International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-7, May 2020

Plant Leaf Disease Detection using Image Processing

Bukka Aswitha, Ravikumar CV, Venugopal P

Abstract*: Agriculture, the primary way to produce the food to the people besides its value added to the economy of the country's gross domestic product. Plants are affected to various diseases and early detection of disease has to be done in order to reduce the social and economical loses. Nowadays farmers are not aware of the type of diseases that affect the plants and the respective measures that have to be taken in order to reduce the effect. This paper focuses on the technique that detects the disease using image processing techniques and providing the measures to the farmers to overcome the disease. The technique is based on the K means clustering which is used to segment the image after that the feature extraction is done based on the Gray level Coocurrence matrix approach then the Support Vector Machine classifier is used to classify the disease with the trained data. We have calculated the percentage of leaf affected and the measurement is done based on it. Here along with disease name its symptoms and measurement are shown.

Keywords: Binary image, Gray level co occurrence matrix, K means clustering, support vector machine.

I. INTRODUCTION

India is a vast country where the main source of economy is Agriculture. Nowadays crops are affected with various bacterial, fungal and many more diseases .Proper care should be taken at the early stages in order to reduce the effect of these diseases on the plants. Many farmers are unaware of the type of diseases and the correct usage of pesticides to that particular disease. The major challenge is to find the disease in time and the proper pesticides and measures to be taken. In this paper we came up with the image processing technique that helps in detection of the disease and providing it with the symptoms and proper medication steps that have to be followed to reduce the effect.

The steps followed in this image processing technique are [9]:

i) Capturing the infected leaf and stem.
ii) Measuring the infected area.
iii) Segmenting the image using K-mean clustering.
iv) Feature extraction using the GLCM approach.
v) Comparison of the feature with the trained data.
vi) Predicting the disease, symptoms and measurements.
vii) Percentage of affected pixel calculation

Images of the leaf or the stem that are affected are taken and the preprocessing is done to remove the noise. Once the preprocessing is done then segmentation is done with the K means clustering method [10] and after this the features are extracted using GLCM [11] and the disease is classified using SVM classifier [12]. Then the disease name along with the symptoms and the measuremts that have to be taken are shown in the message box. This helps the farmers to take necessary steps with in no time to control the effect of the diseases. By using this GLCM technique almost of 22 features are extracted and the classification is done based on this which gives us with the accurate results compared to other techniques. The SVM technique used are much faster than multilayer preceptron networks and accurately predict the target probability scores. From the segmented image the affected region is clustered and converted into binary image and the percentage of the affected pixel is calculated.

II. LITERATURE SURVEY

a. Plant leaf disease using fusion of super pixel, K means, PHOG [1]

The proposed system applies Internet of Things to segment the leaf and recognize the type of disease using combination of K means clustering pyramid of histograms of orientation gradients algorithm and Super pixel clustering algorithm [1].This experiment is conducted on various plant and the technique was effective in providing the result.

b. Detection of plant disease using multiclass vector machine. [2]

The author focuses on integrated approach of combining both image processing and machine learning for disease detection from leaf image. Steps followed in this technique include primarily masking out the green region and background of the leaves, thereby retaining the region of interest (ROI) [2] which is the affected part of the leaf. Then by training support vector machine classifier with particular texture, color property .The segmentation approach and vector machine over 300 images provides an accuracy of 95%.

c. A LDA based segmentation model for plant disease [3]

Here the author approached through a fusion pf Color transformation and machine learning algorithm linear discriminate Analysis (LDA) [3]. At the initial stage different color features of pixel are extracted followed by multivariate representation,

Revised Manuscript Received on May 30, 2020.
***** Correspondence Author
Bukka Aswitha, Department of Electronics and Communication Engineering ,Vellore Institute of Technology, Vellore, India. Email:aswithareddybukka@gmail.com
Ravi kumar C.V , Assistant professor Sr. Department of Electronics and Communication Engineering ,Vellore Institute OfTechnology, Vellore,India. Email:ravikumar.cv@vit.ac.in
Venugopal .P, Assistant professor Sr Department of Electroics and Communication Engineering, Vellore Institute of Technology,Vellore,India. Email:venugopal.p@vit.ac.in

then LDA is used to obtain new feature space. During the testing phase, pixel feature of test sample is replaced to the new space and then 1NN(one nearest neighbor) classifier is used for classification. Finally from the original image the symptoms are separated.

d. Leaf disease detection using Sparse Representation

The proposed system uses three pipelined techniques for cucumber disease recognition. Using k-means clustering method segmenting .Secondly, from lesion information the shape and color of leaf is extracted, and finally classifying diseased leaf image using sparse representation (SR) [4]. The advantage of this approach is cost effective when classification is done in SR space.

e. Monitoring and controlling of leaf disease using image processing techniques

In this paper an automatic system is developed that identifies and classifies the diseased plants from the given data set. Four different types of diseases which include, rice blast, rice sheath rot, rice bacterial blight, rice brown spot are identified and classified. Two features like shape and color of the diseased portion is extracted using the developed algorithm. Minimum distance Classifier (MDC) [5] and k-Nearest Neighbor classifier (k-NN) [5] is used to classify the disease by combining the extracted features. The accuracy obtained using k-NN method is 87.02% and using MDC is 89.23%.

f. An enhancement in SVM to improve disease detection

In this paper an automated system to detect the diseased leaf and also percentage of affected area in leaf is also estimated. Support vector method [6] is used is used with two datasets one for training and others for testing. The steps used for the proposed algorithm includes, Capturing the image, secondly background and black pixels are segmented, later separation of saturation part and hue part of the image is done, finally diseased part of the leaf is segmented and healthy part is separated. SIFT is used for feature matching

g. Disease detection and severity estimation from unconstrained images.

The author focuses on an approach to detect the disease in the cotton plant and estimate the current stage of the disease. The main advantage of this method is that the image was captured in uncontrolled environment by untrained person that have complicated background making segmentation of diseased part more complicated and challenging. The technique uses two cascaded-classifiers which includes initially classifier segments the leaf from the background using local statistical features. Then the next classifier is trained to detect the type of disease and the stage at which the disease is at present using hue-andluminance0from HSV0color0space [7].

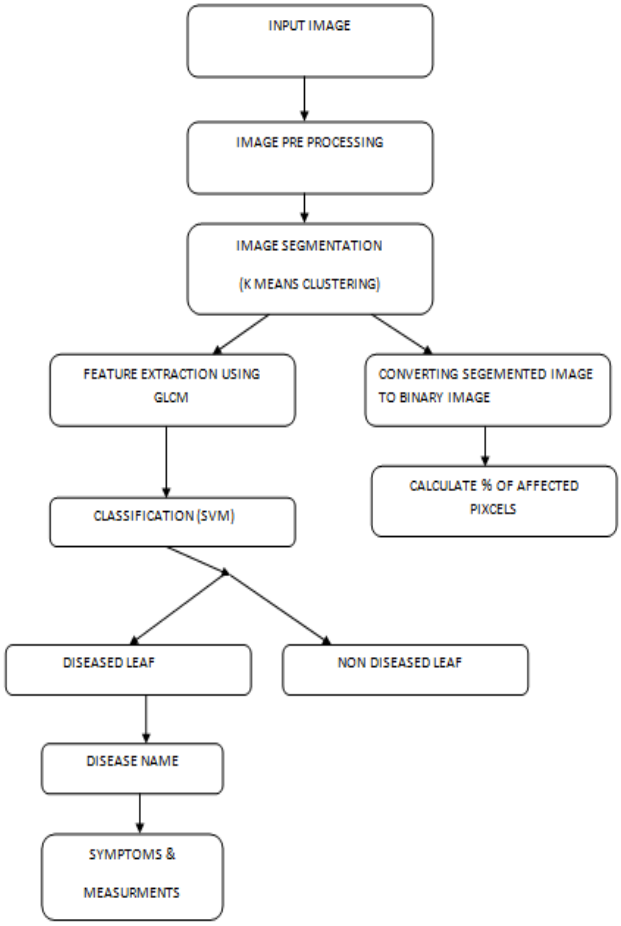
III. PROPOSEDSYSTEM

1.The proposed technique uses k-means clustering process for clustering of the images
2.Then the image properties like entropy, energy, correlation are calculated using GLCM.
3.Then the SVM classifier helps in classifying the diseases by using the trained data and then getting its symptoms and measurements along with the disease.
4.We have calculated the percentage of affected pixels using binary image and this helps in finding the percentage of affected leaf and preventive measures can be taken based on it.

This proposed method is quite efficient and accurate as we are using SVMs. The speeded up robust feature algorithm-based scheme, called GLCM is used which gives better accuracy in less span of time.

- Based on k-means clustering the segmentation and classification can be easily performed. Hence time consumption will be less.
- SVMs are much faster than multilayer perceptron networks and accurately predict target probability score.

IV. FLOW CHART



V. METHODOLOGY

The proposed concept is completely uses image processing techniques to get the output. The first step in the process is input image or image acquisition.

A. Image Acquisition

In this step the image is given as the input to the system. The input can be of any popular format. Normally any input image has lot of noise associated with them.

B. Image Preprocessing

The input image that we has lot of noises associated with it .

This noise is due to the change in the pixels values while taking an image or any other cause .So proper filtration techniques have to be used in order to get the accurate result without losing the information. This involves re sampling, noise reduction, contrast enhancement .Thus the final image obtained is enhanced and the information of each pixel is used to get the accurate results. There are several methods used for noise filtering but here we are using median filtering technique [14] which gives the accurate result without losing any information

C. Image Segmentation

Image segmentation involves in the segmenting the image based on the similar attributes. This is used to easily analyze the image for the disease detection. Here we have used K-means clustering method. This method clusters the images based on the mean values of the pixels. This clustering helps in the identification of the diseased area of the plant leaf.

I) K Means Clustering

This clustering method subgroups the items into k groups of similarities. To calculate this we use Euclidean distance as the basis of measurement.

K means Clustering Algorithm [10]:

1. Divide the points into two clusters based on their similarity.
2. Find the mean of each cluster by giving values to the data points.
3. Calculate the Euclidean distance [10] of each point in the clusters from the corresponding cluster mean co ordinates.
4. Change the cluster points based on its nearest distance to the mean.
5. Repeat the above steps till the distance between the mean and the cluster points doesn't vary much.

Euclidean Distance = $\sqrt{(x2-x1)^2+(y2-y1)^2}$

D. Feature Extraction Using Glcm:

Feature extraction is used to extract the features of the image and then classifying the image by comparing the values obtained to the already obtained of the trained data. There are several techniques used in the feature extraction like Gabor wavelet method, GLCM and many others. Here we have used GLCM method to extract the texture features. Total of 22 second order statistical values are calculated and stored in the form of a matrix. The features extracted here are used to detect the disease affected region and the type of disease based on the similarity with the trained data values for the diseased leaf .If it doesn't match the leaf will be treated as the healthy leaf.

i) Gray Level Co occurrence Matrix (GLCM):

GLCM is a way of extracting the image properties related to second-order statistics like entropy, energy, correlation, Homogeneity which considers the relationship among pixels or groups of pixels. Simple first order statics is not useful in characterizing texture features to detect the disease.

Hence, two-dimensional GLCM is extensively used to find the disease here in this method a total of 22 second-order characteristics are calculated and there values are compared with the trained data values to detect the disease. Here we are making use four texture features namely Correlation, Energy, Entropy and Homogeneity in this paper [11].

Algorithm:

a) Quantize the image data.
b) Create the GLCM.
c) Make the GLCM symmetric:
i) Make a transposed copy of the GLCM
ii) Add this copy to the GLCM itself.
d) Normalize the GLCM.

Calculate the selected features here.

E. Svm Classifier

Support Vector Machines are dominant and flexible supervised machine learning algorithms that are used for both classification and regression. They are very popular because of their ability to handle multiple continuous and categorical variables. This is mainly a representation of different classes in a hyper plane in multidimensional space. The hyper plane will be generated in an iterative manner by SVM so that the errors can be minimized. Then it will choose the hyper plane that separates the classes correctly. Advantages of using this classifier are they proffer great accuracy and work well with high dimensional space and they use a subset of training points so they use less memory.

F.Percentage Of Affected Pixels Calculation

After the segmentation is done using the K means clustering method the image is converted from RGB to Gray scale then the gray scale image is converted into binary image to calculate the number of pixels affected and the percentage is calculated using the formulas:

$$P_L = wp_a + wp_u \quad \text{----------} \quad \textbf{(1)}$$

$$P_A = \frac{wpa}{PL} * 100\% \quad \text{----------} \quad \textbf{(2)}$$

Where PA is the affected pixels, wpa total number of white pixels in the affected region, wpu total number of pixels in the unaffected region, PL is the total pixels of leaf area.

VI. SYSTEM ARCHIETURE:

Initial process is to give the input image to the system. By using the image pre processing techniques mentioned above the leaf has to be diagnosed whether it was affected or a healthy leaf .Once the pre processing is done the image has to be segmented and feature extraction is to be done to find the name of the disease along with the symptoms and precautions to be taken. This project also provides the overall percentage of the affected region of the leaf and the preventive measures will be taken based on the affected percentage.

VII. RESULTS

A. Image Acquisition:

The initial step is to give the input to the computer to find whether the leaf is diseased or not and to calculate the percentage of affected pixels. Here we have taken a paddy leaf as the input.

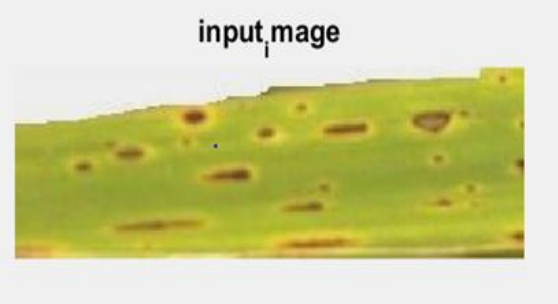


Fig 1: Input image

B. Image Preprocessing:

Once the image is given as the input the preprocessing is done which includes re sampling, noise removal using median filter, contrast enhancement is done and a smooth and enhanced image will be the output from this step. This image will be sent for the further segmentation. Here we have used median filtering technique for the noise removal.

Fig 2: Smooth and Enhanced Image

C. Image Segmentation:

Once the image is pre processing is done the image is sent for segmentation. Here we used K means clustering method for the segmentation process and we get three different outputs as three different clusters are made we make use of the second cluster output to find the disease .

Fig 3: Clustered image

E.Percentage Of Affected Pixel Calculation

i. After the segmentation is done the image is converted from RGB to gray scale image then the gray scale image is converted into binary image .In the below image the white colored portion indicates unaffected part of the leaf. It helps to calculate total number of pixels in the unaffected region.

Fig 4: Unaffected region of the leaf

ii. After this the segmented image with the affected region is converted into binary image where white colored region indicates the diseased area of the leaf. It helps to calculate the total number of affected pixels

Fig 5: Affected region of the leaf

iii.Pixel calculation:

From fig 4

WPu = 31,565

BP1= 15,527

Total number of pixels PL = WPu+BP1

= 47,092

Here BP1 indicates black pixels.

From fig 5

WPa = 8,614

BP2 = 38,478

Total number of pixels PL =WPa+BP2

= 47,092

Here BP2 indicates black pixels.

Now according to (1), total pixels of the leaf area that is obtained will be

PL= 8,614+31,565

= 40,179

According to (2), Percentage of affected pixels can be calculated as

P_A = (8,614/40,179)*100

=21.44%

So P_A =21.44%

So from the above obtained result 21.44% of the leaf is infected with the disease.

E. Feature Extraction:

After the clustering is done the feature extraction is done using GLCM matrix method and by using the SVM classifier we find the matching values with the trained data and the disease is identified along with the Symptoms and measurements of the disease. If the leaf is not diseased we get the output as HEALTHY LEAF. Here we get the percentage of the affected leaf through binary image calculation. Here we get two dialogue boxes stating one as diseased leaf and the symptoms of it.

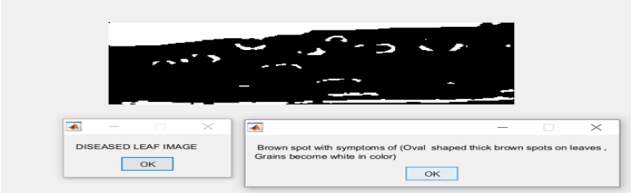


Fig 6: Final Output

This table helps to take the preventive measures based on the percentage of affected pixels

Table 1: Measures to be taken based on percentage of affected pixels

% of infection	Measures
Less than 10%	Flooding of the whole field with the water helps in reducing the affect of the disease from spreading and close observation has to be taken.
10% to 15%	Use Zineb and avoid water stress. Proper nutrition should be given. Excessive use may increase the disease spread in the field
15% to 25%	Silicon based fertilizers like calcium silicate slag should be used. Use Mancozeb and spray it with water.
More than 25%	Use resistant variety Amruth , do not use high nitrogen fertilizers .Use fungicides like(iprodine, propiconazole, azoxystrobin) Spray(1g of ediphenphos and 1 liter water)

Now a comparative study is conducted on the leaf images using the proposed method (PM) and the standard method(SM) which is implemented using the MATLAB function bwarea [9]. The steps followed are:

i. Segment the enhanced image into two clusters .One for affected leaf and the other for unaffected leaf area.

ii. Convert both the images into binary images.

iii. Calculate total area of the objects of the above binary images in pixels

In this method total number of white pixels is treated as the pixels of the objects as

Total =bwarea (BW)

Where BW is the binary image of the input image and Total is the total number of white pixels of the objects of the converted binary image.

Table 2: Experimental results for various figures

Sl no.	Sample Image	Percentage of infection by(SM)	Percentage of infection by(PM)	Deviation
1	Figure 1	17.7306	17.5482	0.1824
2	Figure 2	45.4523	45.4501	0.022
3	Figure 3	3.244	3.164	0.08
4	Figure 4	3.4647	2.4014	0.0633
5	Figure 5	3.2042	2.0311	1.1731

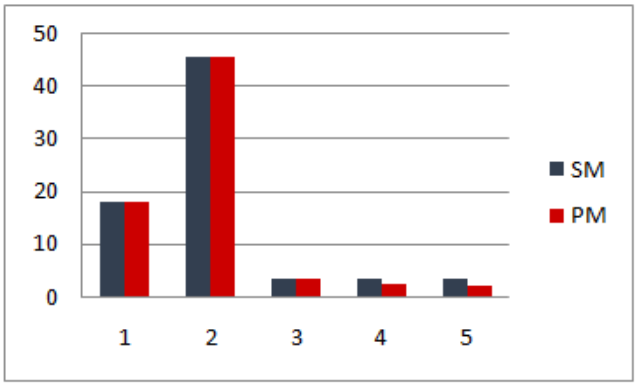


Fig 7: Graphical analysis of % affection of leaf using SM and PM

VIII. PROSPETIVE IMPROVEMENTS

Here we may also connect this to an app on mobile phone where it will be easy for the farmers to cluck pictures and this can be used as data set to find the disease without any delay and it can also provide the reviews of the pesticides they have used and process used by the other farmers to reduce that disease .They may even ask questions related to pesticides and farming.

IX. CONCLUSION

This paper discusses the technique which is very accurate, less time consuming and easily accessible by the farmers. The clustering technique for the image is done into three different clusters that give us with the accurate data to detect the diseases. The percentage calculation of the affected pixels helps to take preventive measures according to the percentage of infection. This technique is inexpensive, less time consuming and almost 95% accurate when compared to other existing techniques. This early detection of plant diseases helps farmers to increase crop yield and in turn helps in improving the India's GDP.

ACKNOWLEDGMENTS

I would like to thank our professor Ravikumar CV, Venugopal P who guided me through the project with his insights. We would also like to thank our institution for providing us with the equipment and opportunity to conduct this project.

REFERENCES

1. Shanwen Zhang, Haoxiang Wang, Wenzhun Huang, Zhuhong You, "Plant diseased leaf segmentation and recognition by fusion ofsuperpixel, K-means and PHOG", International Journal for Light and Electron Optik Volume 157, March 2018, Pp 866-872
2. Islam, M., Anh Dinh, Wahid, K., & Bhowmik, P. (2017). Detection of potato diseases using image segmentation and multiclass support vector machine. 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE).
3. Na Wu, Miao Li, Lei Chen, Yuan Yuan, Shide Song, "A LDA-based segmentation model for classifying pixels in crop diseased images", Proceedings of the 36th Chinese Control Conference , Dalian, China July 26-28, 2017, Pp 11499-11505

4. Shanwen Zhang, Xiaowei Wu, Zhuhong You , Liqing Zhang, "Leaf image based cucumber disease recognition using sparse representation classification", Journal for Computers and Electronics in Agriculture archive, Volume 134 Issue C, March 2017, Pages 135-141
5. Joshi, A. A., & Jadhav, B. D. (2016). Monitoring and controlling rice diseases using Image processing techniques. 2016 International Conference on Computing, Analytics and Security Trends (CAST).
6. Kaur, R., & Kang, S. S. (2015). An enhancement in classifier support vector machine to improve plant disease detection. 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE).
7. Parikh, A., Raval, M. S., Parmar, C., & Chaudhary, S. (2016). Disease Detection and Severity Estimation in Cotton Plant from Unconstrained Images. 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA).
8. Hrushikesh Dattatray Marathe, Prerna Namdeorao Kothe, Leaf Disease Detection Using Image Processing Techniques, International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 3, March – 2013.
9. BWAREA Area of objects in a binary image available https://in.mathworks.com/help/images/ref/bwarea.html
10. K meanns Clustering available at https://in.mathworks.com/help/stats/k-means-clustering.html
11. GLCM algorithm available at https://in.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html SVM available at
12. https://in.mathworks.com/help/stats/support-vector-machine-classification.html
13. Median filtering is available at https://in.mathworks.com/help/images/ref/medfilt2.html

AUTHORS PROFILE

Bukka Aswitha an under graduate student from Electronics and communication engineering, Vellore Institute of Technology, Vellore, India.

Ravi kumar CV Ph.D in communication Networks, M.Tech in Communication Systems , Research area Communication Networks and Machine Learning , Assistant Professor Sr. School of Electronics Engineering from Vellore Institute Of Technology, Vellore, India.

Venugopal P, Ph.D in Power systems, M.Tech in Control Systems, Research area Power Systems and Machine Learning, Assistant Professor Sr.School of Electronics Engineering from Vellore Institute Of Technology, Vellore, India.