International Journal of Computer Applications (0975 – 8887)
Volume 185 – No. 12, May 2023

Classification of Fruit Leaf Disease using Deep Learning and Image Preprocessing Techniques

Nischay Gupta
Pimpri Chinchwad College of Engineering
Pune, India

Sharen Sharma
Pimpri Chinchwad College of Engineering
Pune, India

Shaunak Tripathi
Pimpri Chinchwad College of Engineering
Pune, India

Jaydeep Borde
Pimpri Chinchwad College of Engineering
Pune, India

Kapil Vhatkar
Pimpri Chinchwad College of Engineering
Pune, India

ABSTRACT
One of the primary factors determining any nation's development is agriculture. Approximately 65% of the population in India depends on agriculture. The crops get a variety of illnesses as a result of diverse seasonal circumstances. These illnesses first harm the plant's leaves before spreading to the whole plant, which has an impact on the kind and amount of crops that may be grown. The enormous number of plants in the farm makes it exceedingly challenging for the human eye to identify and categorize each plant's ailment. Additionally, since these diseases may spread, it is crucial to diagnose each plant. Therefore, in this study, we provide an artificial intelligence (AI) based automated plant leaf disease detection and classification system for rapid and simple disease detection, classification, and application of necessary treatments to cure that illness. Our strategy is to boost crop production in the agricultural sector. Several procedures, including picture collecting, image preprocessing, segmentation

General Terms
Picture collecting, image preprocessing, segmentation.

Keywords
Plant Leaf disease detection, Classification, Image Preprocessing, Segmentation, K means Clustering.

1. INTRODUCTION
Any country's economic development depends heavily on agriculture. It is a sector that has a significant impact on a nation's GDP. About 16% of India's GDP is contributed by the agricultural sector. The kind and amount of crops grown are influenced by a number of variables. These plants are susceptible to a variety of illnesses because of local climatic variations and other factors. And if these illnesses are untreated, they might result in significant losses. Around 15 to 25 percent of crops in India are lost to diseases, pests, and weeds. We might also think of the occurrence in Georgia (USA), when plant diseases caused a loss of around 540 USD in 2007.

It is exceedingly challenging for the human eye to accurately identify and categorize each plant since cultivation areas are relatively wide and include a huge number of plants. And doing so is crucial since the illness may spread from just one sick plant. Additionally, the majority of farmers lack sufficient awareness about these illnesses and their effective treatments. The expense of hiring specialists may be prohibitive, and pesticide usage without proper expertise will damage the environment. Because of this, we have created an AI-based solution to address this issue.

The autonomous plant leaf disease detection and classification model's performance will mostly depend on accuracy and speed. The proposed methodology would assist farmers in accurately identifying and categorizing the illness by scanning the leaf and will warn farmers of the disease before it spreads. The model may be broken down into four key parts or phases. In the first, we compile a dataset of various plant leaves, both unhealthy and sick. These are all going to be colored graphics. After removing the noise from the photos in the second stage, the framework for the color modification of the photographs is created. Using existing clustering algorithms, we segment the photos in the third stage. This step is carried out to make it simple to remove the leaf-filled foreground. Now, a collection of images of leaves on a black backdrop is available. The last phase involves training and comparing several machine learning and deep learning algorithms for accuracy, including logistic regression, KNN, SVM, and CNN. The method that performs the best during both training and testing is then taken into consideration.

2. LITERATURE SURVEY
Paper by Chutinan Trongtorkid, Part Pramokchon published about Expert System for Diagnosis Mango Diseases Using Leaf Symptoms Analysis . This research describes the development of an intelligent system for detecting plant pests in the Barracuda mango (Nam-Dok Mai), one of Thailand's key agricultural export products. However, since Thailand is a tropical region, its climate gives rise to a range of plant pathogens that have an effect on mango trees' ability to grow. Due to an agriculturalist's knowledge of the clear identification of plant pathogens, most varieties of agricultural production are reduced. Furthermore, there is no system for offering suggestions on the best approach to avoid or cure the sickness that occurs on their farm. Their remedies of infected plants suffer significantly as a result. Therefore, this approach has been designed to help an agriculture in quickly identifying the diseased plant and resolving the issue. The farmer should possess the software needed for the identification of a plant diseases.[1].

Another paper named Vijai Singh,Varsha,Prof. A K Misra published about Detection of unhealthy region of plant leaves using Image Pro- cessing and Genetic Algorithm.

Agriculture productivity is an important part of the Indian economy. This is among the factors why plant detection of disease is crucial in the industry of agriculture, since the presence of illness in plants is extremely common. If necessary precautions are not followed in this region, plants suffer major consequences, which have an impact on the quality, quantity, or productivity of the corresponding products. The use of an automated method for plant disease detection is advantageous since it reduces the amount of effort required to monitor large crop farms and can identify disease signs as soon as they develop on leaf tissue. This article provides an overview of several diseases classification strategies that may be used to the automated detection and categorization of plant leaf diseases, in addition to an algorithm for picture segments. Using an evolutionary algorithms, image segmentation is an essential step in the disease detection process for plant leaf diseases.[2].

Dr.K.Thangadurai, K.Padmavathi published about "Computer Vision image Enhancement For Plant Leaves Disease Detection" .Compared to the original taken photographs, enhanced images are of greater quality & clarity. Remotely sensed data, pattern classification, and plant leaf disease identification are just a few real-time applications that employ computer vision picture enhancement (Color conversion and Histogram equalization). RGB images make up the original acquired photos. Primary colour's are combined to create RGB visuals (Red, Green and Blue). Due to this color's 0 to 255 range, applications are challenging to implement. Grayscale graphics only have a range between 0 and 1, not 255. So many applications can be implemented quickly. The photographs' clarity is improved via histogram equalization. To identify plant leaf disease, histogram equalization and grayscale conversion are applied.[3].

3. ALGORITHM

The model that is proposed by us to detect and classify the infected plant leaves consists of 4 phases.Those phases are: -

- Dataset Collection
- Image Preprocessing
- Segmentation
- Selection of Classifier

3.1 Data Collection

First off, the leaf photos were gathered from internet sources including GitHub, Kaggle, and other image datasets that have 20,000 photographs split into 19 distinct categories. The collection includes both healthy and diseased leaves, and it includes information on diseases like black rot, rust, bacterial spot, early blight, late blight, leaf scorch, target spot, and mosaic virus that affect various crops including apple, peach, grape, strawberry, corn.

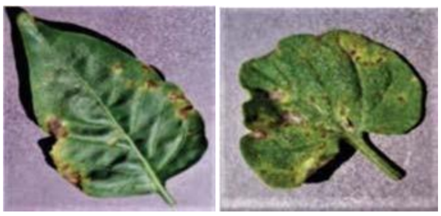

Fig 1: Sample Images from Dataset

3.2 Image Processing

In order to speed up calculations, photos are downsized to lower pixel sizes in this stage. Some noise may be seen in the captured photos. Utilizing filtering methods like Gaussian Blur, this noise is eliminated. The photographs that follow are in RGB format, which is inappropriate for further study since RGB format cannot distinguish between image intensity. As a result, it is transformed to HSV, a different color space that distinguishes between color and intensity.

Fig 2: Images after Preprocessing

RGB to HSV conversion: -

First R, G, B values are divided by max value that is 255

So, R' = R/255

G' = G/255

B' = B'/255

Then C_{max} = max (R', G', B')

C_{min} = min (R', G', B') ẩ = C_{max} – C_{min}

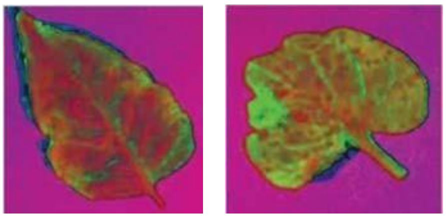

Fig 3: Images converted to HSV color space

3.3 SEGMENTATION

Images are segmented in this stage in order to distinguish the leaves from the backdrop. K-means clustering is used for segmentation, with one cluster centre for the background and one for the foreground. K-means clustering is an unsupervised learning approach used to arrange or cluster the datapoints according to their similarity into a predetermined number (k) of clusters or groups.

K –Means algorithm works as follows: -

Set of inputs: - number of clusters(k), set of datapoints

1. Put k centroids in random location in space.
2. Repeat the following steps until none of cluster location changes: -

a) For every datapoint x_i -

i. Find nearest centroid c_j by argmax D (x_i, c_j) where D = ¥ (™(x_i-y_i))

ii. Assign x_i to the cluster with nearest centroid.

b) For every cluster, new centroid is assigned by taking mean of all data points assigned to that cluster.

Fig 4: Images after K means Clustering

This is a categorization issue since we must categorize the sort of illness present on the plant's leaf. We may thus use a variety of machine learning and deep learning methods on this dataset.

We have decided to start with low complex algorithms and increasing the complexity level in order to increase accuracy of the model. We have selected four classifiers namely – logistic regression, KNN, SVM and CNN.

Logistic Regression

It is the most simple categorization algorithm on the market, yet it is still capable of producing useful results.

The logistic regression uses a sigmoid logistic function to limit the output to a range between 0 and 1. The model's accuracy on the testing set after training on the training set is 66.4%, which isn't too awful when you consider how complicated the technique is and how many classes there are in the dataset.

Knn(K Nearest Neighbour)

The approach is applicable to both classification and regression issues. The method is really straightforward and easy to use. Here, we display every data point in space, and then we calculate the distance between every data point and the input data point to identify the k closest neighbours of the data point that we wish to categorise. Next, the k closest datapoints are selected, and their classes are obtained. The projected class of input is then the class with the highest occurrence. The accuracy of the KNN model on our dataset was 54.5%.

SVM(Support Vector Machine)

Another machine learning method we utilized to categorize the illnesses is SVM. In this approach, all points are spatially mapped such that gaps may be used to separate points of various classes. So that the border may separate them, the gap should be as big as feasible. The extreme data points of the classes are known as support vectors, and this border is referred to as a decision boundary. For nonlinear datasets, kernel techniques are used. There are four types of kernels available: RBF, polynomial, linear, and nonlinear.

In our scenario, the svm performed badly since it only provides accuracy of around 53.4% when employing a linear kernel.

CNN(Convolutional Neural Network)

We have tried many deep learning models to categorize illnesses, but this one is by far the most complicated. Due of its complexity, it also needs strong computing power. It is the neural network that is used the most often to solve picture categorization issues. Convolutional layer, pooling layer, activation function layer, and fully connected layer make up the four layers of the CNN neural network. as shown in figure [5]

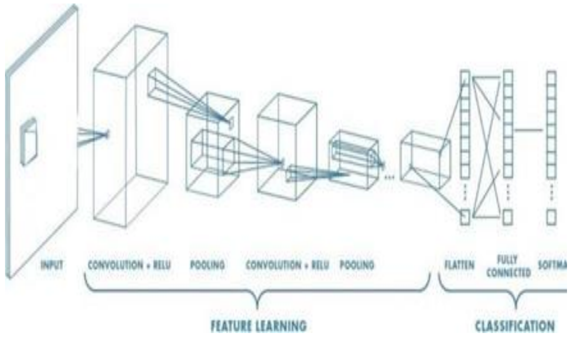

Fig 5: General CNN architecture

Convolutional Layer

It is the most crucial layer in the CNN model and is also in charge of giving the network its name. To get the image's characteristics, certain mathematical operations are carried out at this layer. It comprises of filters with width, height, and depth that are all smaller than the input picture. If a 64*64*3 picture is fed into CNN and there are a total of 10 filters, the layer's output will have a 64*64*10 dimension. There are a total of 5 convolutional layers with, respectively, 32, 64, 64, 128, 128 filters. All layers have a 3*3 kernel size.

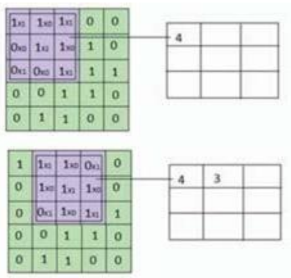

Fig 6: Convolution layer with 5*5 input images and 3*3 filters

Pooling Layer

The output of the preceding layer's size reduction is mostly due to this layer. Although filters of various sizes may be used in this layer, 2*2 size is often recommended. Max pooling and average pooling are the two main types of pooling layers that are used. As the name implies, maximum pooling uses the filter's maximum value, whereas average pooling uses the average. Stride is set to be equal to pool size by default in our model, which uses maximum pooling with a pool size of 3*3.

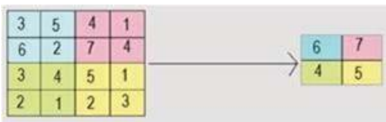

Fig 7: Max Pooling with 2*2 filters and stride 2

Activation Function

The activation layer is crucial in every neural network since it is in charge of the network's nonlinear learning. Different activation functions exist, including sigmoid, tanh, ReLU, and LeakyReLU.With the exception of the output layer, all of the layers in our model employ ReLU; the output layer uses softmax.

Fully Connected Layer

The output from the preceding layers' calculations is fed into a standard neural network for classification purposes. Model has two thick layers, each with 1024 and 19 units.

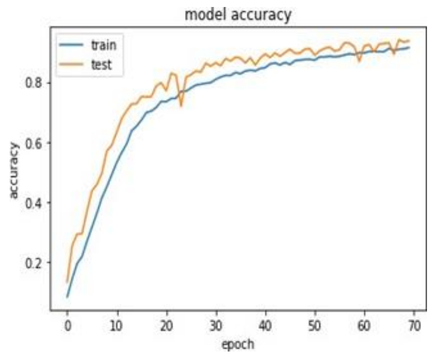

Fig 8: Accuracy vs iterations plot of CNN

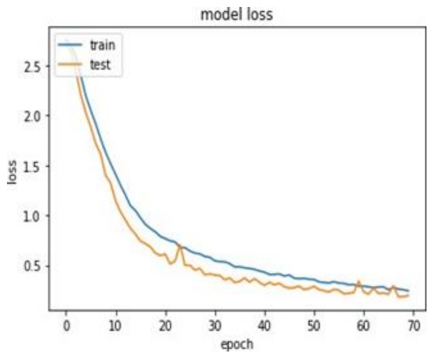

Fig 9: Loss vs iteration plot of CNN

The CNN model is giving accuracy of 99.74% on testing set that is the best among all classifiers with 0.0233 loss.

4. RESULTS

After the use of several picture pretreatment approaches, k-means clustering, and classifier comparison. Given the amount of classes, the logistic regression works pretty well and was able to provide accuracy of 66.4%. Due to its extensive calculations, KNN only provides predictions with a 54.5% accuracy rate. SVM also performs poorly and offers accuracy of just 53.4%. CNN exceeds the competition and provides excellent results with an accuracy score of 99.74%. In identifying and categorising the plant leaf disease, we can detect and categorise the illness with an accuracy of 99.74%. The accuracy score achieved by all classifiers are as follows:

TABLE 1 – Accuracy Achieved of Classifiers

CLASSIFIER	ACCURACY (%)
Logistic Regression	66.4
KNN	54.5
SVM	53.4
CNN	99.7

5. CONCLUSION

The convolutional neural network is used in this study to propose a highly accurate artificial intelligence method for identifying and categorizing various plant leaf diseases.

The dataset utilized in the given model has more than 20,000 photos divided into 19 different groups. The accuracy of the following model may be increased by fine-tuning the hyperparameters and utilizing an even larger dataset with additional illness categories. The model might also contain the treatments for the ailment that is categorized. The model may subsequently be made available on a Windows platform to farmers who can utilize the suggested approach in practice.

Thus, this model will help to solve real world problem to some extent and give suggestion to farmer about the particular disease their plant is suffering from as this will help grow the quantity and quality of the crop.

6. REFERENCES

[1] Trongtorkid C, Pramokchon P. Expert system for diagnosis mango diseases using leaf symptoms analysis. In2018 international conference on digital arts, media and technology (ICDAMT) 2018 Feb 25 (pp. 59-64). IEEE.

[2] Singh, Vijai, and A. K. Misra. "Detection of unhealthy region of plant leaves using image processing and genetic algorithm."In *2015 International Conference on Advances in Computer Engineering and Applications*, pp. 1028-1032. IEEE, 2015.

[3] Thangadurai, K., and K. Padmavathi. "Computer visionimage enhancement for plant leaves disease detection." In *2014 world congress on computing and communication technologies*, pp. 173-175. IEEE, 2014.

[4] Rajneet Kaur , Manjeet Kaur “A Brief Review on Plant Disease Detection using Image Processing” IJCSMC, Vol. 6, Issue 2, 2017

[5] SandeshRaut, AmitFulsunge “Plant Disease Detection in Image Processing Using MATLAB” IJIRSET Vol. 6, Issue 6 , 2017

[6] Sonal P Patel, Mr. Arun Kumar Dewangan “A Comparative Study on Various Plant Leaf Diseases Detection and Classification” (IJSRET),ISSN 2278-08882 Vol. 6 , Issue 3, March 2017

[7] “Plant Leaf Disease Detection and Classification Using Image Processing Techniques” by Prakash M. Mainkar, Shreekant Ghorpade,Mayur Adawadkar in IJIERE Vol. 2, Issue 4, ISSN- 2394-5494 , 2015

[8] C.V. Giriraja, C. M. Siddharth, Ch. Saketa, M. Sai Kiran, "Plant health analyser", Advances in Computing Communications and Informatics (ICACCI) 2017 International Conference on, pp. 1821-1825, 2017

[9] Aparajita, Rudransh Sharma, Anushikha Singh, Malay Kishore Dutta, Kamil Riha, Petr Kriz, "Image processing based automated identification of late blight disease from leaf images of potato crops", Telecommunications and Signal Processing (TSP) 2017 40thInternational Conference on, pp. 758-762,

[10] N. Belsha, N. Hariprasad, "An approach for identification of infections in vegetables using image processing techniques", Innovations in Information Embedded and Communication Systems (ICIIECS) 2017International Conference on, pp. 1-6, 2017.

[11] Trimi Neha Tete, Sushma Kamlu, "Detection of plant disease using threshold k-mean cluster and ann algorithm", Convergence in Technology (I2CT) 2017 2nd International Conference for, pp. 523- 526, 2017.