AI Based Fault Detection on Leaf and Disease Prediction Using K-means Clustering

Sujan Sarkar[1], Sanket Biswas[2], Avinaba Tapadar[3], Pritam Saha[4]

[1,3,4]UG Student, Dept. of Electronics and Communication Engineering
[2]UG student, Dept. of Computer Science Engineering
Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India, 735102

Abstract - *Crop diseases creating a massive problem in food cultivation or in food industry. Day by Day population is increasing; parallel it is very important to increase the food production. But crop diseases creating some big problem. In this paper we have mainly focused on how important to find a fault on a leaf to get rid of it. Our project includes this problem and a suggested process to find the fault area on a defect leaf. Next we have to find the ratio of fault and normal portion of that leaf. The whole process is done on MATLAB 2015a environment. We have created a MATLAB code using K-means clustering method to find that fault area. After that we calculate the ratio of fault area and normal area of that leaf so that we can realise if it is possible to cure that leaf or not. By image processing and image analysis we are trying to detect the disease type of that particular portion using data management and artificial intelligence*.

Key Words: image processing, data management, artificial intelligence, k-means clustering, MATLAB, image segmentation.

1. INTRODUCTION

It is expected that by 2100 earth's population will reach 11.2 billion and with this burning issue there is a crucial need to expand food production. But the main problem is less number of cultivated lands. But to give the food to all over world we have to produce as much as we can. Problem arises when the production is influenced by diseases problem. At this moment it will be a crucial challenge to minimize the losses of crops by the pests and diseases to secure food supply by the aid of technological support. It is noticed that diseases cause heavy crop losses amounting to several billion dollars annually. Most of the cases pests and diseases are detected on the leaves or branches of the plant.

Table – 1: From Food and Agriculture Organisation of the United Nations the total crop loss is given below.

Country (Year)	Estimated Loss (mt)	Loss as Percentage of Expected Output (%)	Value of Production Loss (US$ Million, 1994)
Thailand (1994)	130	58	650
Philippines (1989)	57	96	285
Ecuador (1992)	34	27	170
Indonesia (1991)	50	34	250
China (1992)	180	84	900
Taiwan (1987)	100	72	500
Mexico (1994)	1	8	5
USA (1993)	12	NA[1]	60
India (1994)	25	36	125
Vietnam (1994)	10	20	50
Bangladesh (1994)	5	14	25
Total	**541**	**74**	**3,019**

[1]NA = not available.

Decisive quantification of these optically observed diseases has not been sufficiently studied because of compilation of visual patterns and for that reason the demand for more precise and sophisticated image pattern discerning is increasing. Using image processing techniques we can define images over two dimension (feasibly more) by which we can have more precise image pattern which plays a key role in successful cultivation of crops.

There are several popular techniques which are used in digital image processing like-Anisotropic diffusion, Hidden Markov models, Image editing, Image restoration, Independent component analysis, Linear filtering, Neural networks, Partial differential equations, Pixilation, Principal components analysis, Self- organizing maps, Wavelets.

Image processing for classification of crop diseases was researched by Ying [15]. He mentioned that leaves with marks must be carefully examined in order to carry out intelligent diagnosis on the basis of image processing.

Important image processing methods:

- Image clipping: Classification of leaves with marks.
- Noise reduction: Median filter wipes out noises from the images.
- Thresholding: To segment image into the spot background.

To different methods for the diagnosing of plant diseases were put forward by experts:

- Step by step descriptive methods
- Graphical representation

For the grading process of flue-cured tobacco leaves image feature extraction is useful. Automatic investigation of flue-cured tobacco leaves was mentioned in the report by machine vision techniques [16]. To resolve problems of feature extraction the above mentioned techniques are used.

2. LITERATURE SURVEY

Xanthomonas oryzae, a bacterial disease causes most destructive bacterial blight of rice which causes almost 50% of worldwide annual yield loss [21, 22]. In 1979 and 1980 Punjab and Haryana states of India heavily infected by the disease. The report of Cai and Zhong says that the bacterial blight is also one of the common disease on hybrid rice in Zhejiang Province in China [23]. Magnaporthe grisea causing disease has been found in over 85 countries across the world, which is a fungal disease. It is estimated that every year it destroy enough food to feed more than 60 million people.

Pest damage in pip fruits in fields can be detected by image processing techniques and neural networks that uses wavelets as a means by line detection which was suggested by Woodford, Kasabov and Wearing in his work named "Fruit Image Analysis using Wavelets" [7]. The leaf-roller, codling moth, and apple leaf curling midge were specifically chosen as the main research subjects as they were the predominant pests in the orchards. To detect the main cause Daubechies wavelets are used which are derived from fast wavelets. It's a two-step process- the foremost step include the comparison of the elementary image with the three color component deviation the second stage comprises of the calculation of weighted version of the Euclidean distance between feature coefficient as chosen in the previous stage and the one with the querying image is calculated and one with the least distance are chosen and arranged as identical image to the conditions provided.

Cucumber crop was handled from pest infection by CLASE(central lab of agricultural expert system).four image processing method are used enhancement segmentation , feature extraction and classification to investigate disorder from leaf image .They tested three different disorders such as leaf miner ,powdery and downy. Errors have been highly been reduced between system and use by the following methods. Prasad Babu and Srinivasa Rao worked in recognition of leaf [10] disease by back propagation neural network. To determine the species from the leaf, it was proved that only black propagation network is applicable. Prewitt edge detection and thinning algorithm is used to find back propagation algorithms and leaf tokens as input. Experimentation with large training sets to recognise various leaves with pest and rotten leaves due to insects or diseases can be done as an advancement method as predicted in several reports.

Remote sensing data for segmentation of agricultural landed fields by neural network approach was proposed [12]. Easy to extract features and high efficient recognition algorithm is used to implement a leaf recognition algorithm [12]. For plant leaf recognition Probabilistic Neural Network approach are used. The exactness of this algorithm is 90% on 32 kinds of plants.

For disease recognition based on the diseased images of various rice plants, a software prototype system was explained by Santanu and Jaya in their report[13]. Image growing, image segmentation techniques were used to detect infected parts of plants. For classifying diseased rice images Self Organise Map (SOM) neural network is used.

Otsu's segmentation is used for segmented leaf region. By calculation of the quotient of disease spot and leaf area plant diseases are classified.

Using hybrid intelligent system grape leaf disease is detected from colour imagery [14]. Self-organizing maps and back propagation neural networks are used. Segmentation of grape leaf disease is carried out using advanced self-organizing feature maps with genetic algorithms for optimization and support vector machines for classification and then filtered using Gabor wavelet allowing analyzing and classifying leaf diseases.

A method to investigate plant diseases caused by spores is mentioned in the report [17]. To carry out processing and analysis such as histogram generation, the grey-level correction, image feature extraction and image sharpening, the coloured image is converted to grey image. Using Median filter and canny edge algorithm grey image is processed by edge enhancement. After thresholding binary image captured is operated using morphological features like dilation, erosion, opening etc.

The report proposes and application of image processing and Support Vector Machine (SVM) for the early and accurate detection of rice diseases. Rice diseases marks are classified and their shape and texture feature are extracted. The selected shape and colour texture of disease spot are set as characteristic values of classification. The SVM method was utilized to classify rice bacterial leaf blight, rice sheath blight and rice blast. The affectivity of SVM was 97.2%.

Otsu segmentation, K-means clustering and back propagation feed forward neural network for clustering and classification of diseases that affect plant leaves are used

3. BASIC CONCEPTS OF IMAGE PROCESSING

Image processing techniques are highly beneficial for extracting the desired results from the given input samples. In this work we have used some of the key image processing techniques to find out a solution to the problem.

3.1 Histogram Equalization

To enhance the contrast in the sample data, we have used the histogram equalization technique to stretch out the available range of intensities. The given sample distribution is mapped into a wider and more uniform distribution implemented with a remapping function, the cumulative distribution function (cdf). For a histogram H(i) , it's cdf

$$H'(i) = \sum_{0 \le j < i} H(j)$$

The intensity values of the equalized image output can be obtained from the given equation:

$$equalized(x, y) = H'(src(x, y))$$

3.2 K-means Clustering

To find groups in unlabeled data with the number of groups denoted by variable K, each data point is iteratively assigned to one of the K groups based on the features that are extracted. The clustering of data points are done based on similarity of features. Each centroid of a cluster defines a collection of feature values.

Let Ci be a collection of centroids in set C, each data point x is assigned to a cluster based on:

$$\underset{c_i \in C}{\arg\min}\ dist(c_i, x)^2$$

4. METHODOLOGY AND RESULTS

The fault detection method is same as previous processes. The whole process is done in MATLAB 2015-a simulation software environment. The total work has been classified in some steps, combining the steps we have created an algorithm on basis of specific problem to extract the desire result for further analysis. We use a digital camera to click the picture and then the picture has gone through the processing technique.

Our proposed algorithm shown below:

- Background clipping
- Binary Image conversation
- RGB Image acquisition
- K-means clustering
- Masking green pixel
- Remove masked cell and highlight fault portion
- Masking fault pixel
- Removing masked cell and highlight normal portion
- Histogram plot of original image, fault portion and normal portion
- Comparing both graph

4.1 Preprocessing

The entire framework for our model is based on the techniques of image processing and classification. We have acquired the digital image samples of different varieties of leaves from the environment using a digital camera. Every sample image was then analyzed and processed using our proposed approach. Useful features were then extracted to further classify those samples for our required objective.

(a) (b)

Fig - 1: (a) normal rice leaf, (b) defective rice leaf

For 1st step the background of the sample leaf is clipped. Main reason why we use this process is to avoid the unnecessary noise in the result. As the picture has collected from the natural environment, so if the background remains unchanged then the background colour can change the main result.

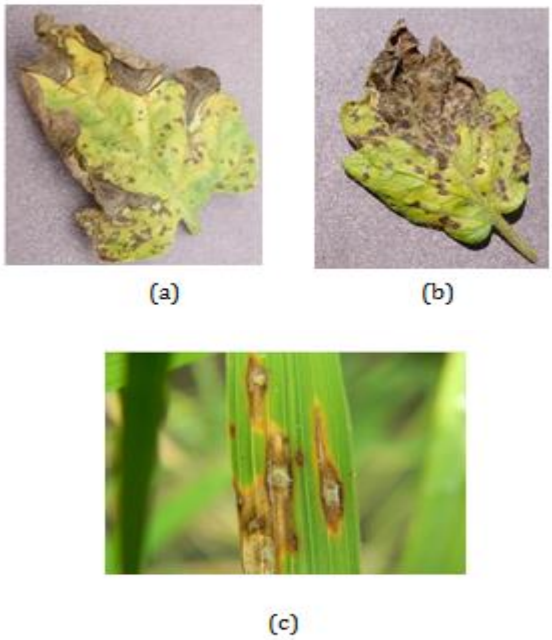


Fig – 2: (a) (b) (c) Sample data of a defect leaf

4.2 Segmentation

Initially, the RGB images of all the leaf samples were studied as shown in figure 3. The samples can be classified into faultless and faulty. In the 2nd step we have done colour transformation by creating a Matlab code to convert digital image to binary image and then the RGB image. Image transformation is most important step as it is the first input of K-means clustering.

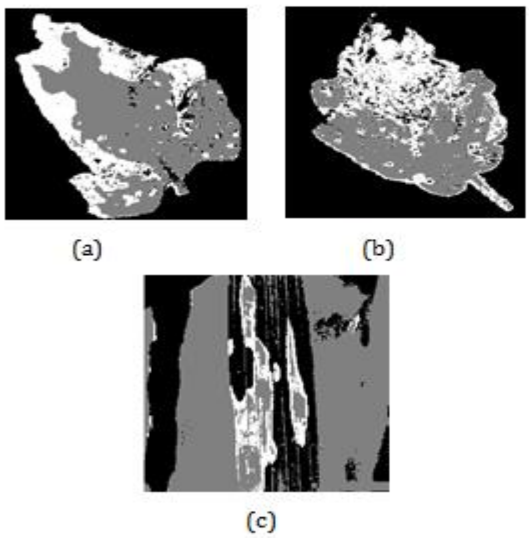


Fig - 3: Colour conversion

4.3 Clustering

The K-Means Clustering algorithm has been used to classify the pixels of sample leaf into K number of clusters based on similar kind of feature weights. This helps to identify the cluster containing the infected portion of the sample leaf afflicted with some specific disease.

Fig.4 illustrates the infected portion of the leaf after clustering has been implemented.

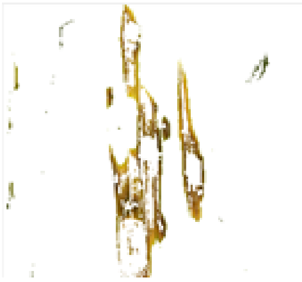

(a) (b)

(c)

Fig - 4: Clustered with threshold identification

Step 5 and 6 is very important towards our destination. This process has done in 2 steps. In first step we have pointed out the green or approximately green pixels from the image. We have applied Otsu's method in order to specify the varying threshold value which chooses the threshold to minimize the intraclass variance of the thresholded black and white pixels to find that point. In the next step we have extract the remaining part after removing the green pixels. W have applied this processes two times to extracts the perfect figure. The result is quite impressive as the clearance is high than the previous methods. The figures are shown below.


(a) (b)



(c)

Fig - 5: K-means clustered and extraction of fault area.

After gone through the step 5 and 6 where we extract the fault area from the picture, now in the steps 7th and 8th we have mainly focused to find the normal portion from the image, so that we can measure the ratio of normal and fault portion. This ratio is very helpful to find the possibilities to cure the leaf. If this ration lies between some expected value then we can cure that leaf by medicine and fertilizer, or if the ration really lies between lower levels then it is better to cut it.

For these steps we have masked the coloured pixel with different values. Now we have detected the pixel values of that fault area. Now we have removed that masked pixel and get this picture where we get the exact area of normal leaf. We have used this method 3 times to get more accurate results. The results are shown below in fig. 6.

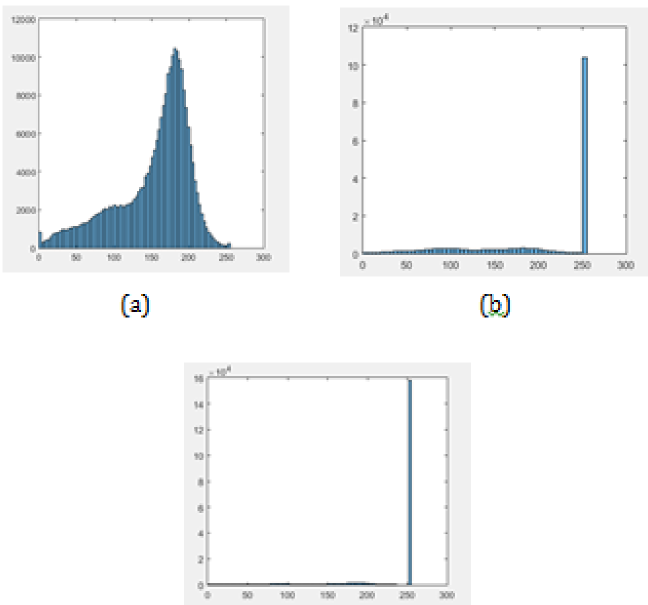

(a) (b)

(c)

Fig - 6: Normal portion extraction after clustering

4.3 Histogram Equalization

We have done this step just verify our results. In the previous steps we have extract the fault area by K-means clustering with higher accuracy. Now we have made sure that the final value is perfect what we actually wanted. In these steps we have done the histogram equalization on three images of a leaf. In the picture shown below fig. 7 (a) we have created the histogram plot of sample leaf and the next picture fig. 7 (b) we extract the histogram plot of fault area and the next pic fig. 7 (c) there is plot of normal portion of that same leaf which we have described in throughout the steps. The results are shown below.



(a) (b) (c)

Fig - 7: Histogram Plot of (a) sample leaf (b) normal portion (c) faulty area

5. CONCLUSION

In this paper we have discussed how important to detect a fault in the leaf and how to improve the mechanism to find out the fault area of a defect leaf. If it is possible to detect the fault then we can cure that portion which will ultimately profitable for agricultural industry. In this paper we have worked on this portion how to detect that fault area with K-means clustering. Behind taking the K-means clustering in our procedure is the accuracy. In the MATLAB 2015-a environment script is written to find the result as our demand. We have work to minimize the problem in previous papers. Now we are working to find the disease type by comparing the previous data input with Data Management. After that it will be possible to find the disease type and give the proper treatment to it. Now we are working on to related projects, we are trying to measure a standard ratio on normal are and fault are of a leaf. If the ratio of a faulty leaf is greater than the standard value it will automatically procedure to cure that leaf and if the ration lies under the standard value then it will suggest to cut that leaf. And the next big project we are working to detect the disease by image analysis and histogram plotting. We are developing an app that will give us the name of that disease and best solution of that disease.

REFERENCES

[1] "World Population Prospects: The 2016 Revision – Key Findings and Advance Tables". United Nations Department of Economic and Social Affairs, Population Division. July 2016. Retrieved 26 June 2017.

[2] Arti N. Rathod, Bhavesh Tanawal, Vatsal Shah "Image Processing Techniques for Detection of Leaf Disease" Computer Engineering, Gujarat Technical University, BVM, V.V.nagar, dist.Anand, Gujarat, India.

[3] Jean Beagle Ristaino: Tracking[Augus2006], "The Evolutionary History Of The Potato Late Blight Pathogen With Historical Collections, Outlooks on Pest Management".

[4] Jon Traunfeld, "Late Blight of Potato and Tomato ,home & garden information center".

[5] Renato B. Bassanezi, José Belasque Junior and Cícero A. Massari "Current Situation, Management And Economic Impact Of Citrus Canker" In São Paulo And Minas Gerais, Brazil.

[6] Jayamala K. Patil1 , Raj Kumar2 "ADVANCES IN IMAGE PROCESSING FOR DETECTION OF PLANT DISEASES".

[7] Brendon J. Woodford , Nikola K. Kasabov and C. Howard Wearing[1999] "Fruit Image Analysis using Wavelets , Proceedings of the ICONIP/ANZIIS/ANNES".

[8] Ahsan Abdullah and Muhammad Umer [2004] "Application of Remote Sensing in Pest Scouting" Evaluating Options and Exploring Possibilities, Proceedings of the 7th International Conference on Precision Agriculture and Other Precision Resources Management, Hyatt Regency, Minneapolis,MN,USA.

[9] Panagiotis Tzionas, Stelios E. Papadakis and Dimitris Manolakis[2005] "Plant leaves classification based on morphological features and fuzzy surface selection technique", 5th International Conference ON Technology and Automation ICTA'05, Thessaloniki, Greece, pp.365-370,15-16 .

[10] M. S. Prasad Babu and B. Srinivasa Rao[2007] "Leaves Recognition Using Back Propagation Neural Network-Advice For Pest and Disease Control On Crops", IndiaKisan.Net: Expert Advisory System.

[11] Alexander A. Doudkin , Alexander V. Inyutin, Albert I. Petrovsky, Maxim E. Vatkin[2007]" Three Level Neural

Network for Data Clusterzation on Images of Infected Crop Field", Journal of Research and Applications in Agricultural Engineering, Vol.52(1).

[12] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu – Xuan Wang Yi – Fan Chang[2007] "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network" , IEEE 7th International Symposium on Signal Processing and Information Technology.

[13] Santanu Phadikar & Jaya Sil[2008] "Rice Disease Identification Using Pattern Recognition Techniques", Proceedings Of 11th International Conference On Computer And Information Technology,25-27

[14] A.Meunkaewjinda, P.Kumsawat, K.Attakitmongcol & A.Srikaew[2008] "Grape leaf disease detection from color imagery system using hybrid intelligent system", proceedings of ECTICON, 2008,IEEE,PP-513-516.

[15] Geng Ying, Li Miao, Yuan Yuan &Hu Zelin[2008] "A Study on the Method of Image Pre-Processing for Recognition of Crop Diseases", International Conference on Advanced Computer Control, 2008 IEEE.

[16] Xinhong Zhang & Fan Zhang[2008] "Images Features Extraction of Tobacco Leaves", 2008 Congress on Image and Signal Processing, IEEE computer society, pp-773-776.

[17] Xu Pengyun & Li Jigang [2009] "Computer assistance image processing spores counting", 2009 International Asia Conference on Informatics in Control, Automation and Robotics, IEEE computer society, pp-203-206.

[18] Di Cui, Oin Zhang, Minzan Li, Youfu Zhao and Glen L. Hartman[2009] "Detection of soybean rust using a multispectral image sensor", SpringerLink – Journal article.

[19] Mew, T.W., 1987. "Current status and future prospects of research on bacterial blight of rice". Annu. Rev. Phytopathol. 25: 359-382

[20] Adhikari, T.B., C.M. Vera Cruz, Q. Zhang, R.J. Nelson, D.Z. Skinner, T.W. Mew and J.E. Leach, 1995. Genetic diversity of Xanthomonas oryzae pv. oryzae in Asia. Appl. Environ. Microbiol. 61: 966-971.

[21] Cai, S. F. and M.Y. Zhong, 1980. "Control experiment of diseases and pests of hybrid late rice [In Chinese]". Zhejiang Nongye Kexue 4: 166-169.

BIOGRAPHIES

Sujan Sarkar is a final year undergraduate student pursuing B.Tech in Electronics and Communication Engineering from Jalpaiguri Government Engineering College. His areas of interest include Digital Circuit Design, VLSI system design, IC fabrication, Communication System, Digital Image processing.

Sanket Biswas is a final year undergraduate student pursuing B.Tech in Computer Science and Engineering from Jalpaiguri Government Engineering College. His areas of interest include Machine Learning, Computer Vision and Deep Learning.

Avinaba Tapadar is a second year undergraduate student pursuing B.Tech in Electronics and Communication Engineering from Jalpaiguri Government Engineering College. His areas of interest include Image processing, Digital System designing, Networking and Cryptography.

Pritam Saha is a final year undergraduate student pursuing B.Tech in Electronics and Communication Engineering from Jalpaiguri Government Engineering College. His areas of interest include Digital Circuit Design, 3-D IC designing, communication and image processing.