Plant Disease Detection using Machine Learning Algorithms

A. Rama Rao, Anji Reddy V, G.V. Narasimha Raju

Department of CSE, Lendi Institute of Engineering and Technology,

Abstract: One of the most exciting research areas in agriculture is the identification of plant diseases from images. Today, the term "machine learning" is a hot topic. It is possible to apply the principles of machine learning to a wide range of fields. Plant disease detection can also benefit from Machine Learning and Deep Learning. For the detection and classification of plant diseases, this paper uses several types of plant leaves as a model. This model framework is created after point-by-point exploratory examination of different methods utilized in picture preparing tasks. We considered three different experimental setups for potato disease detection, Pepper Bell plant disease detection and plant images classification. Potato plant sicknesses specifically Early Blight, Late Blight are detected successfully. Pepper Bell Bacterial Leaf Spot detected with good accuracy rate. To enable accurate extraction of features, we propose Convolutional neural Networks achieved an accuracy of 98.8%, 98.5%, 92.3% for potato disease detection, pepper bell disease detection, plant disease classification respectively. All the experiments are done using python.

Keywords: Machine Learning, Deep Learning, Plant Disease Detection, Kaggle, Python

1. Introduction:

Agriculturists in provincial areas may believe it is difficult to distinguish between diseases that may be present in their fields. Going to the agriculture office to diagnose the plant disease is not an easy operation every time. Going to the agriculture office to diagnose the plant disease is not an easy operation every time. Our main goal is to identify the disease that has been introduced into a plant by observing its shape using image processing and machine learning. A variety of current technologies have arisen to reduce post-harvest processing, strengthen agricultural sustainability, and increase output. Some laboratory approaches, such as hyper spectral techniques, gas chromatography, mass spectrometry, and thermography techniques, are used to identify the plant disease. However, these methods are not cost-effective and take a long time to complete. Machine Learning and Deep Learning are widely using from past few years in agriculture field. The advantage of using Machine Learning and Deep Learning is human intervention can be reduced and efficiency of the prediction system also improved easily. But, the important aspect for applying machine learning/deep learning is selection of dataset. If, dataset has any errors then ML/DL algorithms may not perform well. We used a dataset from standard dataset repository named "Kaggle" in our work.

2. Literature Survey:

Machine Learning for Plant Leaf Disease Detection and Classification was reviewed by Annabel [1] et al. They gave an overview of plant disease detection and classification algorithms that used several machine learning techniques. Sun et al. [2] used a variety of deep learning models to solve plant identification difficulties. Pantazi [3] and colleagues Image feature analysis was used to detect disease in crop species. For Leaf Disease Grading, Sanjeev[4] et. al employed Machine Vision and Fuzzy Logic. For Crop Disease Detection, Anuradha Badage [5] et al used machine learning approaches. They presented their concept in two stages. The dataset is trained in the first step. The second phase focuses on crop monitoring and disease identification utilizing Canny's edge detection method. For cotton leaf disease identification, Sarangdhar[6] et al used a linear regression model. Kranthi[7] and colleagues suggested a machine learning-based approach for Plant Disease Prediction. They used decision trees, Naive Bayes, and Artificial Neural Networks to detect plant diseases. Bauer[8] et al. classified illnesses of sugar beet leaves using multispectral pictures.

3. Proposed Method:

Kaggle [9] provided us with a plant village dataset. Certain procedures must be followed to determine if a leaf is healthy or diseased. Pre-processing, Feature extraction, classifier training, and classification are all examples of this. A number of image-preparation steps were required, such as

reducing the size of all images to a uniform size. We reduced all images to 64x64 size. Later images are converted into array representation. Next, we applied some pre-processing techniques on the arrays using keras predefined library. Then, splitting data into training and testing sets followed by applying CNN classifier. Then evaluation of classifier is done by giving some images as input. The proposed architecture was shown in Figure-1. The original dataset is divided into three datasets. The first dataset contains 2152 images of Potato. These images used for potato image disease detection through potato leaf images. The second dataset contains 2475 images of Pepper Bell with two different classes. These images used for Pepper Bell image disease detection through pepper bell leaf images. The third dataset contains 20,638 images of plant leaves with 15 different classes. These images used for plant disease image classification using leaf images of various plants. The third dataset is nothing but entire dataset. The dataset details are shown in Table-1.

Table-1:
Dataset Description

Plant Name	Image count
Tomato_Septoria_leaf_spot	1771
Potato-Late-blight	1000
Potato-healthy	152
Potato_Early_blight	1000
Tomato_mosaic_virus	373
Tomato_Target_Spot	1404
Tomato-Early-blight	1000
Tomato_YellowLeaf_Curl_Virus	3209
Pepper-bell-Bacterial-spot	997
Pepper-bell-healthy	1478
Tomato_Bacterial_spot	2127
Tomato_Late_blight	1909
Tomato_Leaf_Mold	952
Tomato_healthy	1591
Tomato_Spider_mites_Two_spotted_spider_mite	1676

Figure-1. Proposed Method

4. Experimentation and Results:

All the experiments are done using python. The reason for selecting python is that python provides rich packages for deep learning approaches. We used to google colab for our experiments. Next, we applied CNN for all three datasets created.

4.1.CNN architecture for dataset-1:

The architecture of CNN for first dataset is shown in figure-2.It contains three Convolutional layers followed by dropout layer.The architecture of CNN was shown in Figure-2.

Figure-2. CNN architecture for dataset-1&dataset-2

Three CNN layers used in this model. The purpose of CNN layer is applying convolutional operation. 150,80,50 neurons used in three layers. After CNN layer, an activation function Relu is applied to make data nonlinear. Later, maxpooling layer of 2by 2 applied. The optimizer used in our model is adam. The loss function used is categorical entropy. The output function used in the model is softmax.

4.2.CNN architecture for Dataset-2:

It contains three Convolutional layers followed by dropout layer. We used same network which used for dataset1.

4.3.CNN architecture for dataset-3:

It contains three Convolutional layers followed by dropout layer. But the number of layers used are different. Here, we used more number of neurons, as dataset contains more images.

Figure-3. CNN architecture for dataset 3(classification)

4.4. Comparison of results:

After applying CNN, we need to make a prediction about it. Basically, prediction means identifying given leaf with correct label. The accuracy obtained from three experiments is shown in Table-2 and Figure-4.

Table 2. Comparison of different models

S. No	Model	Dataset		
		Potato	Pepper	Tomato
1	KNN	91.20%	91.40%	86.70%
2	RFM	97%	94.70%	89.50%
3	CNN	98.80%	98.50%	98.50%

Figure-4. Accuracy.

The above comparison states us the 3 different models KNN, RFM and CNN while comparing with the three data sets i.e. Potato, Pepper and Tomato Plant leaf Considering the previous algorithms results i.e., KNN and RFM Proposed Algorithm CNN gives the more Accuracy, so I Considered CNN Algorithm because it gives more Accuracy than KNN AND RFM.

5. Conclusion:

In this work, we proposed a deep learning CNN model for plant disease detection. We created 3 different datasets from plantvillage dataset and applied CNN with different layers. For first dataset (Potato), we achieved an accuracy of 98.8%. For second dataset (Pepper Bell), we achieved an accuracy of 98.5%. For third dataset (whole dataset), we achieved an accuracy of 92.3%. Along with these we also Add a Reference to the Previous Model [10] which Determines the Plant Disease Detection using Machine Learning Models, we Applied the Random Forest Algorithm for 3 Different dataset first dataset (Potato), we Achieved an Accuracy of 97%. For second dataset (Pepper Bell), we achieved an accuracy of 94.7%. For third dataset (Tomato dataset), we Achieved an accuracy of 89.5%. We also Compared with K-Nearest Neighbor Algorithm also, in the above Comparison Table Experimental results shown that our model achieved promising results when compared to traditional machine learning models.

References:

- [1] L.S.P Annabel, T. Annapoorani and P. Deepalakshmi, "Machine Learning for Plant Leaf Disease Detection and Classification – A Review", International Conference on Communication and Signal Processing, April 4-6, 2019.
- [2] Sun, Yu, Yuan Liu, Guan Wang, and Haiyan Zhang, "Deep learning for plant identification in natural environment," Computational intelligence and neuroscience, 2017.
- [3] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers," Computers and Electronics in Agriculture, vol. 156, pp. 96-104, 2019.
- [4] Sanjeev S Sannakki, Vijay S Raj purohit, V B Nargund, Arun Kumar R, Prema S Yallur. 2011. Leaf Disease Grading by Machine Vision and Fuzzy Logic. Gogte Institute of Technology.
- [5] Anuradha Badage "Crop Disease Detection using Machine Learning : Indian Agriculture", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 09 | Sep 2018.
- [6] A.A.Sarandghar , V. R. Pawar , " Machine Learning Regression Technique for Cotton Leaf Disease Detection and Controlling using IoT", International Conference on Electronics, Communication and Aerospace Technology ",ICECA 2017.
- [7] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers," Computers and Electronics in Agriculture, vol. 156, pp. 96-104, 2019.

- [8] Bauer, S. D. Korc and F. Forstner, "Investigation into the classification of diseases of sugar beet leaves using multispectral images," Wageningen Academic Publishers, pp. 229-238, 2009.
- [9] <https://www.kaggle.com/emmarex/plantdisease>
- [10] Md Asif Iqbal and Karnul Hasan Taludker. "Detection of Potato Disease Using Image Segmentation and Machine Learning"