

# ADVANCES IN IMAGE PROCESSING FOR DETECTION OF PLANT DISEASES

**Smita Naikwadi, Niket Amoda**

Department of Electronics and Telecommunication,  
K.C. College of Engineering, Management studies and Research,  
University of Mumbai, India

## ABSTRACT

*We propose and experimentally evaluate a software solution for automatic detection and classification of plant leaf diseases. Studies of plant trait/disease refer to the studies of visually observable patterns of a particular plant. Nowadays crops face many traits/diseases. Damage of the insect is one of the major trait/disease. Insecticides are not always proved efficient because insecticides may be toxic to some kind of birds. It also damages natural animal food chains. The following two steps are added successively after the segmentation phase. In the first step we identify the mostly green colored pixels. Next, these pixels are masked based on specific threshold values that are computed using Otsu's method, then those mostly green pixels are masked. The other additional step is that the pixels with zeros red, green and blue values and the pixels on the boundaries of the infected cluster (object) were completely removed. The experimental results demonstrate that the proposed technique is a robust technique for the detection of plant leaves diseases. The developed algorithm's efficiency can successfully detect and classify the examined diseases with a precision between 83% and 94%, and can achieve 20% speedup over the approach proposed in [1].*

**Index Terms-** Artificial Intelligence, Image Processing

**Keywords-** Plant disease, Trait, Image processing, Color Co-occurrence Method, Neural Network.

## 1. INTRODUCTION

Plant diseases have turned into a dilemma as it can cause significant reduction in both quality and quantity of agricultural products. In India 70% of the population depend on agriculture. Farmers have wide range of diversity to select suitable Fruit and Vegetable crops. However, the cultivation of these crops for optimum yield and quality produce is highly technical. It can be improved by the aid of technological support. The management of perennial fruit crops requires close monitoring especially for the management of diseases that can affect production significantly and subsequently the post-harvest life.

In [2] the authors have worked on the development of methods for the automatic classification of leaf diseases based on high resolution multispectral and stereo images. Leaves of sugar beet are used for evaluating their approach. Sugar beet leaves might be infected by several diseases, such as rusts (*Uromyces betae*), powdery mildew (*Erysiphe betae*).

Disease is caused by pathogen which is any agent causing disease. In most of the cases pests or diseases are seen on the leaves or stems of the plant. Therefore identification of plants, leaves, stems and finding out the pest or diseases, percentage of the pest or disease incidence, symptoms of the pest or disease attack, plays a key role in successful cultivation of crops. It is found that diseases cause heavy crop losses amounting to several billion dollars annually.



**Fig1:** Potato leaf affected by the late blight



**Fig2:** Leaf symptoms of canker on top and bottom of leaf



**Figure3:** Late blight stems lesions

Disease management is a challenging task. Mostly diseases are seen on the leaves or stems of the plant. Precise quantification of these visually observed diseases, pests, traits has not studied yet because of the complexity of visual patterns. Hence there has been increasing demand for more specific and sophisticated image pattern understanding. In biological science, sometimes thousands of images are generated in a single experiment. These images can be required for further studies like classifying lesion, scoring quantitative traits, calculating area eaten by insects, etc. Almost all of these tasks are processed manually or with distinct software packages. It is not only tremendous amount of work but also suffers from two major issues: excessive processing time and subjectiveness rising from different individuals. Hence to conduct high throughput experiments, plant biologist need efficient computer software to automatically extract and analyze significant content. Here image processing plays important role.

## 2. THE PROPOSED APPROACH – STEPBY-STEP DETAILS

The overall concept that is the framework for any vision related algorithm of image classification is almost the same.

Image Acquisition
Image Preprocessing
Image Segmentation
Feature Extraction
Statistical Analysis
Classification Based on Classifier

**Figure 4:** The basic procedure of the proposed imageprocessing- based disease detection solution

First, the digital images are acquired from the environment using a digital camera. Then image-processing techniques are applied to the acquired images to extract useful features that are necessary for further analysis. After that, several analytical discriminating techniques are used to classify the images according to the specific problem at hand. Figure 1 depicts the basic procedure of the proposed vision-based detection algorithm in this research.

The proposed approach step - by - step of the image segmentation and recognition processes is illustrated in Algorithm 1. In the initial step, the RGB images of all the leaf samples were picked up. Some real samples of those diseases are shown in Figure 2. It is obvious from Figure 2 that leaves belonging to early scorch, cottony mold, ashen mold and late scorch have significant differences from greasy spot leaves in terms of color and texture. Also, Figure 3 shows two images; the left image is infected with tiny whiteness disease, and the right image is a normal image. However, the leaves related to these six classes (early scorch, cottony mold, ashen mold, late scorch, tiny whiteness and normal) had very small differences as discernible to the human eye, which may justify the misclassifications based on naked eye.

---

**Algorithm:** Basic steps describing the proposed algorithm.

---

1. RGB image acquisition
2. Create the color transformation structure
3. Convert the color values in RGB to the space specified in the color transformation structure
4. Apply K-means clustering
5. Masking green-pixels
6. Remove the masked cells inside the boundaries of the infected clusters
7. Convert the infected (cluster / clusters) from RGB to HSI Translation
8. SGDM Matrix Generation for H and S
9. Calling the GLCM function to calculate the features
10. Texture Statistics Computation
11. Configuring Neural Networks for Recognition

In details, in step 2 a color transformation structure for the RGB leaf image is created, and then, a device-independent color space transformation for the color transformation structure is applied in step 3. Steps 2 and 3 are inevitable for arraying out step 4. In this step the images at hand are segmented using the K-Means clustering technique [3; 4; 7; 8]. These four steps constitute phase 1 whereas, the infected object (s) is/are determined. In step 5, we identify the mostly

green colored pixels. After that, based on specified and varying threshold value that is computed for these pixels using Otsu's method [5; 6], these mostly green pixels are masked as follows: if the green component of pixel intensities is less than the pre-computed threshold value, the red, green and blue components of the this pixel is assigned to a value of zero. This is done in sense that these pixels have no valuable weight to the disease identification and classification steps, and most probably those pixels represent healthy areas in the leave.

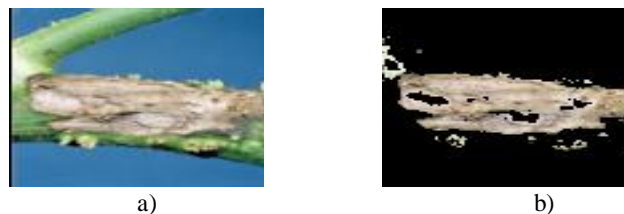
Furthermore, the image processing time should become significantly reduced. In step 6 the pixels with zeros red, green and blue values and the pixels on the boundaries of the infected cluster (object) were completely removed. Steps 5 and 6 form phase 2, and this phase is helpful as it gives more accurate disease classification and identification results with satisfied performance and the overall computation time should become significantly less. The observations behind steps 5 and 6 were experimentally validated.

Next, in step 7 the infected cluster was then converted from RGB format to HSI format. In the next step, the SGDM matrices were then generated for each pixel map of the image for only H and S images. The SGDM is a measure of the probability that a given pixel at one particular gray-level will occur at a distinct distance and orientation angle from another pixel, given that pixel has a second particular gray-level. From the SGDM matrices, the texture statistics for each image were generated. Concisely, the features set were computed only to pixels inside the boundary of the infected areas of the leaf. In other words, healthy areas inside the infected areas were also removed. Steps 7 – 10 form phase 3 in which the texture features for the segmented infected objects in this phase are calculated. Finally, the recognition process in the fourth phase was performed to the extracted features through a pre-trained neural network. For each image in the data set the subsequent steps in Algorithm 1 were repeated.

The Proposed approach for segmentation and classification plant diseases can be divided into three phases:

### **2.1 PHASE 1 - K-MEANS CLUSTERING TECHNIQUE**

There are two 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. In a device independent color space, the coordinates used to specify the color will produce the same color regardless of the device used to draw it. Thus, we created the color transformation structure that 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 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 as the 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 classes. The classification is done by minimizing the sum of squares of distances between the objects and the corresponding cluster or class centroid [3; 4]. However, K-means clustering is used to partition the leaf image into four clusters in which one or more clusters contain the disease in case when the leaf is infected by more than one disease. In our experiments multiple values of number of clusters have been tested. Best results were observed when the number of clusters was 3 or 4. A stem image infected with early scorch and its first cluster (the infected object itself) is shown in Figure 5



**Figure 5:** A stem image infected with early scorch; a) original image b) cluster 1 image

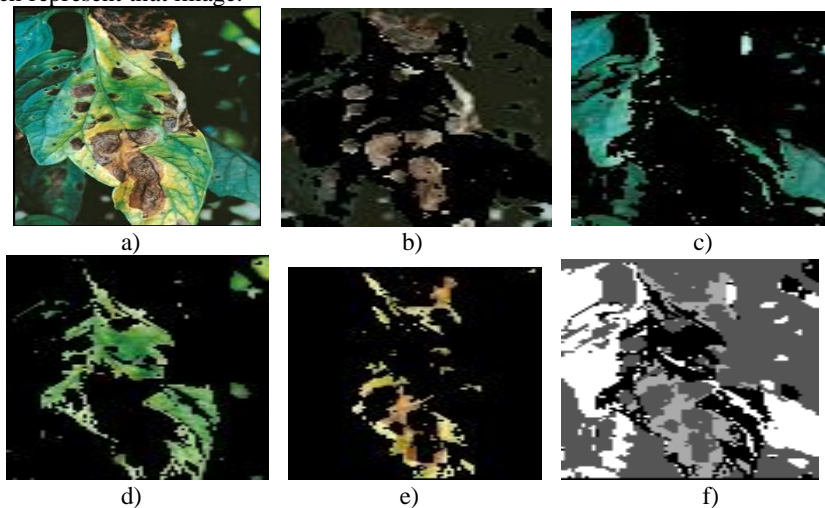
### **2.2 PHASE 2 – MASKING THE GREEN PIXELS AND THE PIXELS ON THE BOUNDARIES**

This phase consists of two steps: The mostly green colored pixels are identified, and then the global image threshold using

Otsu's method [5; 6] has been applied in order to specify the varying threshold value which chooses the threshold to minimize the interclass variance of the threshold black and white pixels. Next, the green pixels are masked as follows: if the green component of pixel intensities is less than the computed 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 zeros components and the pixels on the boundaries of the infected cluster(s).

### 2.3 PHASE 3 – FEATURES EXTRACTION

In the proposed approach, the method adopted 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 and texture of an image are taken into account, to arrive at unique features, which represent that image.



**Figure 6:** An example of the output of K-Means clustering for a leaf that is infected with early scorch disease. (a) The infected leaf picture. (b, c, d, e) the pixels of the first, second, the third and the fourth clusters. (e) a single gray-scale image with the pixels colored based on their cluster index.

#### 2.3.1 CO-OCCURRENCE METHODOLOGY FOR TEXTURE ANALYSIS

The image analysis technique selected for this study was the CCM method. The use of color image features in the visible light spectrum provides additional image characteristic features over the traditional gray-scale representation [9]. The CCM methodology established in this work consists of three major mathematical processes. First, the RGB images of leaves are converted into Hue Saturation Intensity (HSI) color space representation. Once this process is completed, each pixel map is used to generate a color co-occurrence matrix, 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 is based on human color perception [10]. Electromagnetic radiation in the range of wavelengths of about 400 to 700 nanometers is called visible light because the human visual system is sensitive to this range. Hue is generally related to the wavelength of a light and intensity shows the amplitude of a light. Lastly, saturation is a component that measures the “colorfulness” in HSI space [10].

The color co-occurrence texture analysis method was developed through the use of Spatial Gray-level Dependence Matrices (SGDM's) [11]. The gray level co-occurrence methodology is a statistical way to describe shape by statistically sampling the way certain grey-levels occur in relation to other grey-levels. These matrices measure the probability that a pixel at one particular gray level will occur at a distinct distance and orientation from any pixel given that pixel has a second particular gray level. For a position operator  $p$ , we can define a matrix  $P_{ij}$  that counts the number of times a pixel with grey-level  $i$  occurs at position  $p$  from a pixel with grey-level  $j$ . The SGDMs are represented by the function  $P(i, j, d, \Theta)$  where  $I$  represents the gray level of the location  $(x, y)$  in the image  $I(x, y)$ , and  $j$  represents the gray level of the pixel at a distance  $d$  from location  $(x, y)$  at an orientation angle of  $\Theta$ . The reference pixel at image position  $(x, y)$  is shown as an asterisk. All the neighbors from 1 to 8 are numbered in a clockwise direction. Neighbors 1 and 5 are located on the same plane at a distance of 1 and an orientation of 0 degrees. An example image matrix and its SGDM are already given in the three equations above. In this research, a one pixel offset distance and a zero degree orientation angle was used. After the transformation processes, we calculated the feature set for H and S, we dropped (I) since it does not give extra information. However, we use GLCM function in Matlab to create gray-level co-occurrence matrix; the number of gray levels is set to 8, and the symmetric value is set to “true”, and finally, offset is given a “0” value.

## 3. EXPERIMENTAL RESULTS AND OBSERVATIONS

### 3.1 INPUT DATA PREPARATION AND EXPERIMENTAL SETTINGS

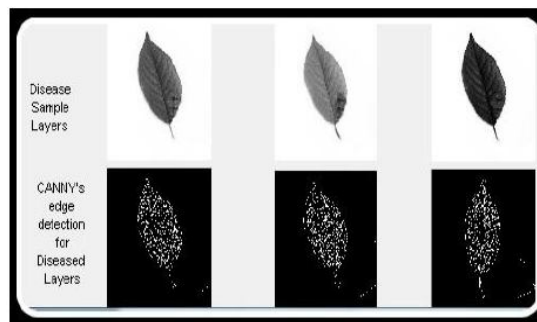
In our experiments, two main files were generated, namely: (i) Training texture feature data, and (ii) Testing texture feature data. The two files had 192 rows each, representing 32 samples from each of the six classes of leaves. Each row

had 10 columns representing the 10 texture features extracted for a particular sample image. Each row had a unique number (1, 2,

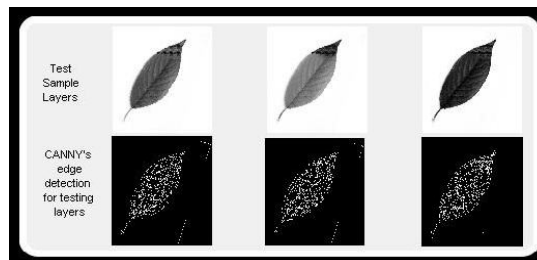
3, 4, 5 or 6) which represented the class (i.e., the disease) of the particular row of data. „1“ represented early scorch disease infected leaf. „2“ represented Cottony mold disease infected leaf. „3“ represented ashen mold disease infected leaf. „4“ represented late scorch disease infected leaf. „5“ represented tiny whiteness disease infected leaf, and „6“ represented normal leaf. Then, a software program was written in MATLAB that would take in .mat files representing the training and testing data, train the classifier using the „train files“, and then use the test file“ to perform the classification task on the test data. Consequently, a Matlab routine would load all the data files (training and testing 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 [5; 6]. 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 early scorch, cottony mold, ashen mold, late scorch, 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  $10^{-5}$ .

There are two main characteristics of plant disease detection using machine-learning methods that must be achieved, they are: speed and accuracy. Hence there is a scope for working on development of innovative, efficient & fast interpreting algorithms which will help plant scientist in detecting disease.

Work can be done for automatically estimating the severity of the detected disease.

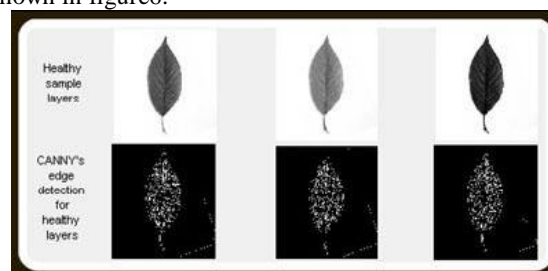


**Figure6: a)** Experiment result of diseased sample



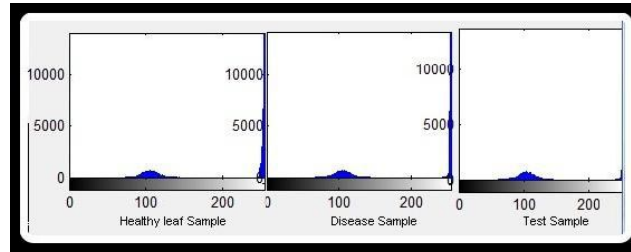
**Figure6: b)** Experiment result of testing sample

The experimentation start with the two samples, one for the healthy leaf sample and second for the diseased leaf sample. The experiment results for the phase-1 which samples are the input to the MATLAB. The training process is started on both the samples. As shown in figure6.

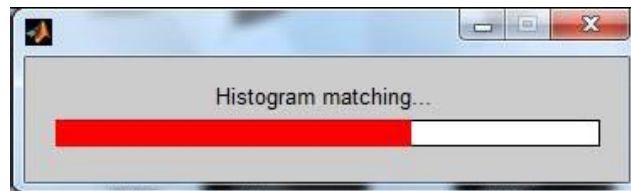


**Figure6:c)** Experiment result of healthy sample





**Figure7:** Histogram for samples

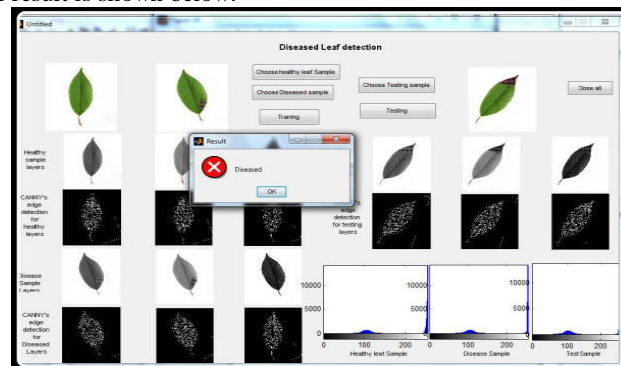


**Figure8:** Histogram matching waiting bar



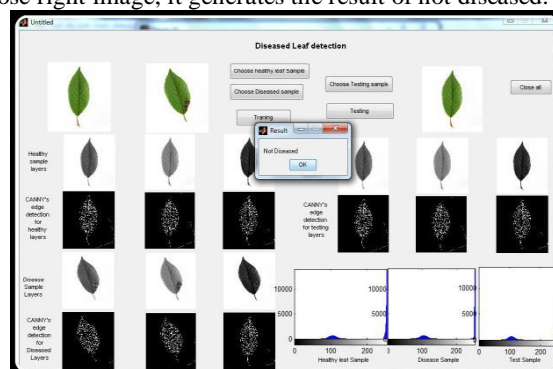
**Figure9:** Shown the result of diseased leaf

The diseased plant full screen result is shown below:



**Figure10:** Full view of GUI of diseased leaf detection

On the other hand, when we choose right image, it generates the result of not diseased. They are shown as below:

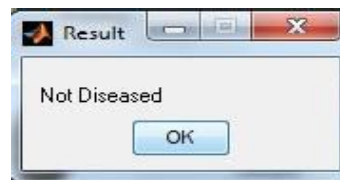


**Figure11:** Full screen result of not diseased image

On the other hand, when we apply this technique on the diseased sample, we obtain the following results that are shown as in figure6a. The training process is necessary for further analysis. If we are going to further analysis without including the training process, it passes a message to us —First we required training process! in the MATLAB command window. Once the first phase is finished the implementation of second phase is started. The second phase start with the implementation of testing image. The implementation result for testing sample is shown in figure 6b. the experiment result on testing sample which conveys the layers separation result and the edge detection on each layer.

When training process for both samples and testing sample are completed. The histogram is generated for healthy leaf sample, diseased leaf sample and for testing sample. Once the histograms for samples are generated the comparison between histogram is started, immediately we applied the comparison technique based on the histogram and edge detection technique. The generation of histogram and the comparison is shown below. The above figure shows the histogram for healthy leaf sample, diseased sample and test sample. The comparison between these histograms is shown by the waiting bar in the MATLAB. The waiting is shown in figure8.

First comparison is made between the test sample and the healthy sample. If the test sample histogram matched with the healthy leaf sample is generated the result that is plant is not diseased. If it does not match with the healthy leaf sample, the comparison is made between the test sample and the diseased sample. If the test sample is matched with the diseased sample, it generates plant is diseased. The healthy and diseased image is shown by the figure9 and figure 12.



**Figure12:** shows the result of not diseased leaf

#### **4. CONCLUSION AND FUTURE WORK**

To wind up all the information discuss above, I should like to concludes that it is a efficient and accurate technique for automatically detection of plant diseased. In this research, plant diseased is detected by using histogram matching. The histogram matching is based on the color feature and the edge detection technique. The color features extraction are applied on samples that are contained the healthy leaf of plant and the diseased leaf of the plant. The training process includes the training of these samples by using layers separation technique which separate the layers of RGB image into red, green, and blue layers and edge detection technique which detecting edges of the layered images.

Once the histograms are generated for both samples and the testing image, immediately we applied the comparison technique based on the histogram. The comparison is firstly with the testing sample and the healthy sample if the testing sample is diseased, it compare testing sample with the diseased sample and these steps take few minute to display the comparison result that is the testing sample is diseased or not. The GUI (graphical user interface) is used to show the overall process. When the comparison is applied the waiting bar is display on our display and results are also shown through the GUI. This is beneficial for us because we are easily understood the processing of implementation phase.

The future work mainly concerns with the large database and advance feature of color extraction that contains a better result of detection. Another work concerns with research work in a particular field with advance features and technology.

#### **REFERENCES**

- [1] Al-Bashish, D., M. Braik and S. Bani-Ahmad, 2011. Detection and classification of leaf diseases using Kmeans-based segmentation and neural-networks-based classification. Inform. Technol. J., 10: 267-275. DOI: 10.3923/itj.2011.267.275
- [2] 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](http://www.precision-crop-protection.unibonn.de/gk_research/project.php?project=3_09)
- [3] 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 of California Press. URL: <http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.bsm/1200512992>.

- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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
- [8] 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>
- [9] 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
- [10] Stone, M. C. (August 2001). "A Survey of Color for Computer Graphics". Course at SIGGRAPH 2001..
- [11] 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>.

## **AUTHOR**



Niket Amoda received his B.E. in Electronics & Communication Engineering from Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal, India, in 2006 and . He completed his Diploma in Advance Computing from ACTS, Pune, India in 2008 and M.E. in Electronics & Telecommunication Engineering from Vivekanand Education Society's Institute of Technology (VESIT), University of Mumbai, India, in 2013.

He worked as a Software Development Engineer in Ford Motor Company, India from 2008 to 2009 and is currently working as a Assistant Professor in Electronics & Telecommunication Engineering Department of KC College of Engineering & Management & Research, Thane, India, since 2010 till date.

He has many publications in international journals and international conferences. His research area is Image processing.



Smeeta Naikwadi received her diploma in Industrial Electronics from Bharti Vidyapeeths' college of Engineering, Kharghar, India in 2010. She is currently pursuing degree in department of Electronics and Telecommunication from KC College of Engineering & Management & Research, Thane, India.

Her areas of interest are Embedded and Telecommunication.