Diagnosis of Tomato Plant Diseases using Random Forest

Meghana Govardhan | Veena M B

ML IA-2 Report - Group1

Group Members	Rollno.
Aayush Malde	1911090
Aditya Malwade	1911091
Rahul Panchal	1911097
Bhavin Patel	1911100

Introduction

Considering the worldwide scenario, plant diseases are hazardous to food security. They also have catastrophic penalties on healthy crops, hence impacting the agriculturists whose livelihoods depend upon. In a developing country such as India, where farmers with small lands cultivate the majority of the agronomical produce, pests and disease are reported to cause sizable damage to the yield.

Based on the agricultural importance, farmers pick crops and the related pesticide to improve the growth of the plant. The irregular usage of the pesticides and the nutritional problems increase the damage in the leaves.

Past methodologies of application of pesticides are enhanced by integrated pest management approaches.
To prevent crop loss due to diseases, several attempts have been made. Identifying the plant diseases when it appears and taking necessary action forms the vital part of the well organized disease management system.

The optical review of the signs of the plant disease indicated an immense level of complication. This complex along with the pathological problems of the widely cultivated plants, repeatedly resulted in senior agronomists to misdiagnose explicit diseases and thus brought about inaccurate inferences and analysis. In this project, the plant leaf diseases are detected using the Random Forest algorithm of ML.

Importance of the Topic

Plant diseases are responsible for economic loss in the agricultural industry, as they destroy the crops. Although pesticides have been used to modify crop production, the excess amount of use of pesticides affects the environment negatively. Hence detection of diseases and differentiating it from nutritional deficiency has a considerable impact in deciding the requirement of pesticides.

The conventional methods for plant diseases identification involve tedious chemical processes to be carried out in the laboratory and is also time consuming. This paper presents an automated system for identification and classification of plant diseases using Machine Learning (ML) and image processing technique.

Literature Review

- **Disease Detection on Ginseng Leaves using K-Means Clustering:** Diseases like black spot, anthracnose and Cylindrocarpon destructans always appear on ginseng leaves. The traditional way to distinguish the disease in farmland is checking by planter in person. This way takes more time and energy. In this paper, according to which cicatrices are different with normal leaf in color, they use K-means clustering algorithm and combine the difference between a and b components in Lab color space to discern the cicatrices in leaf image. Firstly, they can distinguish cicatrices and normal leaves. Secondly, they take out useless parts like petioles. Lastly, they calculate the ratio of area of cicatrices and normal leaf. The ratio can be used as reference of the ginseng disease and its degree. The result of the experiment

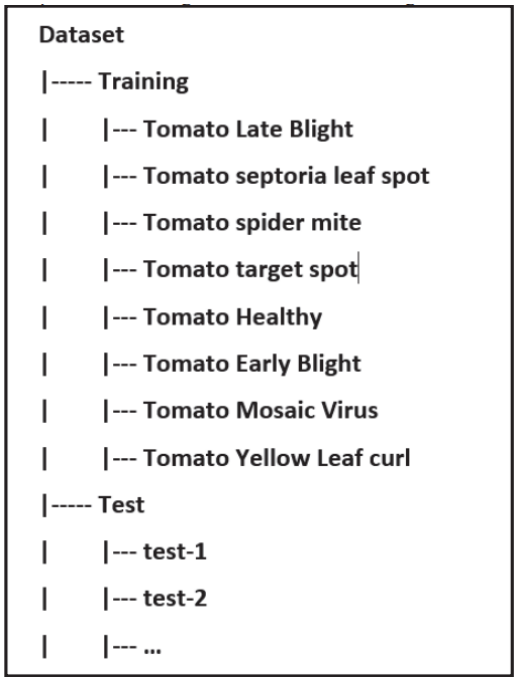
shows that the K-means clustering algorithm can segment the leaf image, and the segmentation is based on the difference of color, then they can calculate the ratio of pixels in each part, and judge the degree of disease.

- **Tomato crop disease classification using pre trained deep learning algorithms(AlexNet, VGG16):**
  The wide scale prevalence of diseases in tomato crops affects the production quality and quantity. In order to counteract the problem, early diagnosis of diseases using a fast reliable nondestructive method will benefit the farmers. In this study images of tomato leaves (6 diseases and a healthy class) obtained from PlantVillage dataset are provided as input to two deep learning based architectures namely AlexNet and VGG16 net. The role of number of images and significance of hyperparameters namely minibatch size, weight and bias learning rate in the classification accuracy and execution time have been analyzed.

- **Automatic Detection of Tomato Diseases and Pests Based on Leaf Images(VGG16):**
  There are many species of tomato diseases and pests, and the pathology of which is complex. It is difficult and error-prone to simply rely on manual identification. For the ten most common tomato diseases and pests in China, This paper explores the detection algorithms on leaf images and constructs the convolution neural network model to detect tomato pests and diseases based on VGG16 and transfer learning. The detection model is trained with Keras/TensorFlow deep learning framework and achieves an average classification accuracy of 89%.

Methodology and Dataset

The methodology implemented in the paper is extraction of features of the leaves of tomato plant followed by applying the ML algorithms to accurately predict the presence of the disease in the plant leaf. As a first step, a publicly available dataset of diseased and healthy leaves is used. A machine algorithm is used to classify the healthy and diseased plant leaves. The classification of healthy and diseased leaves of tomato plants is the main aim of the project. The system is trained with 8 classes of tomato leaf images obtained from the PlantVillage dataset. The same samples are divided into training and testing sets.



The project is divided into two phases. In the first phase, feature extraction is done on all the images and stored in the feature array. During the second phase, Machine Learning algorithms such as :

- K Nearest Neighbors
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM) etc.

are used for training the dataset.

The three global feature descriptors used in the project are shape, color and texture. These features are extracted using the below method:

- **Haralick Texture:** The aspects of an image that are rotationally invariant falls under Texture analysis. The GLCM co-occurrence matrices are rotated at angles such as 0, 45, 90, and 135 degree and are calculated and summed for the texture estimation. It quantifies the texture of the leaf image.

- **Color Histogram:** This is the most prominent technique for portrayal of the color features. It gives the extent of the pixels of each color in a picture.
- **Hu Moments:** These are seven sets of numbers determined utilizing central moments that are invariant to image changes. The initial 6 moments are invariant to reflection, translation, scale, and rotation, while the sign of the seventh moment changes for image reflection. It evaluates the shape of the leaf image.

The Machine learning algorithm that best suits the model with highest accuracy selected in training the dataset and prediction of images unseen by the algorithm during training. The principal reason for utilization of Cross-validation is to evaluate the aptitude of ML model on unseen information. That is, to utilize a restricted example to gauge how the model performs all in all when used to make forecasts on information unseen during the training of the model. It is a prominent technique since it is easy to understand, simple to utilize and has lower inclination contrasted with other techniques.

Expected Results

The proposed system in the paper aims to identify and classify tomato diseases using the Random Forest Machine Learning algorithm. The images are preprocessed and features are extracted which are used for training the ML algorithm. The tomato diseases which the model aims to classify are::

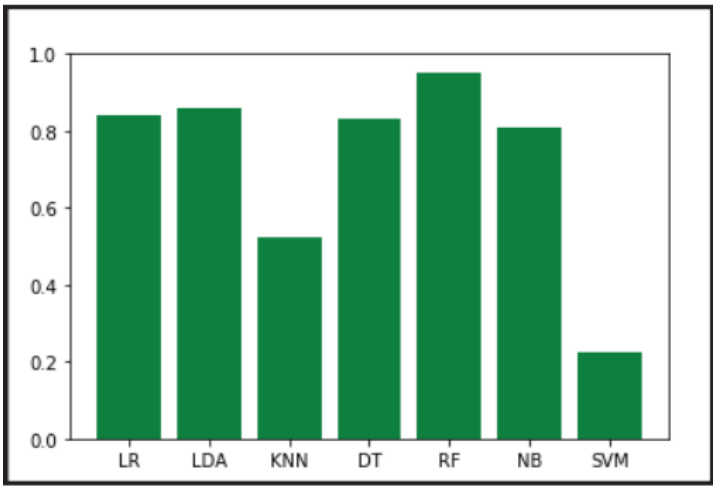
- Early Blight,
- Late Blight,
- Septoria Leaf spot,
- Spider mite,
- Mosaic Virus,
- Yellow leaf curl virus,
- Target spots are diagnosed by the system.

Outcome of the Experiment

The ML algorithms are evaluated using k-fold Cross validation. The accuracy of the different algorithms are shown in the below figure.

```
Logistic Regression : 84.011747
Linear Discriminant analysis : 85.935904
K Nearest Neighbours : 52.385087
Decision Tree : 82.992850
Random Forest : 95.234934
Support Vector Machine : 22.337845
```

It can be inferred from the above image that Support Vector Machine exhibited the lowest accuracy of 22% and Random Forest exhibited the maximum accuracy of 95.2 %. The accuracy comparison of ML algorithms is visualized using Bar Plot

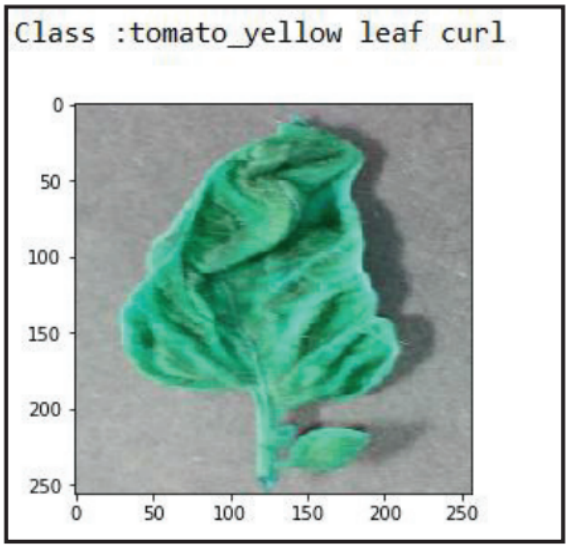


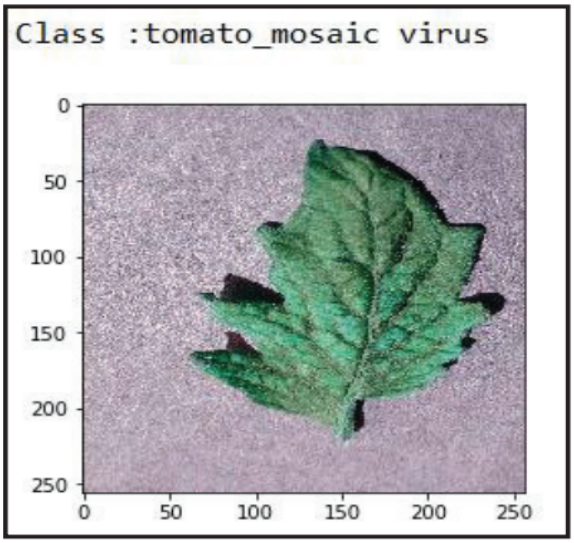
The experimental results show that almost all test images give the best discrimination among the classes of images. The performance of the system is evaluated to be 95%.

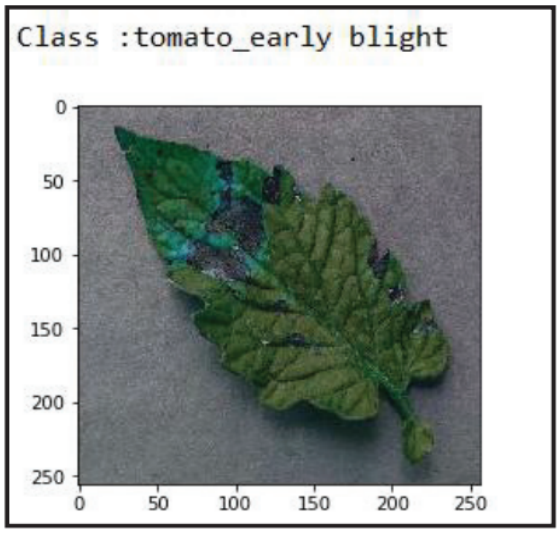
Plant diseases can be hard to recognize as indications seem to be like numerous nutritional deficiencies, ex. Mosaic Virus in Tomato leaves. They differ upon the age of the plant when contamination happens, henceforth ML assumes an imperative job in

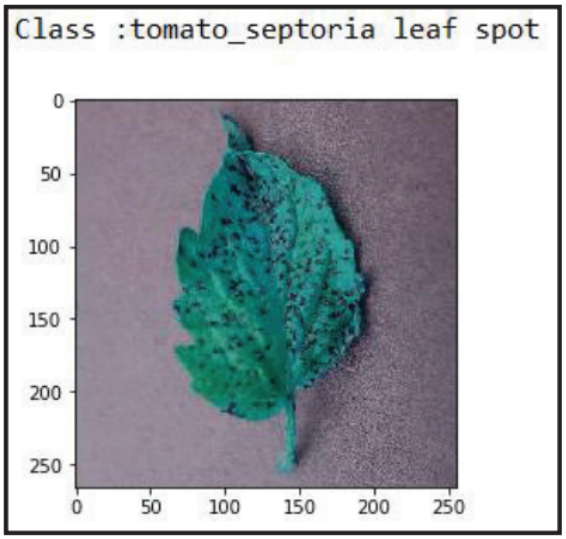
distinguishing the diseases in the leaves. Any leaf diseases can be detected just by training the ML algorithm with a dataset of images.

Prediction of leaf diseases by the model:









References

- Alvaro Fuentes, Sook Yoon, Sang Cheol Kim, and Dong Sun Park, “A Robust Deep-LearningBased Detector for Real-Time Tomato Plant Diseases and Pests Recognition”, Sensors (Basel), DOI: 10.3390/s17092022, September 2017

- Yuan-yuan Liu, Shi-yu Zhou, Jia-hui Sun “Detection of Ginseng Leaf Cicatrices Base on K Means Clustering Algorithm”, International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), DOI:10.1109/cispbmei.2017.8302051, Feb 2018

- Aravind Krishnaswamy, Rangarajan, Raj Purushothaman, AniirudhRamesh,”Tomato crop disease classification using pre trained deep learning algorithm”, Science Direct Procedia Computer Science, vol. 133, pp. 1040-1047, 2018

- Jia Shijie, Jia Peiyi, Hu Siping, sLiu Haibo, “Automatic Detection of Tomato Diseases and Pests Based on Leaf Images”, IEEE Chinese Automation Congress, DOI: 10.1109/CAC.2017.8243388, January 2018

- Monzurul Islam, Anh Dinh, Khan Wahid, Pankaj Bhowmik, “Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine”, IEEE 2017 Canadian Conference on Electrical and Computer Engineering (CCECE), DOI: 10.1109/CCECE.2017.7946594, June 2017

- https://hgic.clemson.edu/factsheet/tomato-diseases-disorders/