Prediction of stress phenotypes in tea plants using Deep Learning

[1]Kochumol Abraham, [2]Dr. J P K Meenakshisundaram,
[3]Win Mathew John, [4]Sijo Thomas

[1]Assistant Professor, [2] Assistant Professor, [3] Assistant Professor
[4]Assistant Professor

[1]P G Department of Computer Applications,
[1]Marian College, Kuttikanam, India

Abstract: Most people in India are farmers. They depend on agriculture. They have aim to increase productivity and quality of product. Vegetables and fruits are the most important agricultural products for from customer view. The economical profit depends on a product quality which is depends on a quality of a soil, seeds and fertilizers. So for increasing the profit farmer mainly focuses on these three main things. Instead there is one more thing which effect on a production that is diseases. To increase profit we have to control these diseases. But it is necessary to detect and control such diseases in a specific period which is at their initial state. These diseases occur due to the pathogens such as fungi, bacteria and viruses, and due to adverse environmental conditions. Therefore, it is necessary to diagnosis a plant disease. For that farmers requires continuous monitor the plant body which is time consuming process. It also the very expensive process for the farmers. So the latest develop method give us machine view for detecting plant diseases which is much accurate and less time consuming. Computer vision and machine learning have the capability of resolving this issue and enabling accurate, scalable high-throughput phenotyping. Among machine learning approaches, deep learning has emerged as one of the most effective techniques in various fields of modern science.

The proposed system is used to record environmental conditions, pest's detection and identify the factors which lead to changes in tea leaves. Extracting patterns and features from this large corpus of data requires the use of machine learning (ML) tools to enable data assimilation and feature identification for stress phenotyping. The primary focus is on providing real-time data acquisition, analysis and monitoring solutions for plant breeders and researchers to boost the productivity. For small scale farmers, early identification of disease is very much possible and able to control the insects by organic pesticides or by the use of minimal amount of chemical pesticides. For large scale farmers frequent monitoring and early identification of disease is not possible and it results in a severe outbreak of the disease and pest growth which cannot be controlled by organic means. In this situation farmers are forced to use the poisonous chemicals to eradicate the disease in order to retain the crop yield. This problem can be solved by automating the monitoring process by use of advanced image processing techniques.

***IndexTerms* - Machine Learning, Segmentation, Deep Learning.**

I. Introduction

Tea is the one of the most widely consumed drinks in the World because of its health-giving, dietetic and even therapeutic qualities. There are numerous beneficial components in green tea, such as vitamins, amino acids, and many inorganic nutritional components. The tea growing environment in Kerala is conducive to a large number of pests and diseases. Leaf disease is impairment to the normal state of the plant that modifies or interrupts its vital functions such as photosynthesis, transpiration, pollination, fertilization, germination etc. The change in the leaf color is the important aspect for the notification. When the physical condition of the tea plant is at a good stage then the color of the leaf will be different but as soon as the leaf is affected by some harming pathogens, the color changes automatically. The emergence of plant diseases has become more common now a days as factors such as climate and environmental conditions are more unsettled than ever. Tea plant diseases are usually caused by many stresses (biotic and abiotic). There are numerous characteristics and behaviors of such plant diseases in which many of them are merely distinguishable. So the ability of tea leaf disease recognition in an earlier stage is an important task. Hence an intelligent system for recognition, prevention and control of tea leaf diseases is needed.

II. RELATED WORK

2.1 Image pre-processing

Plant leaf disease detection and diagnosis is a scientific method. The photographic images are used to implement in the leaf disease detection process. The photographic digital images are transferred into a particular form [32]. Image pre-processing is a method, used to transfer the original images into another form. In plant leaves disease detection, captured photographic images are used. There will be noises in the images, the regions of interest in the image is not clear or other interference appears in the image [33]. The image preprocessing is used to get clear, noiseless enhanced leaves images. This enhanced images are used to leaves diseases detection and analysis process. Various types of images are used in image pre-processing. The selection of image type differs based on the processing area, implementation of mathematical calculation and application. Generally, plant leaves image color and texture are a unique feature, which are used to detect and analyze the diseases and their level [33].

2.2 Segmentation

Wu et al. [17] proposed a Probabilistic Neural Network for leaf recognition using 12 commonly used Digital Morphological Features (DMFs), derived from 5 basic features (diameter, physiological length, physiological width, leaf area, leaf perimeter). The authors collected a publicly available database of plant leaves called Flavia, containing 1907 images of leaves from 32 species. The average accuracy on the current version of the dataset is 93%. Kadir et al. [24] prepared the Foliage dataset, consisting of 60 classes of leaves, each containing 120 images. Results on the Foliage dataset are compared.

The best reported result by Kadir et al. [18] was achieved by a combination of shape, vein, texture and colour features processed by Principal Component Analysis before classification by a Probabilistic Neural Network. S¨oderkvist [25] proposed a visual classification system of leaves and collected the so called Swedish dataset containing scanned images of 15 classes of Swedish trees. Wu et al. [26] introduced a visual descriptor for scene categorization called the spatial Principal component Analysis of Census Transform (spatial PACT), achieving a 97.9% recognition rate on the Swedish dataset. Qi et al. achieve2 99.38% accuracy on the Swedish dataset using a texture descriptor called Pairwise Rotation Invariant Co-occurrence Local Binary Patterns (PRI-CoLBP) [27] with SVM classification. Novotn´y and Suk [13] proposed a leaf recognition system, using Fourier descriptors of the leaf contour normalised to translation, rotation, scaling and starting point of the boundary. The leaf recognition method by Fiel and Sablatnig [12] is based on a Bag of Words model with SIFT descriptors and achieves 93.6% accuracy on a leaf dataset of 5 Austrian tree species. Kadir et al. compare several shape methods on plant recognition [15]. Of the compared methods - geometric features, moment invariants, Zernike moments and Polar Fourier. Transform - Polar Fourier Transform performed best achieving 64% accuracy on a database of 52 plant species. The dataset has not been published.

Kumar et al. [13] describe Leafsnap, a computer vision system for automatic plant species identification, which has evolved from the earlier plant identification systems by Agarwal et al. [19] and Belhumeur et al. [10]. Compared to the earlier versions, they introduced a pre-filter on input images, numerous speed-ups and additional post-processing within the segmentation algorithm, the use of a simpler and more efficient curvature-based recognition algorithm instead of Inner Distance Shape Context (IDSC); a larger dataset of images, and a new interactive system for use by non-expert users. Kumar et al. [13] introduced the Leafsnap database of 184 tree species. On this database, 96.8% of queries have a species match within the top 5 results shown to the user with the used method. The resulting electronic field guide, developed at Columbia University, the University of Maryland, and the Smithsonian Institution, is available as a free mobile app for iOS devices. Although the app runs on iPhone and iPad devices, the leaf images are processed on a server, internet connection is thus required for recognition, which might cause problems in natural areas with slow or no data connection. Another limit is the need to take the photos of the leaves on a white background. The authors also collected a new large leaf dataset called Middle European Woods (MEW) containing 153 classes of native or frequently cultivated trees and shrubs in Central Europe. Their method achieves 84.92% accuracy when the dataset is split into equally sized training and test set. Section 5.3 contains the comparison to our results. One possible application of leaf description is the identification of a disease. Pydipati et al. [30] proposed a system for citrus disease identification using Color Co-occurrence Method (CCM), achieving accuracies of over 95% for 4 classes (normal leaf samples and samples with a greasy spot, melanose, and scab). Kim et al. [28] proposed a tree classification method using a combination of leaf, flower and bark photos of the same tree. The description consists of 20 features of wavelet decomposition with 3 levels for a grey and a binary image for description of bark, 32 features of Fourier descriptor for leaves and 72 features in the HS colour space for flowers. The results were obtained on an unpublished dataset consisting of 16 classes. Recognition accuracy of 31%, 75% and 75% is reported for individual leaf, flower and bark classification and 84%, 75% and 100% accuracy for combinations of leaf+flower, leaf+bark and bark+flower. However, in all cases only a single image per class was tested.

The statistical significance of such result is questionable and may be prone to overfitting and unreliable. PlantNet3 [14] is an interactive plant identification and collaborative information system providing an image sharing and retrieval application for plant identification. It has been developed by scientists from four French research organizations (Cirad, INRA, INRIA and IRD) and the Tela Botanica network. The Pl@ntNet-identify Tree Database provides identification by combining information from images of the habitat, flower, fruit, leaf and bark. The exact algorithms used in the PlantNet-identify web service4 and their accuracies are not publicly documented.

2.3 Classification

Several approaches have been introduced to classify a leaf [1], such as k-Nearest Neighbor Classifier (k-NN), Probabilistic Neural Network (PNN), Genetic Algorithm (GA), Support Vector Machine (SVM), and Principal Component Analysis (PCA). Most of researchers used green color leaves or ignored color information on leaves.For example, Zulkifli [2] proposed General Regression Neural Network to classify 10 kinds of plants with green color leaves. Wu et al. [3] used PNN to classify 32 kinds of green leaves. They shared dataset called Flavia. Several researchers used the dataset to test their classifiers. For example, Singh et al. [4] suggested SVM to implement a classifier that was reported could improve the accuracy, Shabanzade et al. [5] used Linear Discriminant Analysis (LDA) to test part of Flavia. Actually, color as features in leaf identification system has been introduced by Man et al. [6]. They used the first order, the second order and the third of color moments in HSV color space. They claimed that the system can recognize 24 categories of plants with the average accuracy up to 92.2%. The color information also have been inserted into plant retrieval system by Kebabci et al. [7]. Several leaf classification systems have incorporated texture features to improve the performance, such as in [6] that used entropy, homogeneity and contraction derived from co-occurrence matrix came from Digital Wavelet Transform (DWT), in [8] that used lacunarity to capture texture of leaf and in [9] that used GLCM. Recognition of leaves usually refers only to recognition of broad leaves, needles are treated separately. Several techniques have been proposed for leaf description, often based on combining features of different character (shape features, colour features, etc.).

III. PROPOSED SYSTEM

The proposed system aims at capturing images from a tea estate using high throughput phenotyping and storing the data in a server for processing. Both ground and aerial devices can be used to capture images. These devices equipped with multiple sensors can be used to measure the extracting patterns and features from a large mass of data using machine learning (ML) tools which enables data assimilation and feature identification for stress phenotyping.

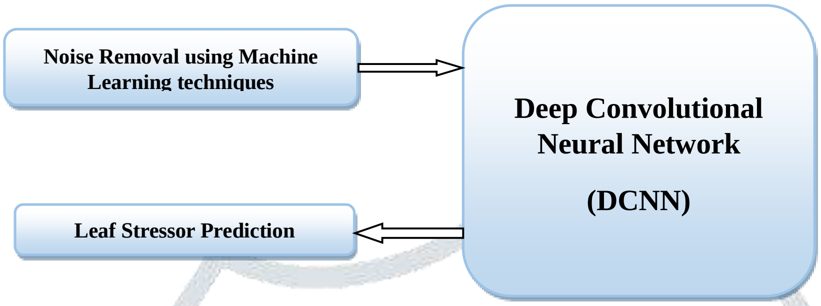

FIGURE 1. PHASES OF PLANT STRESSOR PREDICTION

3.1 Image Capturing

Autonomous platforms such as unmanned aerial vehicles (UAVs) and ground robots equipped with multiple sensors can take pictures in near real-time of the entire experimental plot several times per day, or over the entire season from germination to maturity, resulting in massive amounts of data for analysis and storage. Making sense of all these collected data can be done effectively using ML tools. It is important to emphasize that for ML tools data can be collected from complex, integrated imaging platforms or from simple(r) methods such as crowd-sourced cell phone images

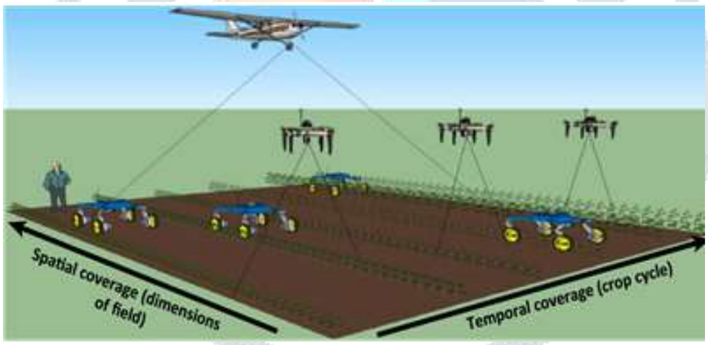

Figure 2. Image Capturing using UAV devices

3.2 DCNN Architecture

Image Classification using CNN takes an input image, process it and classify it into healthy or unhealthy leaf depending on the image resolution. Each input image will pass through a series of convolution layers with filters (Kernels), Pooling, fully connected layers (FC) and Softmax function to classify an object with probabilistic values between 0 and 1. Convolution is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel. The output of convolution is a "Feature Map". Convolution of an image with different filters can perform operations such as edge detection, blur and sharpen by applying filters.

Rectified Linear Unit (ReLU)

ReLU stands for Rectified Linear Unit for a non-linear operation. The output is $f(x) = \max(0,x)$. ReLU's purpose is to introduce non-linearity in the Convolution Neural Network. Since, the real world data would want our Convolution Neural Network to learn would be non-negative linear values.

Pooling Layer

Pooling layers section would reduce the number of parameters when the images are too large. Spatial pooling also called subsampling or down sampling which reduces the dimensionality of each map but retains the important information. Max pooling take the largest element from the rectified feature map. Taking the largest element could also take the average pooling. Sum of all elements in the feature map call as sum pooling.

Fully Connected Layer

The layer we call as FC layer, we flattened our matrix into vector and feed it into a fully connected layer like neural network.

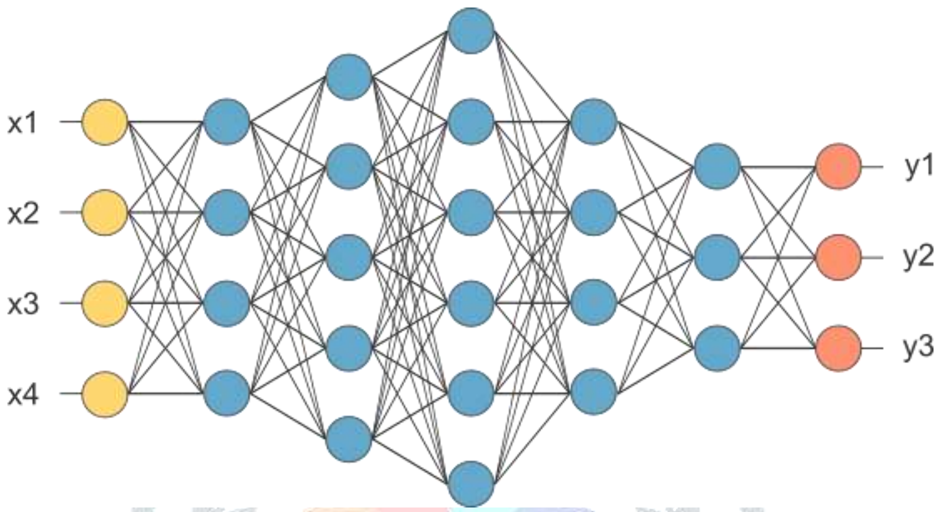

Figure 3. Fully Connected Neural Network

In the above diagram, feature map matrix will be converted as vector (x1, x2, x3, …). With the fully connected layers, we combined these features together to create a model. Finally, we have an activation function such as softmax or sigmoid to classify the outputs as healthy leaf or unhealthy leaf.

3.3. Leaf Stressor Prediction

Through the process of segmentation leaf pixels are extracted from background pixels. This process consists of finding the boundary pixels and obtaining the contiguous segments from boundary estimates. Using a deep convolutional neural network, each pixel is either a boundary or an interior pixel.

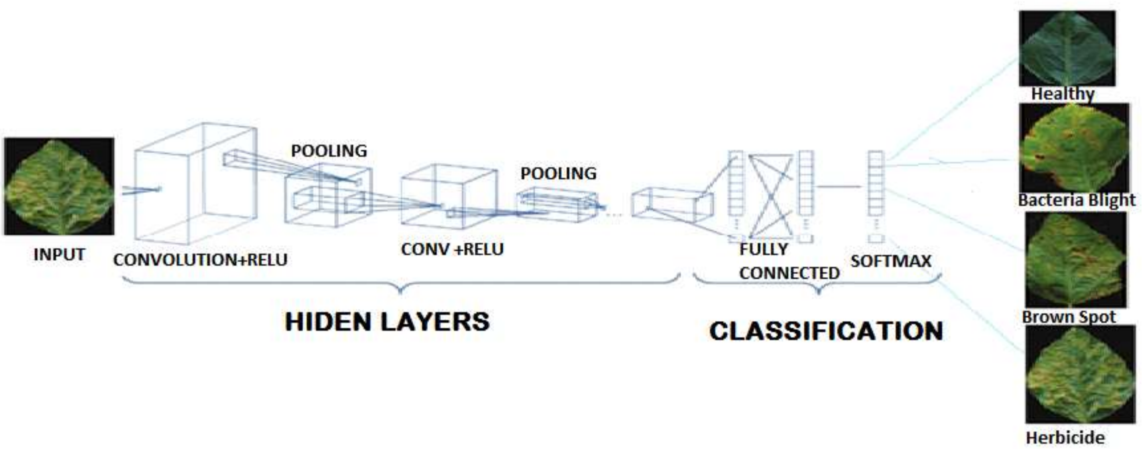

Figure 4. Prediction of Stress Phenotypes in tea plants

The pixel value variation determines whether it is stress affected or not. A change in the interior pixel value determines the type of stress in tea plants. Stress can be due to biotic or abiotic factors. Depending on the factor the type of stress phenotypes varies. A detailed study on the same is required to classify the different types of biotic and abiotic factors in tea plants.

IV.CONCLUSION

The enormous volume, variety, velocity, and veracity of imaging and remote-sensing data generated by real-time platforms represent a 'big data' problem. The data generated by such real-time platforms must be efficiently archived and retrieved for analysis. Although the analysis and interpretation of such (image-based) big data are challenging, the ensuing possibilities that can impact on agricultural production make it a promising approach for High Throughput Phenotyping. ML approaches present a scalable, modular strategy for data analysis, especially for the new application domain of 'plant stress analytics'. Recent studies on High Throughput Stress Phenotyping using images obtained from UAV-based platforms to detect diseases in tea plants using ML algorithms have paved a new path for better stress management practices on spatial and temporal basis.

REFERENCES

[1] M. Kumar, M. Kamble, S. Pawar, P. Patil and N. Bonde. 2011. Survey on Techniques for Plant Leaf Classification. International Journal of Modern Engineering Research.1 (2): 538- 544.

[2] Z. Zulkifli. 2009. Plant Leaf Identification using Moment Invariants & General Regression Neural Network", Master Thesis, University Teknologi Malaysia.

[3] G. Wu, F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang and Q. L. Xiang.2010. A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network, IEEE 7th International Symposium on Signal Processing and Information Technology.

[4] K. Singh, I. Gupta and S. Gupta.2010. SVM-BDT PNN and Fourier Moment Technique for Classification of Leaf Shape, International Journal of Signal Processing, Image Processing and Pattern Recognition.3 (4):67-78

[5] M. Shabanzade, M. Zahedi and S. A. Aghvami.2011. Combination of Local Descriptors and Global Features for Leaf Classification, Signal Image Processing: An International Journal. 2(3): 23-31.

[6] Q. K. Man, C. H. Zheng, X. F. Wang and F. Y. Lin. 2008.Recognition of Plant Leaves using Support Vector. International Conference on Intelligent Computing, Shanghai. pp. 192-199.

[7] H. Kebabci, B. Yanikoglu and G. Unal.2010. Plant Image Retrieval using Color, Shape, & Texture Features. The Computer Journal Advance Access. 53(1).

[8] A. Kadir, L.E. Nugroho, A. Susanto and P. I. Santosa. 2011. Leaf Classification using Shape, Color, and Texture Features, International Journal of Computer Trends and Technology.1(3): 225-230.

[9] A. Kadir, L.E. Nugroho, A. Susanto and P. I. Santosa.2011.Neural Network Application on Foliage Plant Identification, International Journal of Computer Applications. 29(9): 15-22.

[10] Belhumeur, P.N., Chen, D., Feiner, S., Jacobs, D.W., Kress, W.J., Ling, H., Lopez, I., Ramamoorthi, R., Sheorey, S., White, S., Zhang, L.2008: Searching the world's herbaria: A system for visual identification of plant species. In: Computer Vision–. Springer: 116–129

[11] Nilsback, M.E., Zisserman, A. 2009. An Automatic Visual Flora: Segmentation and Classification of Flower Images. PhD thesis, Oxford University.

[12] Fiel, S., Sablatnig, R. 2011.Automated identification of tree species from images of the bark, leaves and needles. In: Proc. of 16th Computer Vision Winter Workshop, Mitterberg, Austria: 1–6.

[13] Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V.2012. Leafsnap: A computer vision system for automatic plant species identification. Computer Vision–ECCV. Springer: 502–516.

[14] Barth´el´emy, D., Boujemaa, N., Mathieu, D., Molino, J.F., Bonnet, P., Enficiaud, R., Mouysset, E., Couteron, P.2009. The pl@ntnet project: A computational plant identification and collaborative information system, Tech. Rep., XIII World Forestry Congress.

[15] Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I.2011. A comparative experiment of several shape methods in recognizing plants. International Journal of Computer Science & Information Technology 3(3).

[16] Agarwal, G., Belhumeur, P., Feiner, S., Jacobs, D., Kress, W.J., Ramamoorthi, R., Bourg, N.A., Dixit, N., Ling, H., Mahajan, D., et al. 2006. First steps toward an electronic field guide for plants. Taxon 55(3): 597–610.

[17] Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F., Xiang, Q.L.2007. A leaf recognition algorithm for plant classification using probabilistic neural network. In: Signal Processing and Information Technology, IEEE International Symposium: 11–16

[18] Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I. 2012. Performance improvement of leaf identification system using principal component analysis. International Journal of Advanced Science & Technology: 44.

[19] Kadir, A., Nugroho, L.E., Santosa, P.I. 2012. Experiments of zernike moments for leaf identification 1(17).

[20] Lee, K.B., Hong, K.S.: Advanced leaf recognition based on leaf contour and centroid for plant classification.2012. International Conference on Information Science and Technology: 133–135.

[21] Lee, K.B., Chung, K.W., Hong, K.S.2013. An implementation of leaf recognition system.

[22] Novotn´y, P., Suk, T. 2013. Leaf recognition of woody species in central Europe. Bio systems Engineering. 115(4) : 444–452

[23] Karuna, G., Sujatha, B., GIET, R., Reddy, P.C. An efficient representation of shape for object recognition and classification using circular shift method.

[24] Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I.2011. Neural network application on foliage plant identification. International Journal of Computer Applications.

[25] S¨oderkvist, O.2011. Computer vision classification of leaves from Swedish trees.

[26] Wu, J., Rehg, J.M.2011. Centrist: A visual descriptor for scene categorization. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33(8): 1489–1501.

[27] Qi, X., Xiao, R., Guo, J., Zhang, L.2012. Pairwise rotation invariant co-occurrence local binary pattern. Computer Vision–ECCV. Springer: 158–171.

[28] Kim, S.J., Kim, B.W., Kim, D.P.2011 Tree recognition for landscape using by combination of features of its leaf, flower and bark. In: SICE Annual Conference Proceedings.

[29] Sulc, M., Matas, J.2013. Kernel-mapped histograms of multi-scale lbps for tree bark recognition. Image and Vision Computing New Zealand (IVCNZ), 28th International Conference Proceedings : 82–87.

[30] Pydipati, R., Burks, T. Lee, W.2006. Identification of citrus disease using color texture features and discriminant analysis. Computers and electronics in agriculture 52(1): 49–59.

[31] Arivazhagan S, Shebiah RN, Ananthi S, Varthini SV. 2013. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int: CIGR Journal. 15(1):211–17.

[32] Anami BS, Pujari JD, Yakkundimath R. 2011. Identification and classification of normal and affected agriculture/horticulture produce based on combined color and texture feature extraction. International Journal of Computer Applications in Engineering Sciences. 1(3):356–60. ISSN: 2231-4946.

[33] Badnakhe MR, Deshmukh PR. 2011. An Application of K-Means Clustering and Artificial Intelligence in Pattern Recognition for Crop Diseases. International Conference on Advancements in Information Technology with workshop of ICBMG, IPCSIT IACSIT Press, Singapore. 20:134–8