…

Rajveer K Shastri
Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering & Technology, Baram…

Anuj Namdeo Fulari
Indian Institute of Technology Bombay

Leaf Disease Detection Using Machine Learning

Utkarsha N. Fulari[1], Rajveer K. Shastri[2], Anuj N. Fulari[3]

Department of Electronics and Telecommunication, Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, India

ABSTRACT

Plant phenotyping is a critical aspect of characterizing plant for plant growth monitoring. This paper introduces an efficient approach to identify healthy and diseased or an infected leaf using image processing and machine learning techniques. Various diseases damage the chlorophyll of leaves and affect with brown or black marks on the leaf area. These can be detected using image preprocessing, image segmentation, feature extraction, and classification using machine learning algorithms. For feature extraction, Grey Level Co-occurrence Matrix (GLCM) is applied. Support Vector Machine (SVM) is one of the machine learning algorithms is used for classification. The Convolutional Neural Network (CNN) resulted in a improved accuracy of recognition compared to the SVM approach.

Keywords:

GLCM, Support Vector Machine (SVM), Convolutional Neural Network (CNN), Confusion matrix.

I. INTRODUCTION

In India, for economic development, agriculture is a valuable source. To increase the production of food, the agriculture industries keep on searching for efficient methods to protect crops from damages. This makes researchers search for new efficient, and precise technologies for high productivity. The diseases on crops give low production and economic losses to farmers and agricultural industries [1,2].

For a successful farming system, one of the essential things is disease identification. In general, by using eye observations, a farmer observes symptoms of disease in plants that need continuous monitoring. Different types of disease kill leaves in a plant. For identifying these diseases, farmers get more difficulties. For disease detection, the image processing methods are suitable and efficient with the help of plant leaf images. Though continuously monitoring of health and disease detection of plant increase the quality and quantity of the yield, it is costly. Machine learning algorithms are experimented due to their better accuracy. However, selection of classification algorithms appears to be a difficult task as the accuracy varies for different input data [3]. The objectives are to detect leaf disease portion from the image, extract features of an exposed part of the leaf, and recognize diseased leaf through SVM. Further, Convolutional Neural Network (Alexnet) is evaluated and compared for accuracy.

The paper is arranged into five sections: the first section gives the introduction, the second section presents the literature survey, the third section discusses methodologies like feature extractions of images, SVM and CNN, the fourth section shows the result of classification, and the fifth section is about the conclusion and future scope.

II. LITERATURE SURVEY

Shruthi et al. presented the stages of general plant diseases detection system and study of machine learning techniques for plant disease detection. They showed that a convolutional neural network (CNN) detects many diseases with high accuracy [1]. P. Srinivasan et al. developed the software to classify and categorize groundnut leaf diseases. In their approach, groundnut crop disease such as Early leaf spot, Late leaf spot, Rust , early and late spot Bud Necrosis were categorized only 4 dissimilar diseasee using image acquisition, image preprocessing, segmentation, feature extraction, and the K Nearest Neighbor (KNN) [2].

L. Sherly reviewed of various types of plant diseases and different classification techniques in machine learning that are used for identifying diseases in different plant leaves along with the pros and cons. This paper summaried different algorithms used for classifying and detecting bacterial, fungal and viral plant leaf diseases [3].

Gurleen Kaur et al. reviewed the methods of plant leaves disease detection The major techniques employed were: BPNN, SVM, K-means clustering, Otsu's algorithm, CCM and SGDM. for image segmentation, feature extraction, and classification [4].

Md. Selim et al. used eleven statistical features and the Support Vector Machine classifier (SVM) to recognize the diseases. The advantages of this increase the efficiency of the detection, identification, and classification process. It gives 93% accurate results in disease classification [5]. Monzurul Islam et al. integrated image processing and machine learning to allow diagnosing diseases from leaf images of potato plants. It gave 95% of the accuracy of disease classification using Color thresholder, GLCM and multiclass SVM [6].

Jobin Francis et al. calculated damaged ratio of leaf for the identification of leaf diseases in pepper plants. Masking and threshold based segmentation carried out to separate leaf from background. Using backpropogarion algorithms two types of diseases were identified [7]. Vijai Singh et al. used genetic algorithm for leaf image segmentation. The advantages of this method are that the plant diseases can be identified at an early stage or the initial stage, and with minimal computational efforts and the optimum results [8].

Mrunmayee et al. describe methods of an image processing and neural network for disease detection and classification. The color images preprocessed, and for segmentation, k-means clustering is used. The texture features are extracted using a gray level co-occurrence matrix (GLCM) method and given to the artificial neural network. The overall accuracy of this method is 90% [9]. Sachin D. Khirade et al. discussed segmentation and feature extraction algorithm used in plant disease detection. For classification of disease in plants neural netowrk methods such as self-organizing feature maps, backpropagation algorithms, SVMs, etc. were proposed [10].

Usama Mokhtar et al. used an color space transformation and gray-level co-occurrence matrix (GLCM) for prepossing and feature extraction respectively. Support Vector Machine (SVM) algorithm with different kernel functions is used for the classification phase. The result shows that the classification accuracy is 99.83% [11]. Melike Sardogan et al. presented a Convolutional Neural Network (CNN) model on three different input matrices have been obtained for R, G and B channels to start convolution for every image in the dataset. reLU activation function and max pooling have been implied to the output matrix. and Learning Vector Quantization (LVQ) algorithm used for four types of tomato leaf disease detection and classification [12].

III. METHODOLOGY

Image acquisition is the first step of a plant disease detection system. By using digital cameras, scanners, or drones, high-quality plant images can be captured. The images can also be taken from the web. Large numbers of image samples were collected from Kaggal datasets, which consists of diseased and healthy leaves. Image Preprocessing is used to increase the quality of leaf image and eliminate the unwanted noise. The segmentation process is used to partition the plant image in various segments to separate the diseased portion of the leaf.

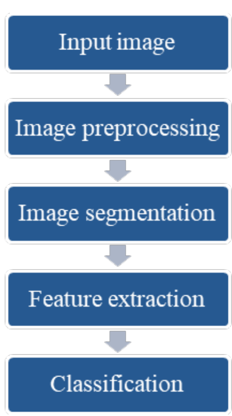

Figure 1: Flowchart for Leaf Disease Recognition

Figure 1 shows the flow chart of operations involved in leaf disease recognition.

Feature Extraction:

The extraction of color, shape, and texture feature of the diseased part of the plant can be done using a grey level co-occurrence matrix (GLCM). There are different techniques of feature extraction that can be employed for growing the system such as gray-level co-occurrence matrix (GLCM), spatial grey-level dependence matrix, color co-occurrence method, and histogram-based feature extraction. The GLCM method is a statistical method for texture classification. In this work, we extracted 6 features (color and textural) from every leaf image of the dataset. The Gray Level Co-occurrence Matrix (GLCM) was used for extracting statistical texture features like contrast, energy, and homogeneity. Besides, numerical indicators like mean, standard deviation, entropy are calculated [6,9].

Some notations-

$g_{ij} = (i, j)^{th}$ entry in GLCM

L-1 = number of distinct gray levels

Contrast- It measures spatial frequency. It is the difference between the highest and lowest values of an adjacent set of pixels. Contrast is given by

$$\sum_i \sum_j (i-j)^2 g_{ij}$$

Energy- Measures the textural uniformity. It reaches a maximum value when the gray level distribution has the same form. It is given by

$$\sum_i \sum_j g_{ij}^{\,2}$$

Homogeneity- It passes the value that calculates the tightness of the distribution of the elements in the GLCM. It is given by

$$\sum_i \sum_j \frac{1}{1+(i-j)^2} g_{ij}$$

Mean is expresses as

$$\sum_{i=0}^{L-1} g(i) P(g(i))$$

and standard deviation is

$$\sqrt{\sum_{i=0}^{L-1} (g(i) - M)^2 P(g(i))}$$

Entropy measures the disorder or complexity in the image. It is large when the image is not texturally uniform and GLCM features have very small values. Entropy is given by

$$\sum_{i=0}^{L-1} P(g(i)) \log_2 P(g(i))$$

Support Vector Machine (SVM) classifier:

To classify the various diseases in plants any of the machine learning techniques can be used. The classification phase suggested deciding if the input image is healthy or diseased. In this paper Support Vector Machine (SVM) classifier has been used because it has some advantages over other classifiers such as effective in high dimensional spaces also in cases where the number of dimensions is greater than the number of samples. It is memory efficient as it uses a subset of training points in the decision function (called support vectors) [5].

SVM is a supervised machine learning algorithm used for both classification and regression. SVM is a discriminative classifier. In this approach, for classification, the SVM technique has been applied. In SVM, each data item is plotted as a point in n-dimensional space; the number of dimensions corresponding to the number of features being classified. The classification is obtained by discovering the hyper-plane that uniquely distinguishes between different groups of scattered data points [5]. By finding the hyper-plane the classification is performed. Hyper-plane differentiates two classes very well. This is shown in Figure 2.

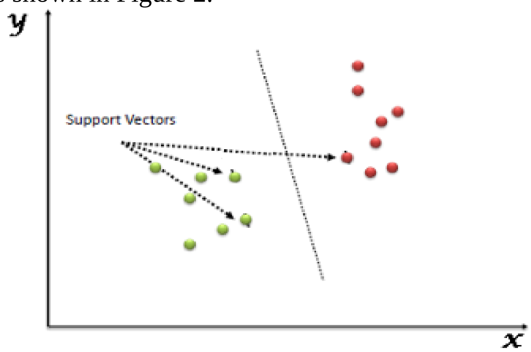

Figure 2: Support Vector Machine

Convolutional Neural Network (CNN):

CNN is a class of deep neural networks. The CNN model comprises an input layer, convolution layer, pooling layer, a fully connected layer, and an output layer shown in figure 3. To classify the disease in plants in a precise manner the images are provided as input. The convolution layer is used for extracting the features from the images. The pooling layer computes the feature values from the extracted features. Depending on the complexity of images, the convolution and pooling layer can be further increased to obtain more details. A fully connected layer uses the output of previous layers and transforms them into a single vector that can be used as an input for the next layer. The output layer finally classifies the plant disease [12].

IV. DATASET

For SVM, 52 images are taken from open source. In that 37 images are used for training and 15 images are used for testing. SVM detects whether a leaf is healthy or infected. For CNN, the source of data is collected from the Kaggle website i.e. “new plant diseases dataset”. From that dataset 12,949 images are used for training. Image dataset containing healthy and unhealthy crop leaves such as apple, cherry, corn (maize), grape, peach, pepper bell, potato, strawberry, tomato, etc. The images thus collated are labeled with different categories of diseases and healthy (to differentiate healthy leaves from affected ones).

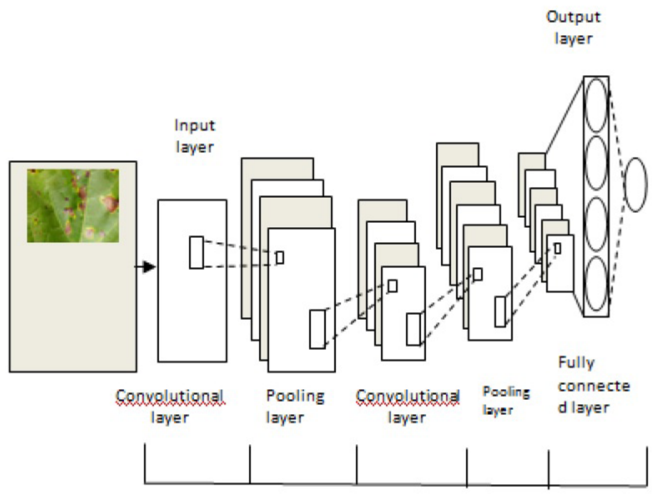

Figure 3: CNN model

V. RESULTS

Figure 4 shows the result of the SVM classifier. The classification shows that whether plant leaf is infected or not. SVM gives 80% accuracy.

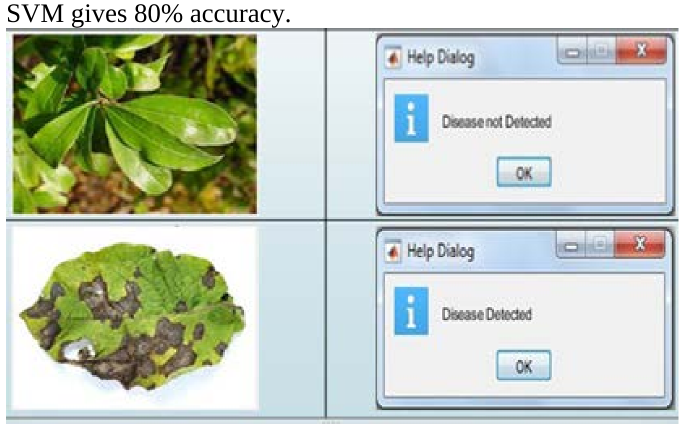

Figure 4: Result of the SVM classifier

The convolutional neural network (Alexnet) is experimented for the identification of leaf diseases with an expection of improved accuracy. The 80/20 splitting ratio is used to divide the database into two datasets i.e. training and testing. CNN detects leaf is healthy or diseased if so predicts the class of disease as well. CNN model was trained using the initial learning rate 0.001 in 10 epoch. Figure 5 presents the performance on the testing dataset for the CNN model, during their training process.

A sample confusion matrix for apple leaves is shown in Figure 6. The overall accuracy for Apples leaves is 99%. The classification accuracy of respective plants is summarized in Table 1. The overall accuracy is found as 97.71%. Figure 7. shows some classification cases of randomly selected images throughout the testing dataset i.e. result of convolutional neural network. Percentage of accuracy for corresponding plant leaves are mentioned on upper right corner of respective images.

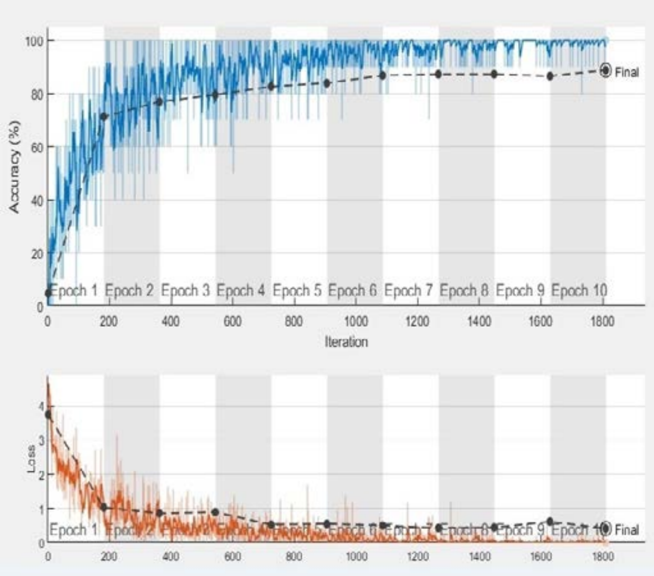

Figure 5: Performance of CNN model

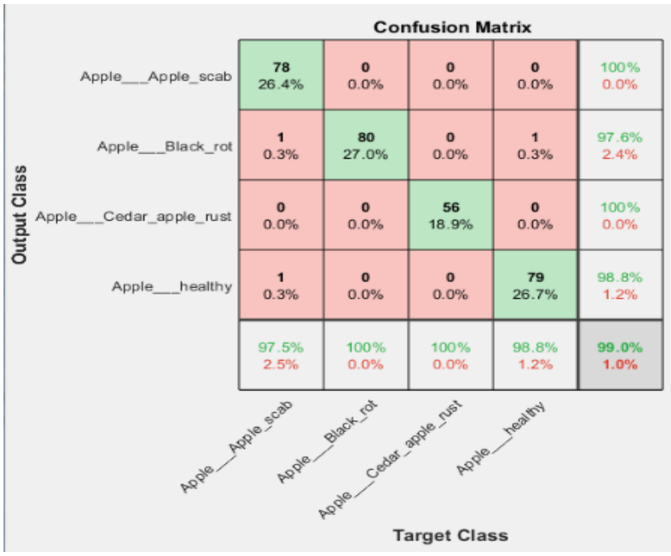

Figure 6: Confusion Matrix for Apple Leaves

VI CONCLUSION

Disease detection and identification is done by using an image processing and machine learning algorithms. The data is collected from the Kaggle website named "new plant diseases dataset" containing different images of healthy and unhealthy crop leaves, with more than 12,949 images. Leaf disease portion is segmented from the image and different features are extracted using a grey level co-occurrence matrix (GLCM). A detected portion of the leaf is recognized through SVM. SVM has given 80% accuracy. To improve on accuracy, the convolutional neural network is used for the identification of plant diseases. CNN has given 97.71% accuracy which is more than the accuracy achieved using hard coding techniques. This work will be useful in automatic identificxation of plant leaf disease and will increase agriculture production by early detection of diseases. The transfer learning and other CNN models can be evaluated to improve accuracy of detection of tomato leaves disease.

Table 1: Classification Accuracy of leaves of plants

Plant name	Classification accuracy
Apple	99.0%
Cherry	99.4%
Corn	95.8%
Grape	99.7%
Peach	97.4%
Pepper bell	99.4%
Potato	98.7%
Strawberry	100%
Tomato	90.1%

References

1. Mrs.Shruthi U, Dr.Nagaveni V, Dr.Raghavendra B K, "A Review on Machine Learning Classification Techniques for Plant Disease Detection," ICACCS, IEEE, 2019.
2. M. P. Vaishnnave, K. Suganya Devi, P. Srinivasan, G. ArutPerumJothi, "Detection and Classification of Groundnut Leaf Diseases using KNN classifier," IEEE, 2019.
3. L. Sherly Puspha Annabel, "Machine Learning for Plant Leaf Disease Detection and Classification – A Review," International Conference on Communication and Signal Processing, IEEE, 2019.
4. Gurleen Kaur Sandhu, Rajbir Kaur, "Plant Disease Detection Techniques: A Review," ICACTM, IEEE, 2019.
5. Md. Selim Hossain, Rokeya Mumtahana Mou, "Recognition and Detection of Tea Leaf's Diseases Using Support Vector Machine," International Colloquium on Signal Processing & its Applications, IEEE, 2018.
6. Monzurul Islam, Anh Dinh, Khan Wahid, "Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine," CCECE, IEEE, 2017.
7. Jobin Francis, Anto Sahaya Dhas D, Anoop B K, "Identification of Leaf Diseases in Pepper Plants using Soft Computing Techniques," IEEE, 2016.
8. Vijai Singh, Varsha, "Detection of the unhealthy region of plant leaves using Image Processing and Genetic Algorithm," ICACEA, IEEE, 2015.
9. Mrunmayee Dhakate, Ingole A. B., "Diagnosis of Pomegranate Plant Diseases using Neural Network," IEEE, 2015.
10. Sachin D. Khirade, A. B. Patil, "Plant Disease Detection Using Image Processing," IEEE, 2015.

11. Usama Mokhtar, Nashwa El-Bendary, “SVM-Based Detection of Tomato Leaves Diseases,” Springer, 2015.
12. Melike Sardogan, Adem Tuncer, “Plant Leaf Disease Detection and Classification based on CNN with LVQ Algorithm,” IEEE, 2018.

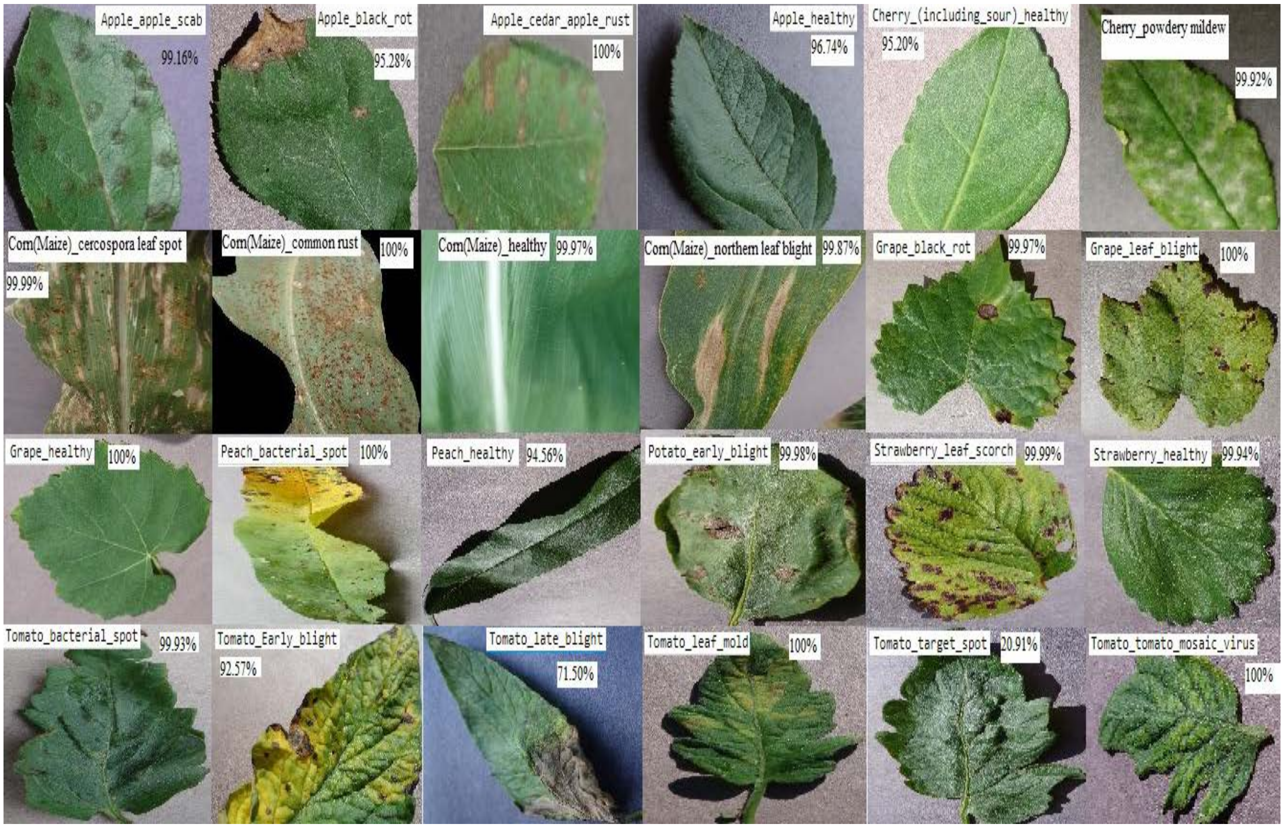

Figure 7: Result of convolutional neural network