Comparative Analysis between Convolution Neural Network and Support Vector Machine for plant disease identification

[1]Jay L. Borade, [2]Navil Rodrigues, [3]Suzanne Tuscano, [4]Rainzil Dabre

[1]Assistant Professor, Department of Information Technology, Fr. Conceicao Rodrigues College of Engineering, Bandra.
[2,3,4]Student, Department of Information Technology, Fr. Conceicao Rodrigues College of Engineering, Bandra.

Abstract: India being incorporated with agricultural tradition needs improvement in the crop yielding mechanism to increase the financial imbalance in rural areas. Diseases are the main consideration constraining crop production and they are usually hard to control. This is one of the important reasons why disease recognition in plants is essential in the agriculture field, as it is quite natural to have a disease in the plant. Existing systems for plant disease detection are mostly expensive because they are composed of hardware. The aim is to help farmers to achieve high crop yields in an inexpensive way by giving an accurate prediction using Support Vector Machine (SVM), a machine learning algorithms and image processing. Recently, deep learning, a machine learning technology gained popularity for classification of objects. Convolution Neural Network (CNN), a deep learning algorithm has great potential to classify objects accurately. The main motto of the paper is a comparative analysis of CNN and SVM classification techniques. To improve accuracy and to obtain the efficient results the paper compares different methods using different classification techniques for disease detection which are tested on the dataset obtained**.**

***Index Terms* - SVM, CNN, Plant disease, Prediction.**

I. INTRODUCTION

About 70% of India's population relies on farming [1]. Plant health monitoring is essential in today's world due to the climatic changes, which affects the growth of the plants and their productivity. Many a times, farmers make a naked eye observation to predict the disease which is not most efficient way of finding the results. In order to achieve accurate results, plant disease through some automatic technique is beneficial and these techniques reduce a large work of monitoring. The existing system is mostly composed of hardware (WSN) which is expensive [2].

The proposed solution made minimal use of hardware, which makes the system inexpensive. The main motive of the proposed solution is to help farmers achieve high crop yielding by providing accurate prediction result based on the input data provided. Prediction can be done in different ways. The proposed work aimed to classify the leaf into healthy or unhealthy. If the leaf is unhealthy the system further classified into a particular disease. The plant Village dataset [3] was used for assessing the model's capability of providing accurate prediction results. The accuracy achieved through CNN model was of 96.85 %. The results provided through our model was the first step for automatic providing the plant disease detection

.

II. METHODOLOGY

The section briefly explains the methodologies of different classification algorithms for plant disease detection.

Classification using SV*M*

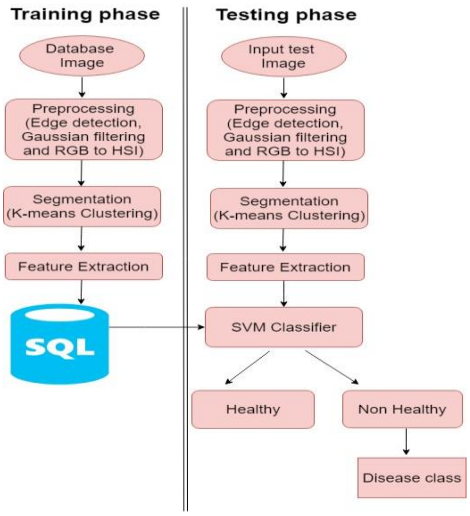

Fig 1. SVM Process flow

Stage 1: Image Acquisition

Image acquisition involves collecting the sample images of the leaves with an affected area that are essential in training the system. Images used for both training and testing the system are taken from pant Village dataset [3] and were labelled with 8 different class labels based on the diseases. In order to distinguish healthy leaves from the leaves that are affected, a separate class for healthy leaves was added.

Stage 2: Pre-processing

Image pre-processing improves the quality of the image and results in the image of interest by removing the undesirable portion from the image. Image improvisation is done by removal of background noise using the Gaussian blur filter. Further, the undesirable portion is removed by using canny edge detection. The image is then converted from RGB – HIS (Hue-Saturation-Intensity).

HUE:

Numerator = $^{1}/_{2}\,[(Red - Green) + (Red + Blue)]$

Denominator = $\left((Red - Green)^2 + \left((Red - Blue) * (Green - Blue)\right)\right)^{0.5}$

Now find theta value for

Saturation:

S = $1 - \left(3/(Red + Green + Blue)\right) * \min[Red, Green, Blue]$

Intensity

I = $^{(Red + Green + Blue)}/_{3}$

The HSI image is formed by combining the three results in a single value [4].

Figure II Pre- processed image

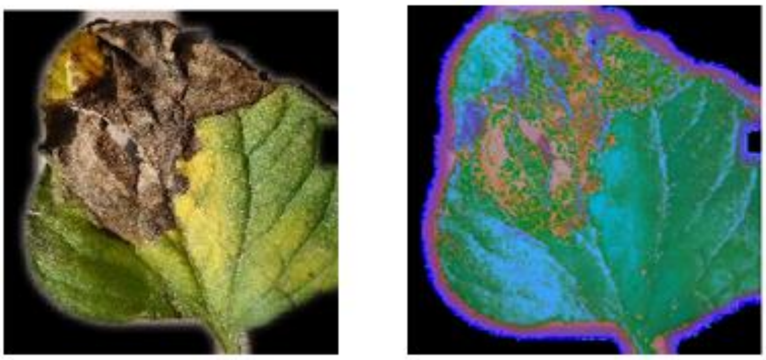

Figure 3 Conversion from RGB to HIS

Stage 3: Segmentation

K-means clustering algorithm is used to obtain different clusters. k clusters are formed, where k is the number of clusters.

Step 1: User is supposed to select the value of k. And the value of means m1, m2….mk are assigned randomly.

Step 2: Each pixel is appointed to its closest centroid (k). Distance is calculated using Euclidean Distance.

Euclidean Distance = $\sqrt{((x1 - x2)\,)^2 + \;(y1 + y2)^2)}$

The means are then updated by taking an average of the pixels in that clusters.

Step 3: The process is repeated for several no. of iterations and clusters are formed.
The cluster with the affected portion is considered for further classification.

Stage 4: Classification

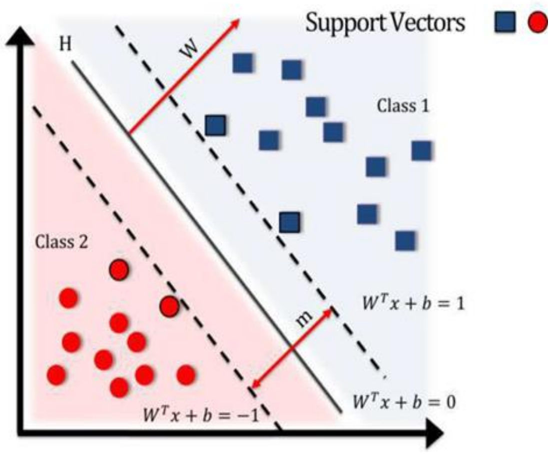

Figure 4 Support Vector Machine [5]

A Support Vector Machine (SVM) is an operative technique for classification of objects. SVM is usually related to supervising learning. To differentiate between two classes SVM uses a hyperplane which is called a decision boundary. The algorithm gradually learns a soft boundary with the aim to cluster the normal data instances using the training set. This hyperplane attempts to divide, one class containing the target training vector which is labeled as +1, and the other class containing the training vectors which is labeled as −1 [1]. SVM classifier uses these labeled training vectors to find a hyperplane that will maximize the margin of separation between the two classes

Classification using CNN

Data source:

The Plant Village [3] is an open access database of over 50,000 healthy and diseased crops images, which have extension of around 38 class labels.

From this dataset, we have selected the images of different diseases and based on which class labels were assigned. The different class labels were healthy, early blight, late blight, Bacterial spot, spider mites, Viral, Black rot and leaf mold.

Image preprocessing:

In the case of deep learning models, the step of preprocessing is very minimal. It requires only 4 basic preprocessing step [6]. In the first step, we resize all the images to standard resolution. In the second step, all the pixel values are divided by 255 to be compatible with the network initial values. In the third step sample wise normalization is performed.

The normalization is done as follows:

a) For each input x, we calculate the mean value m_x and standard deviation s_x

b) Transform the input to

$$x' = (x - m_x)/s_x$$

This is done so that the individual features of the data are distributed normally with zero mean.

In the final step, several random augmentations on images are formed like flipping, rotation, shearing and zooming are applied to the training images so as to prevent over fitting and generalize model better.

Architecture:

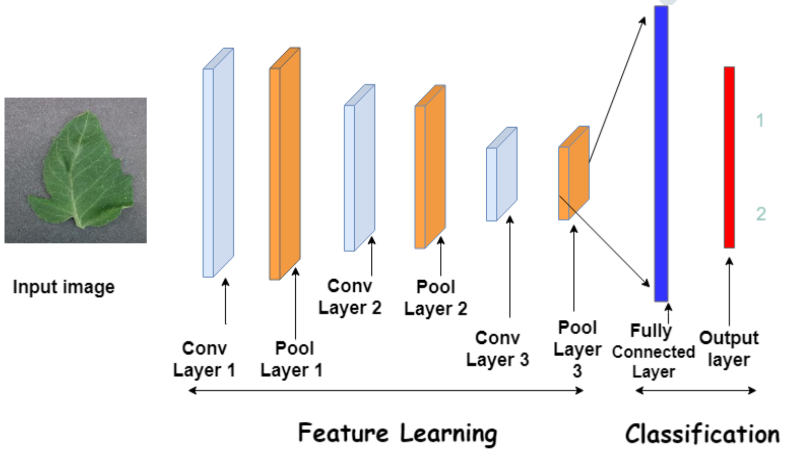

Figure 5. Convolution Neural Network Architecture

The figure above shows the flow of convolution neural network algorithm. We have trained our model with a network of 5 convolution layers. Each convolution layer has 32 filters with the size of 3 × 3, a Rectified Linear Units (ReLU) helps in training the neural network faster by applying element-wise activation function such as ax[0,x] , and all layers are followed by a 3 × 3 max-pooling layer which is useful in performing nonlinear downsampling, With the exception of the last convolution layer with 64 filters. The fully connected layer contains 64 units with a ReLU activation and is accompanied by a dropout layer with a dropout ratio of

80% which was added to reduce the overfitting of the model, the downside of using this layer was that it increased the training time period. The last fully connected layer has 8 outputs, corresponding with the 8 classes, which will compute the class score based on which the output and probability will be calculated

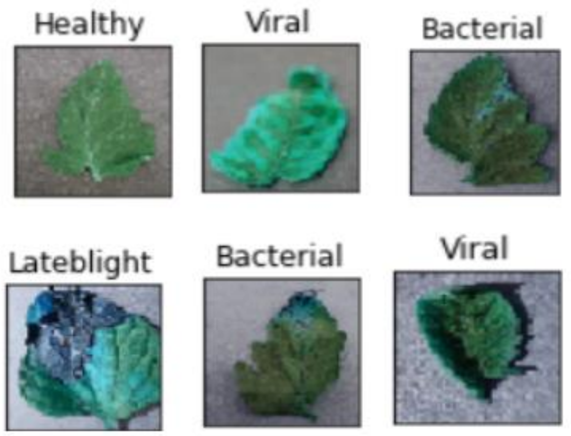

Figure 6. Test results from CNN model

In this work, we used a total of 500 images for testing and validating the prediction results of the model. The figure above shows the results of the first 6 sample test data results which were plotted using Matplotlib - a python 2D plotting library.
The accuracy we obtained was of 96.36% with 8 epochs.

Implementation

Training and testing of a plant disease detection model were done on Google Colab which provides a free cloud service based on Jupyter Notebook that supports free GPU. It comes with inbuilt libraries like Keras, pytorch, tensorflow and OpenCV.
Since it provides a shared environment for the pool of resources, every training iteration of the model took approximately 2 hours.

III. RESULTS

Parameters	[1]	Proposed system
Classifier used	SVM	CNN
Total samples	133	500
Feature used	18 (9 Texture & 9 Color)	9 (Texture)
% Accuracy	88.89%	96.36%

Table 1-Comparative Analysis between CNN and SVM

The above Table shows the comparative analysis between the proposed CNN system and previous SVM system. By comparing the features of both algorithms, we get a clear idea about the accuracy i.e. CNN gives more accuracy in classification than SVM

IV. CONCLUSION

There are several methods available for the detection of plant disease still the research phase in this domain is lacking. An accuracy of 96.36% was achieved using Convolution neural network, a deep learning algorithm and SVM showed 88.89 accuracy. The accuracy results of Convolution neural network were compared with Support vector machine - A classification algorithm and it was found that CNN algorithm certainly achieved higher accuracy results. In future work more data will be collected for identifying different disease at different stages. It can be done using multispectral camera and infrared camera.

REFERENCES

[1] Padol, P. B., & Yadav, A. A. (2016). SVM classifier based grape leaf disease detection. 2016 Conference on Advances in Signal Processing (CASP).doi:10.1109/casp.2016.7746160
[2] G. Deepika and P. Rajapirian, "Wireless sensor network in precision agriculture: A survey," *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, Pudukkottai, 2016, pp. 1-4. doi: 10.1109/ICETETS.2016.7603070
[3] https://github.com/spMohanty/PlantVillage-Dataset
[4] Radha, Suja. (2017). Leaf Disease Detection using Image Processing. Journal of Chemical and Pharmaceutical Sciences
[5] D. O. Shamkuwar1, Gaurav Thakre2, Amol R. More3, Kishor S. Gajakosh4, Muktanand O. Yewale5 "An Expert System for Plant Disease Diagnosis by using Neural Network",2018
[6] Wang, G., Sun, Y., & Wang, J. (2017). Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. Computational Intelligence and Neuroscience, 2017, 1–8. doi:10.1155/2017/2917536