Detection and Classification of Plant Diseases

Mr. N.S. Bharti, Prof. R.M. Mulajkar.

[1] *ME Student, Electronics and Telecommunication, J.C.O.E. Kuran, Maharashtra, India*
[2]*Professor, Electronics and Telecommunication, J.C.O.E. Kuran, Maharashtra, India*

Abstract – We proposed software solution for automatic classification and detection of plant leaf diseases. Which is an improvement to the solution proposed in [1] as it will be able to provide quick and more accurate solution. The process consists of four main phases as mentioned in [1]. The following extra two steps are required to add successively after the segmentation phase. In the first step we find the mostly green colored pixels. And in second step, these green pixels are masked based on their specific threshold values which will computed using Otsu's method, then those mostly green pixels are masked. The other additional step is that the pixels with zeroes R.G.B. values and the pixels on the boundaries of the infected cluster are completely removed. The experimental results indicate that the proposed technique is a fast and accurate technique for the detection of plant leaves diseases. The proposed approach can successfully detect and classify the examined diseases with a precision between 83% and 94%, and able to achieve 20% speedup over the approach proposed in [1].

Key Words: SGDM Matrix, HSI, Neural Networks, Color Co-occurrence Method, K-means.

1. INTRODUCTION

Plant diseases have turned into a crucial as it can cause significant reduction in both quality as well as quantity of agricultural products [20]. It is estimated that 2046 plant disease affected in Georgia (USA) is approximately $1039.74 million. Of this amount, around 185 million USD was spent on controlling the leaf diseases, and the rest is the value of damage caused by the diseases.

The naked eye observation of experts was the main approach adopted in practice for detection and identification of plant leaf diseases [20]. However, this method requires continuous monitoring by experts which might be large expensive in large area of farms. Further, in some developing countries, farmers may have to go long distances to contact experts for identification, this makes consulting experts too expensive and time consuming [14; 5; 8].
Automatic detection of plant diseases is a very important research topic as it may prove benefits in monitoring large fields of crops at early stage, and thus automatically detect the symptoms of diseases as they appear on plant leaves [1; 18; 8]. Therefore; looking for accurate and automatic, less expensive and accurate method to detect plant disease cases is of great realistic significance [14; 5]. Machine based on detection and recognition of plant diseases can provide clues to identify and treat the diseases in its early stages [8; 18].

Three are two main characteristics of plant-disease detection software based methods that must be achieved, they are: speed of detection and accuracy in finding the disease. In this paper an automatic detection and classification of leaf diseases is proposed, this method is based on ANNs as a classifier tool using and K-means as a clustering procedure. This method is advanced version of proposed method in [1]

2. THE PROPOSED APPROACH.

The overall concept used for image classification is almost the same. First, the digital images are acquired from the environment using a data storage device or by digital camera. Then image-processing techniques which will be applied to the acquired images to extract useful features that are necessary for further observations. After that, several techniques are apply to classify the images according to the specific disease. Figure 1 shows the basic procedure of the proposed approach.

Image acquisition
Image preprocessing
Image segmentation
Feature extraction
Statistical analysis
Classification based on a classifier

Chart -1: The basic procedure of the proposed image processing- based disease detection solution

The proposed steps are illustrated in Algorithm 1. In the initial step, the RGB images of all the samples of leaf images were picked up. Some real samples are shown in Figure 1. We can identify by Figure 1 that leaves belonging to early scorch, cottony mold, ashen mold and late scorch have significant differences form greasy spot leaves in terms of *color* and *texture*. Also, Figure 2 shows two images; the right side image is a normal image and the left side image is infected with tiny whiteness disease. However, the impacted pattern related to these six classes (early scorch, cottony mold, ashen mold, late scorch, tiny whiteness and normal) had very small differences as it cannot identify to the necked human eye, which may justify the misclassifications.

a) Ashen mold

b) Late scorch

c) Early scorch

d) Cottony mold

Fig -1: Sample images from our dataset indicating with type of disease

Fig -2: A leaf image infected with tiny whiteness disease (left) and a normal leaf image (right).

In step two a color transformation structure is created for the RGB leaf image and then, a device-independent color space transformation for the color transformation structure is applied in step three. Steps 2 and 3 are needed for carrying out step 4. In this step the images in database are segmented using the K-Means clustering technique [10; 7; 3; 9]. Same four steps combines in phase 1 whereas, the infected objects are determined. In step five, we are identifying the mostly green colored pixels. After that, based on specified and varying threshold value which is computed for current pixels using *Otsu's method* [12; 13], these mostly green pixels are masked as follows: if the green component intensities of pixel is less than the pre-computed threshold value, the RGB components of the this pixel is assigned to a value equal to zero. This is done because that these pixels have no valuable information to the disease identification and classification steps, and most probably those pixels represent healthy areas in the leave. Hence, the image processing time should become significantly reduced.

In step 6 the pixels with zero RGB values and the pixels on the boundaries of the infected cluster (object) were completely removed out. Phase 2 formed by steps 5 and 6, and this phase is very helpful to gives more accurate disease classification and identification results with satisfied performance and the overall computation time should become vary less. The observations behind steps 5 and 6 were experimentally validated. Next, in step 7 the infected cluster is then then converted to HIS format from RGB format. In the next step, the SGDM matrices are then computed for each pixel map of the image for only H and S images. The SGDM is a measure of the probability that the given pixel at one particular gray-level will occur at a distinct distance from other pixel and orientation angle from another pixel, given that pixel has a second particular gray-level other than previous one. By using the SGDM matrices, the texture statistics for each image were generated. Concisely, the features set are computed only for pixels inside the boundary of the infected areas of the leaf. In other words, healthy areas also inside the infected areas were removed. Steps 7 – 10 form phase 3 in which the identification of texture features for *the segmented infected objects* in this phase are calculated. Final step is, the recognition process in the fourth phase was performed to the extracted features by using a pre-trained neural network. For each image in the data set the all required steps were repeated. The Proposed approach for segmentation and classification plant diseases can be divided into four phases:

2.1 Phase 1 - K-means Clustering Technique.

There will be the two main preprocessing steps that are needed in order to implement the K-means clustering algorithm: The phase starts first by creating device-independent color space transformation structure as mentioned in phase 1. In a device independent color space, the coordinates used to specify that the color will produce the same color regardless of the device used to draw it. Thus, we created the color transformation structure which will defines the color space conversion. Then, we applied the device-independent color space transformation, which converts the color values in the image to the color space specified in the color transformation structure. The color transformation structure specifies various no of parameters of the transformation. A *device dependent color space* is the one where the resultant color depends on the equipment used to produce it. For example the color produced using pixel with a given RGB values will be altered in the form of brightness and contrast on the display device used. Thus the RGB system is a color space that is dependent.

The K-means clustering algorithm tries to classify *objects* (pixels in our case) based on a set of features into K number of sub classes. The classification is done by minimizing the *sum of squares* of distances between the objects and the corresponding cluster or class *centroid* [10; 7]. However, by using K-means clustering leaf image partitioned into four clusters in which one or more then one clusters contain the disease in case when the leaf is infected by more than one disease. In our experiments multiple number of values clusters have been tested. Best results were observed when the number of clusters was 3 to 4. A stem image infected by early scorch and its first cluster is shown in Figure 3. For the identification of such type of leaf disease we are put forward the six clustered area which contain the impacted area as well as the background of the captured image. All five are finally clustered on a single gray-scale image with the pixels colored based on their cluster index.

a) Original image b) cluster 1 image
Figure 3: A stem image infected with early scorch;

2.2 Phase 2 – Masking the green pixels and the pixels on the boundaries

In this phase we work on two steps: Identification of the mostly green colored pixel, and then threshold the global image using *Otsu's method* [12; 13] has been applied in order to specify the varying threshold value which chooses the threshold for minimize the variance of the threshold black and white pixels. Next step, the green pixels are masked as follows: if the green component of pixel intensities is less than the available threshold value, then, the red, green and blue components of the this pixel are cleared. The next step in this phase is focused on deleting both the pixels with zero component values and the pixels on the boundaries of the infected cluster/ clusters.

2.3 Phase 3 – Features Extraction

This method is implemented for extracting the feature set is called the *Color Co-occurrence Method* or CCM method in short. It is a method, in which both the color as well as texture of an image are taken into account, to reach at unique disease pattern, which is represented in that image.

2.3.1 Co-occurrence Methodology for Texture Analysis

The CCM methodology for this work will consists of three major mathematical processes. First, the RGB images of leaves are converted into HSI color space representation. Once this process is completed, a color co-occurrence matrix for each pixel map is generate, resulting in three CCM matrices, one for each of the H, S and I pixel maps. (HSI) space is also a popular color space because it will be based on human color perception ability [17]. Color spaces can be transformed from one space to another easily.

While transformation processes is done, we calculated the feature set for H and S, we dropped (I) since it does not provide extra information. However, we use GLCM function in MATLAB to form gray-level co-occurrence matrix; the gray levels is set to 8, and the symmetric value is set to "true", and finally, offset is given a" 0" value.

2.4 Texture Features Identification

The following features set were computed for the components H and S:

The angular moment (E) is used to measure the homogeneity of the image, and is defined as shown in Equation 1.

$$E = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} [p(i,j)]^2 \quad \text{.............. (1)}$$

The product moment (cov) is analogous to the covariance of the intensity co-occurrence matrix and is defined as shown in Equation 2.

$$cov = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i - I_2)(j - I_2) P(i,j) \quad \text{.............. (2)}$$

The sum and difference entropies (and) which are computed using Equations 3 and 4 respectively.

$$se = \sum_{k=0}^{2(N_g-1)} P_{x+y}(k)\ lnP_{x+y}(k) \quad \text{.................... (3)}$$

$$de = \sum_{k=0}^{N_g-1} P_{x-y}(k)\ lnP_{x-y}(k) \quad \text{..................... (4)}$$

The entropy feature (e) is a measure of the amount of order in an image, and is computed as defined in Equ. 5

$$e = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j)\ lnP(i,j) \quad \text{....................... (5)}$$

The information measures of correlation (*inf2*) is defined as shown in Equation 6.

$$inf2h = \left[1 - e^{-2(Hxy-e)}\right]^{1/2} \quad \text{........................... (6)}$$

Where:

$$Hxy = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_x(i)P_x(j)\ ln[P_x(i)P_x(j)]$$

Contrast (*id*) of an image can be measured by the inverse difference moment as shown in Equation 7.

$$id = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{P(i,j)}{1+(i-j)^2} \quad \text{........................... (7)}$$

Correlation (is a measure of intensity linear dependence in the image and is defined as shown in Equation 8.

$$c = \frac{\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} ijP(i,j) - I_2^{2}}{I_3} \qquad \text{.......................... (8)}$$

2.5 Phase 4 – Neural Networks

In this paper, neural networks are used in the automatic detection of leaf as well as leaves diseases. Neural network will be used as a Classification tool due to its well-known technique we can used it as successful classifier for many real applications. The training and validation processes are the two important steps in developing an accurate process model using NNs. The dataset consists of two parts for training and validation processes; the training feature set which are used to form the NN model; while a testing features sets are verify the accuracy of the trained NN model.

Before the data can be apply to the ANN model, the proper network design must be configured, including type of the network and method of training. Which can be followed by the optimal parameter selection phase. However, this same phase was carried out simultaneously with the network training phase, in which the network was trained using the feed-forward back propagation network. In the training phase it is required that connection weights were always updated until they reached the defined iteration number or acceptable error. Hence, the capability of ANN model to respond accurately was improved using the Mean Square Error (MSE) criterion to emphasis the model validity between the target and the network output.

3 EXPERIMENTAL RESULTS AND OBSERVATIONS

3.1 Input Data Preparation and Experimental Settings

In our execution, two main files are generated, namely: (i) Training texture feature data, and (ii) Testing texture feature data. Those two files had 192 rows each, which representing 32 samples from each of the six classes of leafs. Each row had 10 columns representing the 10 texture features extracted for a particular sample image. Each row had a unique index number (1, 2, 3, 4, 5 or 6) which represented the disease type of the particular row of data. First one represented early scorch disease infected leaf. Second one represented Cottony mold disease infected leaf. Next represented ashen mold disease infected leaf. Next will represented late scorch disease infected leaf. Next represented tiny whiteness disease infected leaf and Last one represented normal leaf. Then, a code is written in MATLAB that would take in .mat files representing the training and testing data, train the classifier using the "train files", and then to perform testing use the "test file" to perform the classification task on the test data. Simultaneously, a MATLAB routine would load all the data files and make modifications to the data according to the proposed model chosen. In the experimental results, the threshold value for each of the above categories is constant for all samples infected with the same disease. This threshold is a global image threshold that is computed using Otsu's method [12; 13]. The architecture of the network used in this study was as follows. A set of 10 hidden layers in the neural network was used with the number of inputs to the neural network (i.e. the number of neurons) is equal to the number of texture features listed above. The number of output is 6 which is the number of classes representing the 5 diseases studied along with the case of normal (uninfected) leaf. Those diseases are cottony mold, early scorch, late scorch, ashen mold, tiny whiteness. The neural network used is the feed forward back propagation with the performance function being the Mean Square Error (MSE) and the number of iterations was 10000 and the maximum allowed error was 10^{-5}.

3.2 Experimental Results

The NN classification strategy for testing samples were given in Table 1. These results are obtain by using a NN classifier for different set of five diseases. In particular, model implemented in same achieved the highest overall classification accuracy, in which it achieved an overall accuracy of 94% which is high as compared to the 89.5% accuracy achieved in [1]. Also, it can be concluded that this model M1 is the best in overall model in this classifier in terms of accuracy and in computational time for both training and classification. In conclusion, Table 2 reported better classification accuracies for all the data models.

Table -1: Percentage classification of various diseases

Model	Color Features	Early scorch	Cottony mold	Ashen mold	Late scorch	Tiny whiteness	Normal	Overall average
M1	HS	98	96	89	91	92	100	94.33
M2	H	90	92	86	89	93	98	91.33
M3	S	90	89	85	89	81	98	88.67
M4	I	92	89	84	88	86	99	89.67
M5	HSI	81	84	78	79	81	99	83.67

The recognition rate for NN classification strategy of all models was also computed for all disease mentioned above based upon all the working phases, the obtained results for M 1 and M5 are reported in Table 2.

Table -2: Recognition rates of individual plant diseases

Model	Color Features	Early scorch	Cottony mold	Normal	Overall Average
M1	HS	99	100	100	99.66
M5	HIS	96	98	100	98.00

It can be seen from Table 2 that Model M1 which has used only the H and S components in computing the texture features, has observed as the best model among the various models. Furthermore, it can be observed from Table 2 that Model M1 has recognition rate more than Model 5; this is because of the elimination of the intensity from computing the texture features in Model M1. As a matter of fact, elimination of intensity is implemented in this study because it nullifies the effect of intensity variations. The numbers of leaf samples that were classified into each of the five tested disease categories using model M1 with specific threshold value are shown in Table 3. It is observed from Table 4 that only few samples from late scorch and tiny whiteness leaves are get misclassified, also, three test images were misclassified for the case of late scorch infected leaves. Similarly, in the case of tiny whiteness images, only two test images from the class were misclassified. In average, accuracy of classification using same approach was 94.67 compared to 92.7 in case of using the approach presented in [1].

Table -3: Classification results per class for neural network with back propagation.

species	Early scorch	Cottony mold	Ashen mold	Late scorch	Tiny whiteness	Normal	Accuracy
Early Scorch	25	0	0	0	0	1	100
Cottony Mold	24	0	1	0	0	0	96
Ashen Mold	0	0	25	0	0	1	100
Late Scorch	0	0	0	22	1	0	88
Tiny Whiteness	0	1	0	0	23	0	92
Normal	0	0	0	2	1	23	92
Average							94.67

The convergence curve by the neural Network in the proposed system is better than that of [1] as shown in Figure 4. In Figure 4 Approach 1 represents the study presented in [1], while Approach 2 represents our proposed study.

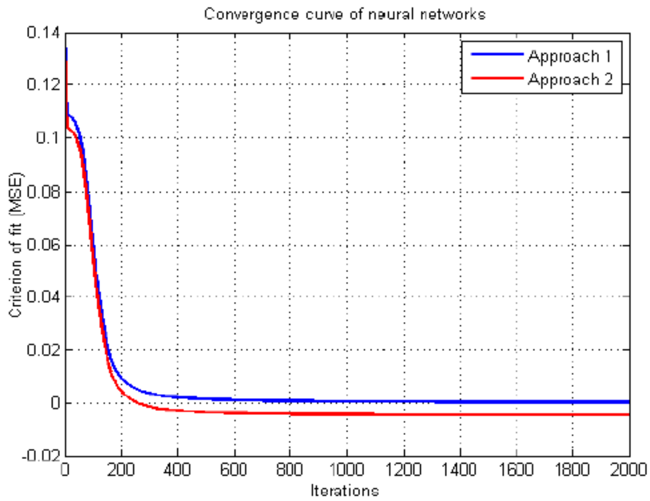

Fig -4: The convergence curve of [1] that is "Approach 1", and our approach "Approach 2"

3. CONCLUSIONS

In this paper, respectively, the applications of K-means clustering with Neural Networks (NNs) had been implemented for clustering and classification of diseases that effect on plant leaves. Recognizing the leaf disease or leaves disease is mainly the purpose of the proposed approach. Thus, the proposed Algorithm was tested on five diseases which influence on the plants; they are: ashen mold, Cottony mold, early scorch, late scorch, tiny whiteness. The experimental results indicate that the proposed approach is accurate approach, which can able to support an accurate detection of leaf diseases in a little computational effort.

REFERENCES

[1]. Al-Bashish, D., M. Braik and S. Bani-Ahmad, 2011. Detection and classification of leaf diseases using Kmeans-based se mentation and neural-networks-based classification. Inform. Technol. J., 10: 267-275. DOI: 10.3923/itj.2011.267.275

[2]. Aldrich, B.; Desai, M. (1994) "Application of spatial grey level dependence methods to digitized mammograms," Image Analysis and Interpretation, 1994., Proceedings of the IEEE Southwest Symposium on, vol., no., pp.100-105, 21-24 Apr 1994. DOI: 10.1109/IAI.1994.336675

[3]. Ali, S. A., Sulaiman, N., Mustapha, A. and Mustapha, N., (2009). K-means clustering to improve the accuracy of decision tree response classification. Inform. Technol. J., 8: 1256-1262. DOI: 10.3923/itj.2009.1256.1262

[4]. Bauer, S. D.; Korc, F., Förstner, W. (2009): Investigation into the classification of diseases of sugar beet leaves using multispectral images. In: E.J. van Henten, D. Goense and C. Lokhorst: Precision agriculture "09. Wageningen Academic Publishers, p. 229-238. URL: http://www.precision-crop-protection.unibonn. de/gk_research/project.php?project=3_09

[5]. Camargo, A. and Smith, J. S., (2009). An image processing based algorithm to automatically identify plant disease visual symptoms, Bio systems Engineering, Volume 102, Issue 1, January 2009, Pages 9-21, ISSN 1537-5110, DOI: 10.1016/j.biosystemseng.2008.09.030.

[6]. Ford, A. and Roberts, A., (2010) Colour Space Conversions, August 11, 1998. http://www.poynton.com/PDFs/coloureq.pdf, viewed on May 2010.

[7]. Hartigan, J. A.; Wong, M. A. (1979). "Algorithm AS 136: A K-Means Clustering Algorithm". Journal of the Royal Statistical Society, Series C (Applied Statistics) 28 (1): 100–108.

[8]. Hillnhuetter, C. and A.-K. Mahlein, Early detection and localization of sugar beet diseases: new approaches, Gesunde Pfianzen 60 (4) (2008), pp. 143–149.

[9]. Jun, W. and S. Wang, (2008). Image thresholding using weighted parzen-window estimation. J. Applied Sci., 8: 772-779. DOI: 10.3923/jas.2008.772.779. URL: http://scialert.net/abstract/?doi=jas.2008.772.779

[10]. MacQueen, J.. (1967). Some methods for classification and analysis of multivariate observations. In L. M. LeCam and J. Neyman, editors, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, volume 1, pages 281--297, Berkeley, CA, 1967. University http://projecteuclid.org/DPubS?service=UI&version=1.0 &verb=Display&handle=euclid.bsmsp/1200512992.

[11]. Martinez, A., (2007). 2007 Georgia Plant Disease Loss Estimates.http://www.caes.uga.edu/Publications/display HTML.cfm?pk_id=7762. Viewed on Saturday, 15, January, 2011.

[12]. Otsu, N. (1979). "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber. 9: 62–66. DOI:10.1109/TSMC.1979.4310076.

[13]. Otsu, N., "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, 1979, pp. 62-66.

[14]. Prasad Babu, M. S. and Srinivasa Rao , B. (2010) Leaves recognition using back-propagation neural network - advice for pest and disease control on crops. Technical report, Department of Computer Science & Systems Engineering, Andhra University, India. Downloaded from www.indiakisan.net on May 2010.

[15]. Sezgin, M. and Sankur, B. (2003). "Survey over image thresholding techniques and quantitative performance evaluation". Journal of Electronic Imaging 13 (1): 146–165. DOI:10.1117/1.1631315.

[16]. Soltanizadeh, H. and B.S. Shahriar, 2008. Feature extraction and classification of objects in the rosette pattern using component analysis and neural network. J. AppliedSci.,8:4088-4096. OI:10.3923/jas.2008.4088.4096. URL:
ttp://scialert.net/abstract/?doi=jas.2008.4088.4096.

[17]. Stone, M. C. (August 2001). "A Survey of Color for Computer Graphics". Course at SIGGRAPH 2001.

[18]. Rumpf, T., A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, L. Plumer, Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance, Computers and Electronics in Agriculture, Volume 74, Issue 1, October 2010, Pages 91-99, ISSN 0168-1699, DOI: 10.1016/j.compag.2010.06.009.

[19]. Wang, X., M. Zhang, J. Zhu and S. Geng, Spectral prediction of Phytophthora infestans infection on tomatoes using artificial neural network (ANN), International Journal of Remote Sensing 29 (6) (2008), pp. 1693–1706.

[20]. Weizheng, S., Yachun, W., Zhanliang, C., and Hongda, W. (2008). Grading Method of Leaf Spot Disease Based on Image Processing. In Proceedings of the 2008 international Conference on Computer Science and Software Engineering - Volume 06 (December 12 - 14, 2008). CSSE. IEEE Computer Society, Washington, DC, 491-494. DOI= http://dx.doi.org/10.1109/CSSE. 2008.1649.

BIOGRAPHIES

Nilesh Bharti received his BE from D.Y. Patil College Of Engineering in 2011. Now he perceiving Master Degree from J.C.O.E. Kuran. SPPune University.