PLANT DISEASE DETECTION USING IMAGE PROCESSING AND MACHINE LEARNING ALGORITHM

Article · January 2022

DOI: 10.37896/jxu14.7/012

Banupriya .N
RMK Engineering College

PLANT DISEASE DETECTION USING IMAGE PROCESSING AND MACHINE LEARNING ALGORITHM

Rajaneni Deepika Sai Chowdary#1, *Yogeshwari V*#2, *Varsha*#3, *M.R.Banu Priya*#4 *Swathy Chinnappan*#5

Department of Computer Science and Engineering, R.M.K Engineering College

ABSTRACT — **The identification of the plant disease is crucial to obtain a good crop yield along with a good quantity of agricultural products. Detection of plant illness includes the research work of many farm-related factors such as organic farming, constant plant monitoring, and recognition of all diseases. In farms that contain entirely different crops, plant diseases cannot be tracked manually. This requires an enormous amount of work, plant disease expertise, and also a substantial amount of time. The image processing along with k-means clustering and convoluted neural networking algorithms could be used for the accurate prediction of the disease. The detection of the disease includes methods including image segregation, pre-processing data, fragmentation of the image, detection, and recognition of characteristics. This paper also examines the binding segmentation and retrieval functions of two different plant diseases.**

KEYWORDS— **image processing, plant disease detection, k-means clustering algorithm, Convoluted Neural Network.**

I. Introduction

Farming accounts for approximately 17%of total GDP[1], providing more than 60 % of the population with employment. The recognition of plant diseases plays an important role throughout the agricultural climate. Indian farming involves plants such as maize, wheat, and so on. With its root and leaf energies, each of these plants is cultivated.

For research in plants with visually recognizable trends, the plant disease experiments apply. The control of plant health and diseases plays an important part in the effective cultivation of plants. In the early times, the person with experience in this field was assigned responsibility to track and examine plant disease manually. This requires a lot of work and considerable time for processing. Image processing methods can be used to diagnose plant disease, and algorithms can be used to predict two different plant diseases. The plant disease experiments apply to research on the plants with visibly identifiable trends. In this article, we have performed a survey on various diseases of plants and specific specialized techniques to diagnose these conditions.

II. LITERATURE SURVEY

In the past few years, various developments within the declining agriculture field have arisen and that fetched an honest supply of financial gain for the farmers. And one in every one of them will be an image process with machine learning algorithms. Pomegranate (Punica granatum) may be a deciduous tree fully grown in arid and semi-arid regions [2].

It develops well in areas of 25-35 degrees temperature and 500-800 millimeter annual downfall. Diseases have resulted in Brobdingnagian in developed pomegranate in recent years. Micro-organisms like fungi, microbes, and viruses are sometimes liable for these diseases. Microorganism blight, seed stain, plant red, and leaf plot are the diseases [3].

Potato plants are straightforward to grow. they've fully grown virtually all told elements of the planet however many diseases have an effect on potato plants, however, the foremost common diseases are a blight, fungus wilt, and Rhizoctonia canker. These diseases are simply known and if treated early enough, the plants could also be saved. If the diseases don't seem to be caught early enough, the complete plant ought to be removed. These diseases are contagious and that they unfold from plant to plant simply. The diseases inflicting substantial yield loss in potato are Phytophthora infestans (late blight) and Alternaria solani (early blight). Early detection of those diseases will permit preventive measures and mitigate economic and production losses [4].

Over the last decades, the foremost practiced approach for detection and identification of disease is the optic observation by consultants. However, in several cases, this approach proves impracticable to the excessive time interval and inaccessibility of consultants at farms settled in remote areas [5]. Dhakate M & Ingole A. B. (2015). Used neural networks for the identification of the pomegranate plant diseases.

Observation of plant diseases exploitation machine learning is associated with the Nursing rising field and has gained a huge impact in precise farming within the last decade. It will be used in the event of measurements of upper quality combined with advanced algorithms and therefore the improved chance of mixing many image sources into datasets [6].

The photos are sometimes delivered via satellite imagination, sensors, or digital cameras put in in fields or mounted on drones. During this case, segmentation may be a crucial part that primarily includes 2 stages: the primary stage includes separating the leaf (foreground) from different images related to sections (background). The second part, additionally referred to as the realm of interest or ROI segmentation, includes distinctive symptom region(s) from the leaf [6].

Islam, M. AnhDinh, Wahid, K., & Bhowmik, P. (2017). Detection of potato diseases exploitation image segmentation and multiclass support vector machines. He performed segmentations on the leaf of a potato plant to extract the latent characteristics by masking out the background and therefore the inexperienced region of leaves and thereby he extracted the Region of Interest **(ROI).** And by coaching the multiclass Support Vector Machine **(SVM)** the author can find the diseases of the plants [4].

Anand Singh Jalal et al. [8] used complete native Binary Pattern (CLBP) for illness detection in apple fruit. His projected approach consists of the k-means cluster rule for feature extraction and pictures were classified as an exploitation of the Multiclass support vector machine [7].

Gittaly Dhingra describes the application of agriculture exploitation laptop vision technology to spot and classify the illness of the plant leaf. The paper deals with the correlation between illness symptoms and therefore the impact on product yield. It additionally deals with increasing the amount of coaching information and testing to get smart accuracy [8].

III.TYPES OF DISEASES IN CLASSIFICATION

A. EARLY BLIGHT POTATO LEAF

Figure 1: Potato early blight leaf

Early Blight in potato plants- The illness chiefly affects leaves and stems, however below weather conditions, and if left unrestrained, it will contribute to important defoliation and increase the prospect of infection with the tuber.[10]

B.LATE BLIGHT POTATO LEAF

Figure 2:late blight leaf

Late Blight in potato plants-Late blight is evoked by Phytophthora infestans, a fungus-like oomycete microorganism. This doubtless serious illness can infect foliage of potatoes and tubers at any purpose of growing crops [11].

C.ALTERNARIA of POMEGRANATE LEAF

Figure 3: Alternaria of pomegranate leaf

C. Alternaria: little achromatic circular spots seem on the leaves.

D.ANTHRACNOSE of POMEGRANATE LEAF

Figure 4: Anthracnose of pomegranate leaf.

D.Anthracnose: seems like a little regular or irregular boring violet or black leaf spot with yellow halos. Leaves flip yellow and fall out.

E.BACTERIAL BLIGHT

Figure 5:Bacterial Blight of pomegranate leaf.

E.Bacterial blight: look of 1 to many little water-soaked, dark-colored irregular spots on leaves.

IV.PROPOSED WORK

In our proposed system firstly we perform the image acquisition in which the images are captured using the digital camera.

And secondly, image processing techniques such as image improvement and image segmentation is performed on the leaf to enhance its affected region and to eliminate the noise from the provided image.

The k-means clustering algorithm is used to fetch the required features from the leaves. And the features of the infected part of the leaves are enhanced by obtaining the contrast image of the leave and from the neural networks.

And finally, the classification of the diseased leaves is completed using the convolution neural network. For all the images in the data set the following steps are performed repeatedly.

Image segmentation is performed on each image using k-means clustering a sample clustered image set which consists of three other clusters.

After the detection of the infected leaf, the textural statistics such as mean, variance, entropy smoothness, skewness the leaf are obtained from the Gray-Level Co-occurrence Matrix (Fig three provides the diagram for testing within the planned system, step-by-step that starts with the test image, the process of the given image to spot it's sort of plant. And from that, the clustering-based segmentation takes place that then is carried to get the feature extraction using an image process, and therefore the ROI of the leaf will be obtained.

classification takes place once the feature extracting from the testing the image process play its role of image acquisition, image improvement and segmentation take place before the neural network classification that is trained with the k-means cluster rule by looking for the center of mass of the pixels and it divides the pixels into a cluster that then represent the clustered image so provides the divided output.

V.FEATURE EXTRACTION

Feature extraction may be a vital and essential step to extract regions of interest. In our planned methodology, the fundamental options are mean, variance, entropy, IDM, RMS, variance, smoothness, skewness, kurtosis, contrast, correlation, energy, and homogeneity are calculated and thought of as options values. Then we've created the feature vector for these values. The divided methodology shows totally different values for pictures using Gray Level Co-occurrence Matrix (GLCM). The Mathematical formulas for feature extraction are used to calculate the accuracy of the images. [4]

Feature	Mathematical Formulation
Contrast	$\sum_i \sum_j (i-j)^2 g_{ij}$
Correlation	$\frac{\sum_i \sum_j (ij) g_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y}$
Energy	$\sum_i \sum_j g_{ij}^{\ 2}$
Homogeneity	$\sum_i \sum_j \frac{1}{1+(i-j)^2} g_{ij}$
Mean, M	$\sum_{i=0}^{L-1} g(i) P(g(i))$
Standard Deviation, S	$\sqrt{\sum_{i=0}^{L-1} (g(i)-M)^2 P(g(i))}$
Skewness	$\frac{1}{S^3} \sum_{i=0}^{L-1} (g(i)-M)^3 P(g(i))$
Entropy	$\sum_{i=0}^{L-1} P(g(i)) \log_2 P(g(i))$
Kurtosis	$\frac{1}{S^k} \sum_{i=0}^{L-1} (g(i)-M)^k P(g(i))$
RMS	$\sqrt{\frac{1}{L*L} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (g(i,j)-I)^2}$

Fig 6: Mathematical Formulation of features [4]

VI. CLUSTERING USING K-MEANS CLUSTERING ALGORITHM

K-means clustering algorithm is used to cluster the images according to their classes. It is an unsupervised algorithm, it has k centers one for each of the clusters present.

It always aims at minimizing the squared error function.

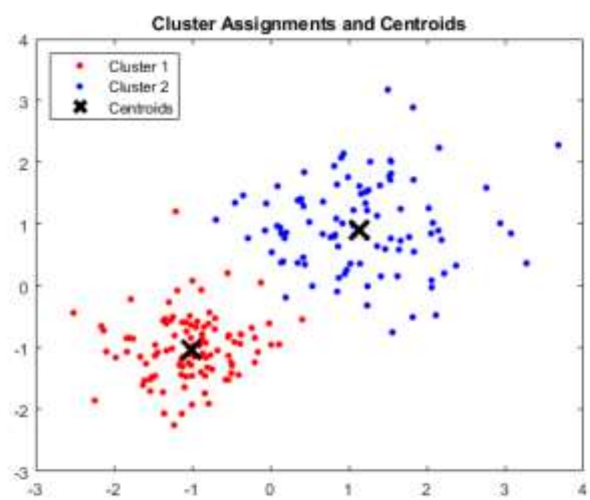

Figure 7: K-mean clustering

VII. CLASSIFICATION USING NEURAL NETWORK

A Convolutional Neural Network (CNN) is a scientific paradigm that's galvanized by the means biological nervous systems, like the brain, process data

images into a simple-to-process form, without leaving out the features that are crucial for obtaining better results.

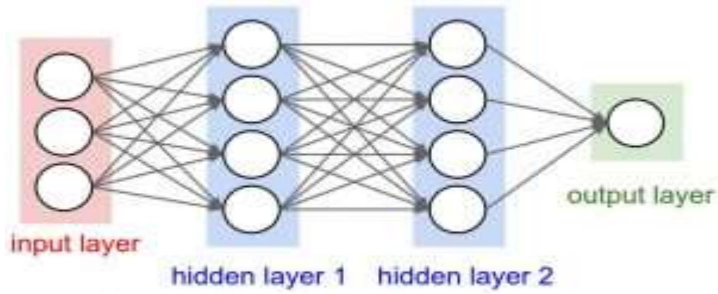

Figure 8: Convoluted Neural Network

- Input Layers: Input is given to the model in this layer. The number of neurons in this layer is equal to the total number of features in our data (number of pixels in case of an image).
- Hidden Layer: The input from the is then uploaded into the hidden layer. hidden layers vary according to our model and data size. Each hidden layer can have different numbers of neurons which are generally greater than the number of features. The output from layers is calculated by matrix multiplication of output of the previous layer with learnable weights of that layer and then by the addition of learnable biases by activation function makes the network nonlinear.
- Output Layer: The output from the hidden layer is then fed into a logistic function like sigmoid or softmax which converts the output of each class into the probability score of each class.

VIII.ROLE OF CLUSTERING & CLASSIFICATION

The picture processing system and k-means clustering algorithm were used to diagnose five pathogens, including early blight, late blight in potato leaves and Alternaria, anthracnose and bacterial blight affected pomegranate leaf. The K-means was used to cluster photos of the diseased leaf. The clustered photos were then transferred into a classifier NN. The outcome was that the NN classifier was much more accurate. This approach leads greatly to the precise and automated diagnosis of leaf diseases.

Using MATLAB software the training and testing of the leaves are done using the train files and test files that are represented using the mat files. The software would get all the files according to the given input it would process the mat files and provide the output file. The system architecture of our proposed work is the neural network consisting of 10 hidden layers and there are 3 clusters of the leaves. The neural network used is a forward backpropagation algorithm. The performance is calculated using the Mean Square Error **(MSE)** and the Epoch given was 2000 which is to obtain better accuracy.

Fig 9: original image of a pomegranate leaf[12]

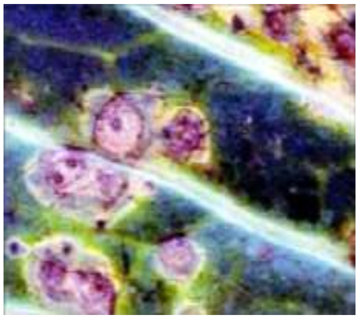

Fig 10: contrast image of a pomegranate leaf[12]

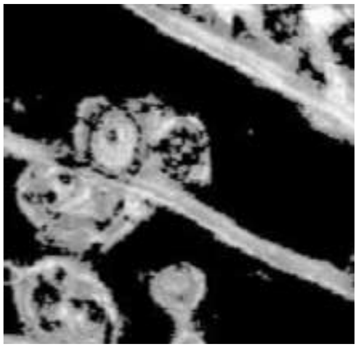

Figure 11: HSI image of pomegranate leaf[12]

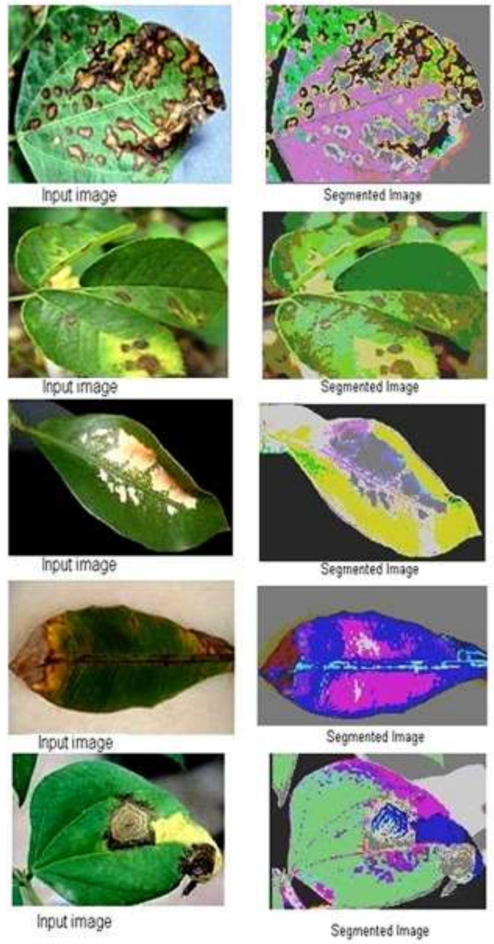

Figure 12: input and segmented image of potato plants[14]

XI. ALGORITHM

Segmentation by k-means clustering operation Input: Pomegranate leaf image.

Output: segmented clusters of pomegranate leaf image.

Step1. Scan input image

Step2. Input pictures converted regenerate to grayscale images.

Step3. Apply enhancement.

Step4. Resize the image.

Step5. Apply a K-Means clustering operation.

Step6. Find the centroid of the pixels.

Step7. Divide the pixels into clusters.

Step8. Represent the clustered image.

Step9. Segmented output.

Stop

step6: Once all the measurements have been allocated to the clusters, measure the score by summarizing all the euclidean distances between every data point and the respective centroid.

$$\text{Total distances} = \sum_{j=1}^{k} \sum_{i=1}^{n} \| x_i^{(j)} - c_j \|^2$$

step7: Define each cluster's new centroid by measuring the mean of all points assigned to the cluster. The definition (n is the number of points allocated to the cluster) is as follows:

x_i / n

X.RESULTS AND DISCUSSION

This research was experimented by using a software called MatLab when provided with the diseased leaf images they are segmented using the image processing techniques and segregated into different clusters belonging to there disease. The functionality of co-occurrence is calculated after mapping the R, G, B components of the given leaf image to the given threshold. The image classification is achieved first for the minimum distance criterion with K-Mean Clustering with k=3 And then the characteristic features are extracted from the clusters.

samples	early blight	late blight	bacterial blight	Alternaria	anthracnose
5 test images(bacterial blight leaf)	-	-	yes	-	-
5 test images (early blight)	yes	-	-	-	-
5 test images (late blight)	-	yes	-	-	-
5 test images (Alternaria)	-	-	-	yes	-
5 test images(anthracnose)	-	-	-	-	yes

Table 1:Classification of Different pomegranate and potato leaves

The proposed system for prediction of diseases of both plants mostly depends on the image processing, K-means, and neural networks based on the input given with 25 leaves and 5 diseases. The output is provided in a Mean Square Error **(MSE)** graph obtained using the hybrid algorithms.

The neural network is having the 10 hidden layers which are processed with the epoch of 2000 to obtain the higher accuracy.

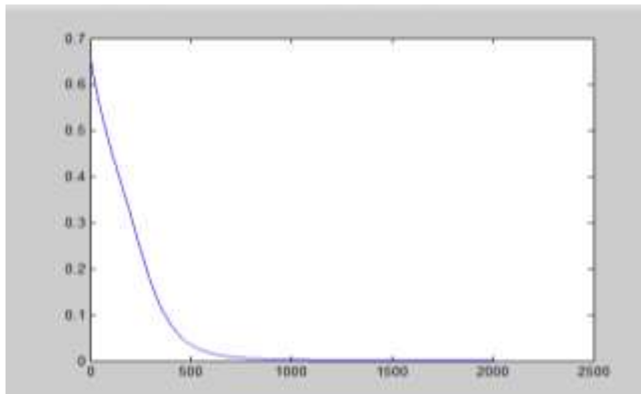

Fig.9 pomegranate disease prediction graph in Mean Square Error

The overall accuracy results obtained for both the leaf disease analysis using image processing, CNN and K-means clustering algorithm is approximately 89.8% for disease detection on pomegranate leaves and also 91% disease detection of potato leaf disease which is approximately 90% disease detection for both the plants.

XI.CONCLUSION

A new extended automatic area of hybrid algorithms has been enforced. The results indicated the excellence of the projected k-mean cluster for automatic segmentation of unhealthy symptom regions exhibiting the anatomy element, therefore enhancing its favored use in catching the mandatory border regions, tailored for finer extraction of different options with neural networks. The present rule, with the utilization of GLCM that implements specific color homogeneity by threshold and morphology, is easy to implement as a part of an entire disease detection system. The k-means clustering algorithm is used in the classification of the diseases of both potato and pomegranate plants. A neural network algorithm is used to predict the diseases of both plants. The epoch of 2000 is given to obtain much more accuracy in disease detection in plants.

REFERENCES

[1] "Indian agriculture economy.". Available: https:// statistics times.com/economy/sectorwise-gdp-Contribution-of India. PHP

[2] V.T.Jadhav," Vision -2025", National Research Centre on Pomegranate (Indian Council of Agricultural Research), August 2007.

[3] Khirade, S. D., & Patil, A. B. (2015). Plant Disease Detection Using Image Processing. 2015 International Conference on Computing Communication Control and Automation.

[4] Islam, M., Anh Dinh, Wahid, K., & Bhowmik, P. (2017). Detection of potato diseases using image segmentation and multiclass Support Vector Machines. 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE).

[5] Vibhute A, Bodhe SK., "Applications of image processing in agriculture: a survey," International Journal of Computer Applications. 52(2) 2012, pp. 34-40.

[6] S. Kaur, S. Pandey, and S. Goel, & quot; Plants Disease Identification and Classification Through Leaf Images: A Survey," Archives of Computational Methods in Engineering, journal article January 19, 2018

[7] Shivram Dubey and Anand Singh Jalal, "Detection and classification of apple fruit diseases using complete local binary patterns", Third International Conference on Computer and Communication Technology, vol.13(no.8): pages.346-351, August 2012.

[8] Mohammed Brahimi, Kamel Boukhalfa & Abdelouahab Moussaoui," Deep Learning for Tomato Diseases: Classification and Symptoms Visualization," vol. 31, no.4, 299–315, Taylor & Francis, 2017

[9] Vikaspedia - This portal has been developed as part of the national level initiative - India Development Gateway (InDG), dedicated to providing information/knowledge and ICT based knowledge products and services in the domain of social development. InDG is a Ministry of Electronics and Information Technology (MeitY), Government of India initiative, and is executed by the Centre for Development of Advanced Computing, Hyderabad.

[10]vegetablemdonline.ppath.cornell.edu/factsheets/Potato_EarlyBlt.htm

[11] Andy Robinson, Assistant Professor and Potato Extension Agronomist, NDSU/University of Minnesota Gary Secor, Professor, Plant Pathology Department, NDSU; Neil Gudmestad, University Distinguished Professor, Plant Pathology Department, NDSU

[12] plantvillage.psu.edu/topics/pomegranate

[13] Disease Detection in Pomegranate Leaf Using Image Processing Technique, GM sowmya, chandan,sampath kini

[14]Detection of plant leaf diseases using image segmentation and soft computing techniques Vijay SinghaL.K.Misrab

AUTHORS PROFILE

Rajaneni Deepika Sai Chowdary is a final year student pursuing B.E at RMK Engineering College under the stream of computer science department. Her area of interest includes Data Science and Artificial intelligence .

Yogeshwari V is a final year student pursuing B.E at RMK Engineering College in the department of Computer Science and Engineering. Her area of interest is DBMS and image processing.

Varsha G is currently pursuing B.E in the department of Computer Science and Engineering.Her area of interest is Machine Learning and Data Mining.

Banupriya N is currently working as Assistant Professor in CSE Department of R.M.K Engineering College. She has completed M.E. in 2011. She has experience of 8.5years of teaching. She has guided mini projects and Published 1 paper in international conferences and 2 in journals. Her area of interest is Machine learning.

Swathy Chinnappan completed her B.E. in 2006. She has 10 years of experience. Currently working in TCS Limited her current role is Senior Data Architect. Her area of interest is in Data science