Detection of plant diseases and nutritional deficiencies from unhealthy plant leaves using Machine Learning techniques

[1] Debolina Nath ,[2]Dr. Pushan Kumar Dutta, [3]Anup Kumar Bhattacharya

[1][2]Amity School of Engineering and Technology, Amity University Kolkata

[3] Division of Post Harvest Management, ICAR-Central Institute for Subtropical Horticulture, Rehmankhera, P.O. Kakori, Lucknow- 226101, UP

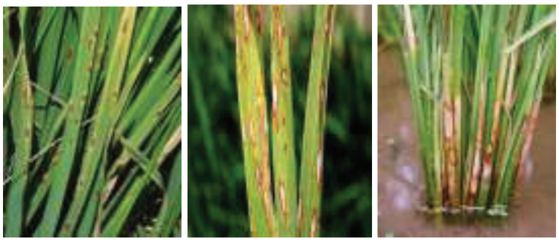
Email: [1]debolinanath8689@gmail.com , [2]pkdutta@kol.amity.edu, [3]dadabhatu@gmail.com

Abstract- In countries like India, which are majorly dependent on an agricultural economy, detection of crop-disease and classification has immense significance. To support agriculture, automatic disease detection and classification of plant/crop disease has become an important aspect today. Manual disease detection requires enormous amount of labour and involves too much processing-time. Therefore, image-processing plays important role for the detection of crop/plant disease from images of diseased leaves. The proposed method is based on the automatic segmentation and classification of different diseased regions of leaf from the plant/crop images, which works in three steps: In the first step, pre-processing is applied to the captured leaf image that is subjected to median filters and contrast stretching in order to remove the digitization noise and enhance the contrast of the image, respectively. In the second step, K-means clustering algorithm is employed on L*a*b* colour model of leaf image that segments pixels into uniform sets. These segments then are converted into HSV colour model and the segment that has maximum hue value among other segments are extracted as a disease segment. Finally, to the extracted diseased segment we will apply machine learning, i.e., Multi SVM classifier to classify crop image into various categorises of diseases. The same methodology is used to predict the Chlorophyll-Nitrogen content of the leaf and use the SVM classifier to show whether there is an 'efficiency' or 'deficiency' of nutrients in the crop.

Index Terms: Image Processing, K-means, Machine learning, SVM

I.INTRODUCTION

Agriculture is major source of livelihood of people in our nation. Surveys reveal that about seventy percent of the Indian population is solely dependent on agriculture for their living, out of which eighty-two percent are poor or marginal farmers [1]. Due to climatic alterations that is occurring across the globe and several other factors that affect agricultural produce, there has been a considerable rise in the number of plant or crop diseases. These diseases are responsible for restricting plant growth and also depreciating the quality as well as the quantity of agricultural produce. Hence, detection of plant diseases or nutrient deficiency in initial stages is of utmost importance. If the disease can be accurately detected, ways can be suggested subsequently to protect the crop from massive damage or crop failure. There are various diseases that may affect crops such as bacterial blight, alternaria alternate, cercospora leaf spot, anthracnose etc. Crop diseases are basically identified by observing different pattern on the parts of the crop like leaf, fruit, and stem. These diseases are mostly found on rice leaves, soyabean leaves, carrot leaves, mango leaves etc. and can be identified from the images of leaves. Often failure of crops also occurs due to nutritional deficiencies. This can also be predicted based on Chlorophyll and Nitrogen content of the leaves. The colour of the leaf is associated with the level of nitrogen and chlorophyll content present in the leaf. Chlorophyll and Nitrogen affects the green colour of the plant and determines their biomass yield [2]. Plants that are sufficiently supplied with Nitrogen are green and very healthy, while those insufficiently supplied with Nitrogen are pale green or yellow in colour and remain small and stunted. Hence, leaf colour has led to exploit this property by using image processing analysis to detect Chlorophyll and Nitrogen content in plants. This can be done using modern technology that will benefit the farmers and save them from suffering major financial losses. Since paddy(rice) is India's most important food and various varieties of paddy is extensively *cultivated across country, therefore in this paper we will be mainly concentration upon the detection and classification of diseases in paddy from the images of the unhealthy leaves of the crop by image processing using machine learning technique. Rice suffers from various diseases such as Rice Blast, Brown Spot of rice, Neck Blast, Sheath Blight of Rice etc. caused by their respective pathogens [3].*



(a)Brown spot (b)Rice blast (c)Sheath Blast

Fig .1: Different diseases in Rice crops

Manual disease detection is an extremely cumbersome task requiring a lot of processing time and manpower. Identification of these leaf diseases through naked eye is often prone to high error rates and faulty classification. In this scenario, modern farming techniques play a significant role and thus image processing [4][5] can be used for the crop disease detection with greater accuracy. In our paper, we aim to capture the image of the diseased part of the plant or the diseased leaf and use this image for the detection. We will be applying image pre-processing to enhance the quality of the captured image and remove noises from the images. This will be followed by segmentation using K-means and subsequent classification done by SVM technique that will finally provide us the result of our search. SVM is adopted in our system because SVM's are very fine when one has no idea on the data and work well with semi-structured data like Images. Once the disease gets detected and properly identified, hence measures can be suggested to save the crop from further damage. It is pertinent to mention that along with the detection of disease, we will be predicting the chlorophyll content of the leaf as well as the nitrogen content to the plant and then subjecting it to SVM to categorize whether it falls under "efficiency" or under "deficiency". This is help to reflect is the plant is suffering from any kind of nutritional deficiency and the farmers can add the required fertilizers in adequate amount or take other necessary measures. The detailed working of our proposed mechanism is discussed in section III of the paper. Thus, our system aims to provide a high-speed, accurate and inexpensive method in detecting and classifying leaf diseases. The remainder of the paper is organized as follows: In section II, we have discussed some previous works carried out in this field as Literature Review. In section III comprises the proposed mechanism of our project. In section IV, implementation of the mechanism has been discussed followed by the Results and Analysis in section V and Conclusion and Future Works mentioned in section VI.

II. LITERATURE REVIEW

Even in today's age of technological advancement, when technology has entered every domain of human life and existence and is being extensively used in all fields starting from healthcare to education, farmers still use the traditional methods of disease detection in crops in most parts of the world. Thus, to facilitate agriculture, several works have been conducted previously in this domain as automatic disease detection in plants is one of the major technological advancements in the field of agriculture. In all the research works that have been conducted over the years, the primary aim was to develop a mechanism for easy and accurate detection of diseases in plants or crops. Thus, in this section we will discuss some of the interesting works that have been conducted in this field. In [4], the authors proposed a Vision-based detection algorithm with masking the green-pixels and colour co-occurrence method. They describe that there are mainly four steps in developed processing scheme, out of which, first one is, for the input RGB image, a colour transformation structure is created, because this RGB is used for colour generation and transformed or converted image of RGB, that is, HSI is used for colour descriptor. In second step, by using threshold value, green pixels are masked and removed. In third, by using pre-computed threshold level, removing of green pixels and masking is done for the useful segments that are extracted first in this step, while image is segmented. And in last or fourth main step the segmentation is done. Spatial gray-level dependence matrices (SGDM) method has been used for extracting statistical texture features. While [4] mainly concentrated on the colour of the leaves and utilized colour co-occurrence method for the segmentation, [5] mainly concentrates on the texture of the leaves for the segmentation and consequent detection, though they follow a similar approach towards the problem. Here they consider edge based, region based and threshold-based segmentation and feature based clustering. An overview of various techniques employed in the segmentation as well as the classification procedure have also been discussed. [6] presents an algorithm for image segmentation technique used for automatic detection as well as classification of plant leaf diseases and survey on different diseases classification techniques that can be used for plant leaf disease detection. In this paper, Image segmentation is done by using genetic *algorithm. In [7], the methodology used for detecting plant-diseases early and accurately using diverse image processing techniques has been proposed where Gabor filter has been used for feature extraction and ANN based classifier has been used for classification. In [8], the method of K-means clustering along with BP neural networks was applied for the clustering and the classification of diseases that can be identified from plant leaves. The proposed algorithm has been tested on five diseases viz. Early and late scorch, cottony and ashen mould, tiny whiteness. In [9], it was proposed that a Machine learning based recognition system will prove to be very useful as it saves efforts, money and time too. The approach given in this for feature set extraction is the colour co-occurrence method. For automatic detection of diseases in leaves, neural networks were used. The approach proposed can significantly support an accurate detection of leaf, and seems to be important approach, in case of stem and root diseases, putting fewer efforts in computation. A similar approach for disease detection is applied in [5] which also uses neural network-based classification technique. In [10], principal component analysis (PCA) has been performed for reducing dimensions in feature data processing, then back-propagation (BP) networks, radial basis function (RBF) neural networks, generalized regression networks (GRNNs) and probabilistic neural networks (PNNs) has been used as the classifiers to identify diseases. In [12] median filtering method was applied to obtain the filtered and segmented image from the captured image of the leaf. The pre-processed image is converted to its binary image and then feature extraction is carried out followed by the final classification. The scope of application of [12] is mainly restricted to pomegranate leaves and [13] is restricted to grape leaves but in both of these, an image processing technique was primarily applied followed by the use of the SVM classifier. However, in comparison to all the works already done in this particular field of automatic disease detection in plants, our approach is unique is its own way as we are not only classifying the plants/ crops on the basis of the disease they are affected by but at the same time our system determines the chlorophyll as well as nitrogen content to the leaves and determines whether they are nutritionally efficient or deficient.*

III.PROPOSED MECHANISM

Detection and classification of disease from leaf image. The mechanism that is proposed by us is discussed in a series of steps that will assist in the segmentation and classification of the diseased leaves from the crop images. The methodology will be dealt with in the following three phases:

(a) Pre-processing of image

(b) Segmentation based on K-means clustering

(c) Classification of diseases based on Support Vector Machine (SVM)

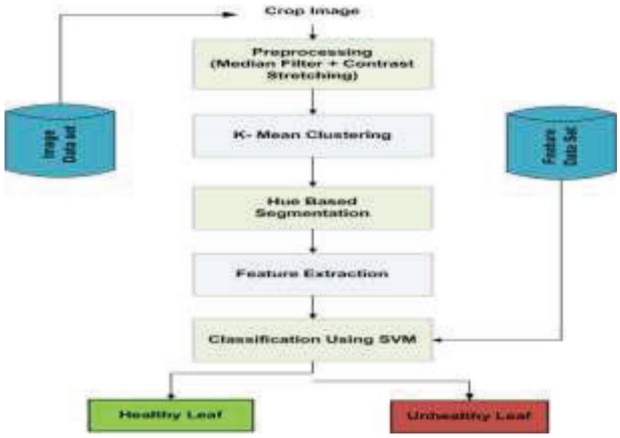


Fig.2: Flowchart of the Proposed Mechanism

The initial step is to acquire the image of the diseased leaf. This image is then passed for pre-processing. The pre-processing of the image is carried out to enhance the quality of the captured image of the diseased leaf. This will be followed by the removal of noise from the image. By the use of median filters, the salt and paper noise is removed. Further the contrast of the image is also increased to demarcate the different regions of the diseased leaf in a better way. In the second phase of the mechanism, the obtained image of the diseased leaf is split into multiple clusters using K-means clustering.

III a. K-Means clustering

K-means method has a good effect on the segmentation of plant leave disease. It is an unsupervised classification algorithm that needs the manual requirement to enter the number of classes plus randomly distribution of initial centroids of each class. In our approach, the k-means clustering is exercised to assign labels to every pixel in a input image. RGB, or HSV Colour model does not provide a correct depiction of an allotment of colour areas. plane a and plane b has Euclidian distance directly proportional to the visual similarity of the colour so the clustering is performed only in "a", "b" space,. The "L" component stands for luminous.

*Working of k-means clustering is given as follows: Input the image as L*a*b* colour model.*

Decompose the image into L , a *and* b *colour component.*

Convert component a *and* b *into 1D array.*

Combine the color plane a and b into matrix such as x axis represent a component and y axis represent b component.

Choose the number of clusters (say K). Repeat steps 6 and 7 until cluster membership become stable.

Make a new partition by allocating each point to its nearby cluster centre.

Compute new cluster centres using the Euclidean distance metric.

$$d = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (2)$$

Where d is the distance between a point $X(x1, x2, \ldots xn)$ and $Y(y1, y2, \ldots, yn)$

(1) *Repeat the step 5 to 7 three times to avoid local minima.*

(2) *Finally we have every pixel assigned with cluster label.*

III b. Feature Extraction

Extracting the features from the given image is one of the important steps required for classification of objects into various classes. Colour, texture, shape, morphology etc. are the primary features of the image by which we can easily classify plant disease effectively [15]. In this paper we have chosen some of these features for disease detection.

III c. Colour Features

First order, second order and third ordered colour moments known as mean, standard deviation and skewness, respectively, are selected as a colour features for detecting plant disease due to their effective representation of colour distribution of the image. for RGB colour image we calculate these features separately for each channel. So in total we have 3 x 3 = 9 features for given image.

Suppose the colour value of the i^th plane positioned at the jth image pixel is ρ_{ij}, then the colour moments are calculated as given below:

Moment 1 : Mean

$$E_i = \frac{1}{N} \sum_{j=1}^{N} \rho_{ij} \quad (3)$$

Moment 2: Standard Deviation

$$\boldsymbol{\sigma_i} = \sqrt{\left(\frac{1}{N}\sum_{j=1}^{N}\left(\rho_{ij} - E_i\right)^2\right)} \quad (4)$$

Moment 3: Skewness

$$S_i = \sqrt[3]{\left(\frac{1}{N}\sum_{j=1}^{N}(\rho_{ij} - E_i)^3\right)} \quad (5)$$

III d. Texture Features

Texture can be thought of as repetitive blueprint of pixels on top of spatial domain.

Energy:

$$f_1 = \sum_i \sum_j \{p(i, j)\}^2$$

Contrast:

$$f_2 = \sum_i \sum_j |i - j|^2 \, p(i, j).$$

Correlation:

$$f_3 = \frac{\sum_i \sum_j (i, j)\, p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}.$$

(8)

Where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the means and standard deviations of p_x, p_y

Homogeneity:

$$f_4 = \sum_i \sum_j \frac{p(i, j)}{1 + |i - j|}.$$

(9)

$$f_5 = M^2 - \frac{\sum_{(i,j)\in B} I(i, j)}{Max_{(i,j)\in B} I(i, j)}$$

Where B specifies block size of $M \times M$. Where $I(i, j)$ indicates the value at a location (i, j) in the image.

III e. Support Vector Machine

SVM is a binary classifier that classifies two classes based on hyper-plane which is nothing just decision boundary between two classes [18]. In SVM the objective is to maximize the boundary (margin) between two classes. Those points which are closest to the boundary, and they will be selected for determining the hyper plane is called *support vectors. Fig 3 shows the concept of SVM. The standard form of SVM was intended for two-class problems. Multi class SVM focus on labelling the instances by obtaining the decision boundary from a finite collection of various instances. It can easily be done by converting the single multiclass problem into multiple binary classification problems.*

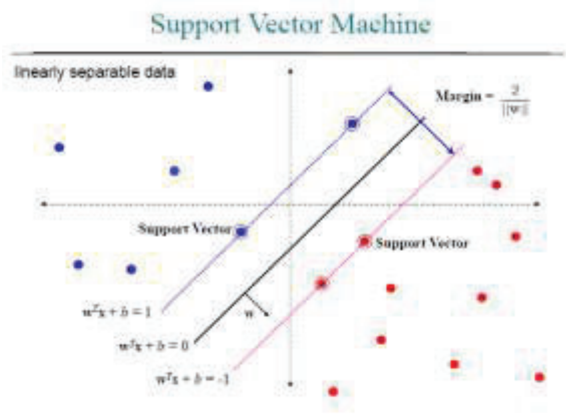


Fig.3: SVM Diagram

There are two way for doing multiclass classification by either using one-to-one or one-to-many. Winning class be determined by the output function that has highest value. For example, in one-to-one approach binary SVM is trained for every two classes of data to construct a decision function. Hence there are k (k−1)/2 decision functions for the k-class problem. Suppose k = 10, 45 binary classifiers need to be trained. This suggests large training times. In the classification stage, a voting policy is exercised where the testing point is chosen to be in class having the highest number of votes. Now the main aim is to identify the cluster containing the diseased part of the leaf. In order to do so, the all the clusters obtained from the image are converted from RGB (red green blue) colour model to the HSV (Hue Saturation Value) colour model because in HSV model an object with a specific colour can be easily identified and can overcome the effect of light intensity from the exterior. From this model, the mean hue values of all the clusters that are obtained are determined. The mean hue value of the cluster containing the disease portion of the leaf will be the maximum. Therefore, the cluster which has the greatest mean hue value out of all the clusters is segmented as the diseased cluster. In the final phase, the classification of the disease is to be done. In order to do so, multi SVM classifier is applied. It will classify the captured image into the various classes of diseases based on various crop disease images. Prediction of nutritional efficiency or eficiency on the basis of chlorophyll content and nitrogen content of the plant from leaf image.The initial step is to acquire the image of the leaf. This image is then passed for pre-processing. The pre-processing of the image is carried out in order to enhance the quality of the captured image of the diseased leaf. This will be followed by the removal of noise from the image. By the use of median filters, the salt and paper noise is removed. Further the contrast of the image is also increased to demarcate the different regions of the diseased leaf in a better way. For the calculation of the chlorophyll content, this image is converted from RGB model to the Greyscale model. The threshold value of the Greyscale level is set and then this Greyscale image is converted into binary image. The image is then inverted and it is cropped to extract the leaf and its various features. The aspect ratio is then calculated and the eccentricity of the leaf is determined. Then from the RGB model of the image, the mean hues of R, G and B are determined and from the obtained values, the chlorophyll content can be calculated by the its mathematical formula. Following a similar procedure, the nitrogen content of the leaf can also be determined. In this case, the image is converted from the RGB model to the HSI (Hue, saturation, intensity) model. From this, the respective mean values of R, G and B are determined using their respective mathematical formulas which are further utilized for the calculation of the nitrogen content of the plant. Once the nitrogen and chlorophyll percentages are determined, then in the final phase, multi SVM classifier is applied for the classification which shows two classifications, either "efficiency" and "deficiency". Based on the results obtained, the farmers can take necessary steps to safeguard their crop or to enhance their quality.

IV.IMPLEMENTATION

The mechanism discussed in section III is implemented using GNU Octave / MATLAB. The source code is run on GNU Octave / MATLAB and the results are obtained in a separate output window.

V.RESULTS AND ANALYSIS

Fig 4(a)

Fig 4(b)

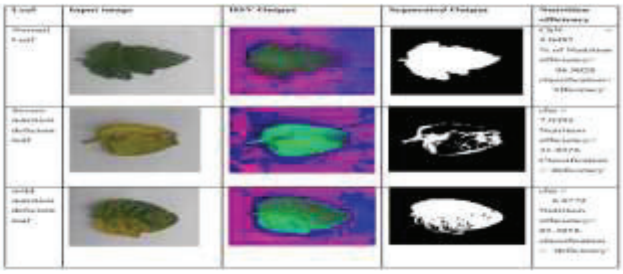


Fig 4(c)

VI. CONCLUSION AND FUTURE WORK

In the field of agriculture, major losses are encountered due to crop diseases which lead to frequent crop failure and decreased production. The aim of our earlier work in the 3rd smart cities conference[16] was identification of the classifiers for leaf spot identification but we found that the nutrition and chlorophyl content will be useful tool to categorize leaves even with spots which we have identified in this paper. The future work of this paper is to improve the system to identify more nutrient contents with one algorithm and to use multiclass classifiers such as neural networks and neuron fuzzy classifiers to classify the leaves.

REFERENCES

[1] http://www.fao.org/india/fao-in-india/india-at-a-glance/en

[2] Sheng Wang, Kaiyu Guan, Zhihui Wang, Elizabeth A Ainsworth, Ting Zheng, Philip A Townsend, Kaiyuan Li, Christopher Moller, Genghong Wu, Chongya Jiang, "Unique contributions of chlorophyll and nitrogen to predict crop photosynthetic capacity from leaf spectroscopy", Journal of Experimental Botany, Volume 72, Issue 2, 2 February 2021, Pages 341–354, Published:16 September 2020

[3] Francisco Elazegui , Zahirul Islam, "Diagnosis of Common Diseases of Rice", International Rice Research Institute, 2003.

[4] Prof. Sanjay B. Dhaygude, Mr.Nitin P.Kumbhar, "Agricultural plant Leaf Disease Detection Using Image Processing" IJAREEIE, vol. 2(1), pp. 599-602, January 2013.

[5] Kiran R.Gayhale , U.Gawande "An Overview of the research on plant leaves disease detection using image processing techniques" , IOSR Journal of Computer Engineering, Published 2014

[6]Vijai Singh ; Varsha ; A K Misra "Detection of unhealthy region of plant leaves using image processing and genetic algorithm", International Conference on Advances in Computer Engineering and Applications, March 2015

[7] Anand H. Kulkarni, Ashwin Patil R. K., "Applying image processing technique to detect plant diseases", International Journal of Modern Engineering Research, vol.2, Issue.5, pp: 3661-3664, 2012.

[8] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases" IJCA, vol. 17(1), pp. 31-38, March 2011, IEEE-2010

[9]Mrunalini R. Badnakhe, Prashant R. Deshmukh, "Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering" IJARCSSE, vol. 2(3), pp. 449-452, 2012.

[10] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li , Image Recognition of Plant Diseases Based on Principal Component Analysis and Neural Networks, 8th International Conference on Natural Computation, pp-246-251, Chongqing, China, 2012

[11] Saradhambal G, Dhivya R, Latha S, R. Rajesh, "Plant Disease Detection and Its Solution Using Image Classification", International Journal of Pure and Applied Mathematics, Volume 119, No. 14, 2018.

[12] Sowmya G.M, Chandan V, Sampath Kini, "Disease Detection in Pomegranate Leaf, using Image Processing Technique", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 6, Issue 3, March 2017

[13] Pranjali B. Padol, Prof. Anjali A. Yadav, "SVM Classifier Based Grape Leaf Disease Detection", 2016 Conference on Advances in Signal Processing (CASP), Cummins College of Engineering for Women, Pune, June 9-11, 2016.

[14] Manoj Mukherjee, Titan Pal, Debabrata Samanta, "Damaged Paddy Leaf Detection using Image Processing", Journal of Global Research in Computer Science, Volume 3, No. 10, October 2012.

[15] M.N. Abu Bakar, A.H. Abdullah, N. Abdul Rahim, H. Yazid, S.N. Misman, M.J. Masnan, "Rice Leaf Blast Disease Detection Using MultiLevel Colour Image Thresholding", Journal of Telecommunication, Electronic and Computer Engineering, Volume 10, No. 1-15.

[16] V. K. Trivedi, P. K. Shukla and P. K. Dutta, "K-mean and HSV model based segmentation of unhealthy plant leaves and classification using machine learning approach," 3rd Smart Cities Symposium (SCS 2020), 2020, pp. 264-270, doi: 10.1049/icp.2021.0953.