Machine Learning Strategies for Predicting Crop Diseases

Vidya Ubalanka¹, Anu Jose², Dhanya Viswanath³
Department of Information Science and Engineering, CMR Institute of Technology, AECS Layout, Bengaluru-560037

Email: arnavidya@gmail.com¹, anujmattam@gmail.com², dhanyaarun0705@gmail.com³

Abstract. Prevalence of crop diseases is a major hindrance for successful crop production. These diseases can be identified in less time and more accurate using Machine Learning (ML) strategies as compared to any manual approach. Agronomy plays a key role in anticipating crop diseases at an early stage. With the advent of computer vision, plants can be classified as diseased or healthy by extracting architectural characteristics of a leaf using various image processing techniques. Support Vector Machines (SVM) classification technique is used in distinguishing between diseased and healthy leaf from the datasets that are publicly available. SVM method exhibited high fitting and predictive precision. The proposed paper is organized in various steps such as identifying the features, extraction of features using a computer vision technique known as Scale Invariant Feature Transform (SIFT), model training and testing. Predominantly, crop diseases on a larger scale are predicted by harmonizing speed and accuracy using computer vision and machine learning strategies.

1. Introduction

The Indian economy relies immensely upon the efficiency of farming, its productivity and sustainability. The key challenge to farmers these days are protecting the crops from pests and diseases. So, predicting the crop diseases plays a key role in the field of horticulture. Different modern technologies have emerged in order to minimize post-harvest processing, strengthen agricultural sustainability and maximize productivity. Automated Bots and Drones are deployed in farms to monitor and serve the crop. Digital apps are designed to identify diseases attacked in the farm by measuring the NPK (Nitrogen, Phosphorus and Potassium) values of the plant. Computer Vision, Artificial Intelligence, Machine Learning, Deep Learning and IoT technologies are adopted by developers and tech companies to boost crop yields. For technologies like Machine Learning, deep learning etc., a large volume of images that constitute complete picture agriculture environment can be considered as Dataset to address various challenges through Computer Vision

2. Literature Survey

[1] In this paper authors proposed different Machine Learning techniques to predict crop pests. They have summarized the features of machine learning techniques based on crop diseases.
[2] Authors proposed various image processing techniques for classifying the plants using pattern recognition and digital image processing.

[3] Authors proposed how colour features are represented in RGB to HIS. Seven invariant moment are taken as shape parameter. SVM classifier is used for detecting disease in wheat plant.

[4] In this paper, authors proposed different strategies for pest detection and identification on tomato plants. Computer vision and machine learning techniques are used for detection.

[5] Modifications to the SIFT descriptor is proposed by the authors in order to improve its robustness against spectral variations. It is based on fact, that edges remain well preserved in multispectral imaging for achieving better image matching results. This is implemented by boosting the contribution of local edges in the SIFT descriptor construction process.

3. Framework Proposed

To predict if a plant is diseased or healthy a framework is followed in this paper. The architecture of the system proposed is as shown in Fig1. The design flow includes the following steps:

(i) Acquisition of leaf image
(ii) Pre-processing of image
(iii) Extracting the features
(iv) Feature Matching
(v) Prediction using SVM

(i) **Image Acquisition**: Leaf images of various plants like Potato, Pepper bell, Tomato which are healthy or diseased are acquired from PlantVillage [6] dataset which is publicly available. These images are in RGB scale of 100X100X3. Fig2 shows some samples of the dataset used for experimental purposes.
Fig2. Sample images from PlantVillage dataset

(ii) **Preprocessing of image:** To reduce the complexity of the model, the RGB leaf image is converted into a grey scale of 100X100X1. Comparative to RGB image, grey scale images have less noise, increases the speed of processing and visualizing. Fig3 shows some samples of healthy and diseased RGB images of leaves converted to grey scale images.
Diseased leaf: Colour

Diseased leaf: Grey

**Fig3. RGB to grey scale conversion of a leaf**

Histogram Plot[7] provides a definition of the pixel count present in the image. Fig4 shows the histogram plot of colour and grey images with their respective histograms.

**Fig4. Histogram Plot**

(iii) **Extracting the features using SIFT:** Extracting the features of an image defines identifying the characteristics of an image. This reduces the number of resources required for describing
large dataset. Features extracted will be helpful in representing significant classification and prediction of data. Scale invariant feature transform (SIFT) is the algorithm used for feature extraction. This algorithm has two major modules, i.e., key point detection module and descriptor generation module [8].

SIFT calculation utilizes Difference of Gaussians which is an estimate of DoG. Distinction of Gaussian is acquired as the distinction of Gaussian obscuring of a picture with two diverse $\sigma$, let it be $\sigma$ and $k\sigma$. This procedure is accomplished for various octaves of the picture in Gaussian Pyramid. It is as shown in Fig5. When this DoG are discovered, pictures are looked for neighbourhood extrema over scale and space.

![Fig5. Image of Gaussian Pyramid](image1.png)

Fig5. Image of Gaussian Pyramid

![Fig6. Key points extracted using SIFT algorithm for a healthy leaf](image2.png)

Fig6. Shows the Key points extracted using SIFT algorithm for a healthy leaf.

![Fig7. Key points identified for a healthy leaf](image3.png)

Fig7. Shows the Key points extracted using SIFT algorithm for a diseased leaf.
Next comes Histogram of Oriented Gradients (HoG). The guided gradient histogram is a feature descriptor used for the purpose of object detection in computer vision and image processing. The technique counts occurrences of gradient orientation in localized portions of an image[5]. The HoG features of the healthy leaf (Fig6) and diseases leaf (Fig7) are shown in Fig8 and Fig9 respectively.

(iv) **Feature Matching**: Feature matching is a part of many computer vision applications that define similar features between the images. In this paper we use Euclidian distance between key points to identify the similar features. Fig10 shows the feature matching extraction.
Fig10. Feature Matching based on Euclidian distance

(v) **Prediction using SVM:** In the proposed work the classifier used is Support Vector Machine (SVM). It is a supervised algorithm for machine learning, used for classification and regression. SVM classifier is trained with the leaf image features extracted. It is tested for new images to predict the desired results. The input leaf images for the system taken are grape black rot, peach bacterial spot, Tomato healthy, grape leaf blight, pepper-bell bacterial spot, tomato bacterial spot and tomato late blight. The results conclude the prediction of a crop if its healthy or diseased.

Two types of SVM kernels are used for classification. They are Linear Kernel and the Radial Basis Function (RBF) kernel. Linear Kernel is a commonly used kernel, which can be used when the dataset can be linearly separable by a single line. RBF is used for non-linear classification of dataset. It is also known as kernel trick.

4. **Results and Discussion**

In this experimental study, 100 samples for training and another 100 samples for testing has been used in the system. The attributes are normalized to ensure that they all have same scale. Scaling used is Min-Max scaling. The formula for it is as follows:

\[ X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \]

The classification for training data is as shown in Fig11. The results of SVM model on testing data is shown in Fig12. The lines that are dividing defines the maximization of classification.
Fig11. SVM model Training Results

Fig12. SVM model Testing Results

**Linear Kernel Results:** Fig13, Fig14, Fig15 shows the support vectors plotted on the hyperplane using SVM linear Kernel

Fig13. Support Vectors
Fig 14 shows support vectors for different sizes of dataset i.e., N=50,100, where N is the No. of datapoints

Fig 15. Hyperplanes plotted with different ‘C’ values

In SVM, parameter ‘C’ known as regularization parameter determines the influence of misclassification on the objective function. Larger values of ‘C’ leads to a small margin, ignoring the hard constraints. Smaller values of ‘C’ leads to large margin, ignoring easy constraints. Fig 15 shows hyperplanes using different ‘C’ values.

**RBF Kernel Results:** Fig 16, Fig 17, Fig 18 shows the support vectors plotted on the hyperplane using SVM RBF Kernel
Fig 16. Support Vectors

5.

Fig 17 shows support vectors for different sizes of dataset i.e., N=75,100, where N is the No. of datapoints
5. Conclusion

In this proposed work, Computer Vision and Supervised machine learning algorithms are used for predicting and classifying if a crop is healthy or not. The computer vision technique used is SIFT. SVM is a supervised machine learning algorithm used for classification and regression using two different kernels i.e., Linear and RBF. The dataset used are the images of leaves obtained from a publicly available dataset known as PlantVillage dataset. The results show that RBF kernel provides more accurate results than Linear Kernel if the dataset is non-linear. Optimum results are obtained with very less computations.

6. References

[1] Yun Hwan Kim, Seong Joon Yoo, Yeong Hyeon Gu, Limb Dongil Han, Wook Baik “Crop Pests Prediction Method Using Regression and Machine Learning Technology: Survey”
[2] Goutum Kambale, Dr. Nitin Bilgi “A Survey Paper on Crop Disease Identification and Classification Using Pattern Recognition and Digital Image Processing Techniques”
[3] Tian, Chunjiang Zhao, Shenglai Lu and Xinyu Guo,” SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases,”.
[4] Aitor Gutierrez, Ander Ansuatagi, Loreto Susperregi, Carlos TubioIvan Rankič, and Libor Lenža “A Benchmarking of Learning Strategies for Pest Detection and Identification on Tomato Plants for Autonomous Scouting Robots Using Internal Databases”.
[5] Saleem S., Sablatnig R. (2013) A Modified SIFT Descriptor for Image Matching under Spectral Variations. In: Petrosino A. (eds) Image Analysis and Processing –ICIAP 2013. ICIAP 2013. Lecture Notes in Computer Science, vol 8156. Springer, Berlin, Heidelberg.
[6] Hughes, David P. and Marcel Salathé. “An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing.” ArXiv abs/1511.08060 (2015): n. pag.
[7] Caglayan, A., Guclu, O., & Can, A. B. (2013, September). “A plant recognition approach using shape and color features in leaf images.” In International Conference on Image Analysis and Processing (pp. 161-170). Springer, Berlin, Heidelberg.
[8] M. Yuvaraju, K. Sheela Sobana Rani, “Feature Extraction of Real-Time Image Using Sift Algorithm”, International Journal of Research in Electrical & Electronics Engineering Volume 3, Issue 4, October-December, 2015, pp. 01-07

[9] Tony Lindeberg (2012) Scale Invariant Feature Transform. Scholarpedia, 7(5):10491

[10] Wikipedia contributors. "Histogram of oriented gradients." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 14 Jan. 2020. Web. 31 Jan. 2020.