2018 IEEE 14th International Colloquium on Signal Processing & its Applications (CSPA 2018), 9 -10 March 2018, Penang, Malaysia

Recognition and Detection of Tea Leaf's Diseases Using Support Vector Machine

Md. Selim Hossain[1], Rokeya Mumtahana Mou[1], Mohammed Mahedi Hasan[2], Sajib Chakraborty[1], M. Abdur Razzak[1,*]
[1]Department of Electrical and Electronic Engineering, Independent University, Bangladesh, Dhaka, Bangladesh
[2]Department of Electrical and Electronic Engineering, Southeast University, Dhaka, Bangladesh
[*]E-mail: razzak@iub.edu.bd

Abstract—**Tea is a popular beverage all around the world, and in Bangladesh the cultivation of tea plays a vital role. Many diseases affect the proper growth of tea leaves leading to its reduction,thus hindering of the production of tea. However, if the disease is identified at an early age it would solve all the above mentioned problems through the application of appropriate treatment, or through the pruning of the diseased leaves to prevent further spread of the disease. To solve this problem image processing is the best option to detect and diagnose the disease. The main goal of this research is to develop an image processing system that can identify and classify the two most widespread tea leaf diseases in Bangladesh, namely brown blight disease and the algal leaf disease, from a healthy leaf. Disease identification is the first step; there are many methods that have been used for identifying the leaf disease. In this paper, Support Vector Machine classifier (SVM) is used to recognize the diseases. Eleven features are analyzed during the classification. These features are then used to find the most suitable match for the disease (or normality) every time an image is uploaded into the SVM database. When a new picture is uploaded into the system the most suitable match is found and the disease is recognized. The approach is novel since the number of features compared by the SVM classifier is reduced by three features compared to previous researches, without adversely sacrificing the success rate of the classifier, which retains an accuracy of more than 90%. This also speeds up the identification process, with each leaf image taking 300ms less processing time compared to previous research using SVM, thus ensuring a greater number of leaves can be processed in a given time frame. The proposed solution increases in efficiency of the detection, identification, and classification process will enable the tea industry in Bangladesh to become more competitive globally, by reducing the losses suffered due to diseases of the leaf, and thus increasing the overall tea production rate.**

Keywords—Image processing, Disease detection, Disease recognition, Feature Extraction, Support Vector Machine

I. INTRODUCTION

Tea is one of the essential beverages in Bangladesh. Most of the Bangladeshi people start their day with a cup of tea. Bangladesh has become an important tea producing country. Today the country has 172 commercial tea estates [1]. The districts that produce tea are Maulvibazar, Habiganj, Sylhet, Chittagong, Panchagarh, Brahmanbaria, and Rangamait [2]. Almost the entirety of the district of Sylhet is the standard tea garden area and Srimangal is known as the tea capital of Bangladesh [3]. Tea is the second largest export based cash crop of Bangladesh. The industry accounts for 1% of the national GDP of Bangladesh [4]. Tea production in Bangladesh is greatly hindered due to a number of pests and diseases, caused by a variety of insects, mites, nematodes, bacteria, algae, fungi, weeds, and other diseases which are caused due to the environmental condition of that particular region [2].

Bangladesh is an agricultural country where more than 75% population rely on agriculture directly or indirectly [4]. Approximately 20% to 30% of the tea leaves are lost due to various diseases each year [5]. Farmers in the field judge the identification of tea leaf diseases with their naked eye and previous experience. Many a times, experts are needs to be called in to analyze the tea leaves when there is ambiguity in detecting the diseases by local farmers; this process is not only time consuming, but also costly.

It is important to catch the spread of the disease in its early stages before they reach epidemic proportions; otherwise the disease can spread quickly throughout the entire plantation, resulting in huge losses for the farmers. To aid the farmers in the crucial task of identifying tea leaf diseases in their infancy, it is practical to have an intelligent system of detection, identification, and classification system in place as a preventative measure. The first sign that something is wrong with the leaf is usually indicated by a change in color from a healthy dark green hue. When the tea leaf is healthy the color is distinct, but when the leaf is affected by disease, the color of the leaf changes drastically. Each disease usually has a distinguishable leaf color and texture as symptoms.

The latest trends of research in agriculture are toward the use of gene technology to develop disease resistant variant of the plant, and to increase food quality and productivity of the plant with reduced expenditure [6]. Numerous technological improvements are responsible for the progress in crop management techniques in recent times; including advances in information technology, remote sensing technology, and image processing and pattern recognition [6-7].Therefore, now it is possible to develop and deploy an autonomous system for detection, identification, and classification of diseases in crops in very large fields with minimal manual input. A search through recent literature have identified research in various types of crop diseases including diseases in rice, citrus, Betel vine, and wheat leaf to name a few [8]. However, research into diseases of tea leaves is one area that has not yet seen any significant efforts. Therefore, there should be a way to develop tea leaf disease recognition and detection to help the tea industry in Bangladesh. In this paper, Support Vector Machine classifier (SVM) is used to recognize the diseases of tea leafs.

II. Literature Review

Many types of diseases of the leaf have been investigated including disease of the rice leaf, citrus leaf, wheat leaf, and Betel vine plant.Various papers describing the methods suggesting ways to implement the detection of diseases will be discussed here.

Kholis Majid, *et al.* [7], has added to a portable application for paddy plant malady identification framework utilizing fuzzy entropy and Probabilistic neural system classifier that keeps running on Android Versa Tile's framework. It includes the identification for all sorts of maladies, in particular brown spot, leaf blast, tungro and bacterial leaf blight. The exactness of paddy sicknesses distinguishing proof is 91.46 percent.

Qing Yao *et al.* [9] has proposed segmentation of rice disease spots, and extracting the shape and texture features from these segments. Then SVM method was employed to classify rice bacterial leaf blight, rice sheath blight, and rice blast. The results showed that SVM could effectively detect and classify these disease spots to an accuracy of 97.2%.

Elham Omrani *et al.* [10], used Support Vector Regression (SVR) based on radial basis functions to identify and classify diseases of the apple tree. It is a three step process. First, the captured images of the leaves had to be changed into a device independent color space, such as CIELAB, from a device depended format such as Red-Green-Blue (RGB) color space. Then, the image was segmented to extract the infected area from the overall leaf image. The segmentation technique employed was a region-based one using K-means clustering, wavelet, and grey-level co-occurrence matrix. This features extracted using this type of segmentation are the color, shape, and texture. These types of segmentation techniques are normally used for region description. Finally, the segmented image is classified using SVR.

Phadikar *et al.* [11], used SVM to identify and classify diseases in rice crop, such as leaf blight, sheath blight, and rice blast. The SVM classifier is used to extract features based on shape and texture. In addition to the SVM classifier, they have also proposed using pattern recognition techniques for identifying rice disease based on various infected pictures of rice plants. Digital cameras were used to capture the images of various types of infections in rice plants. Then segmentation techniques were used to detect and separate the infected part of the plant from the overall image, and finally SVM was used to classify the infection.

Haiguang Wang *et al.* [12], used Principal Component Analysis (PCA) and neural networks for disease identification and recognition in wheat crop and grape plants. PCA was used to extract twenty one color, four shape, and twenty five texture features from the plants. After the feature extraction process, various types of neural networks were used for identification and classification of the diseases including: BackPropagation (BP-NN), Radial Basis Function (RBF-NN), Generalized Regression (GR-NN), and Probabilistic Neural Networks (PNN).The results from the different neural networks were compared against each other.

Dheeb Al Bashish *et al.* [13] developed a framework for detecting and classifying leaf and stem diseases of plants. In their paper, they first converted the color space from RGB to Hue-Saturation-Intensity (HSI) format, then used K-means clustering technique to segment the images and extract the color and texture based features, and finally used a statistics based neural network classifier to identify the diseases.

V. A. Gulhane *et al.* [14] also used a self-organizing feature map together with BP-NN to extract color and spot features from the leaf images. This technique was used detect and diagnose diseases in cotton leaves.

P. Revathi *et al.* [15] developed a rather complex algorithm to detect cotton leaf spot disease. The algorithms employs two different methods of feature selection, one being the Particle Swarm Optimization (PSO) method using skew divergence for edge extraction, and the other being the Genetic Algorithm (GA) method based on color and texture variances. Both methods are employed within the Cyan-Magenta-Yellow-Black (CMYK) color space. The extracted features are then input into a SVM BP-NN for classification.

Finally, S. Arivazhagan *et al.* [16] developed an algorithm for detecting unhealthy region of plant leaves and identifying the disease using texture features.The algorithm has three essential steps. First, a color transformation structure is created for the image, which is initially in the RGB color space format. Then, the green pixels are masked inside the structure, after which shape and texture features are extracted. Finally, SVM classifier using the Minimum Distance Criterion is used to identify the disease.

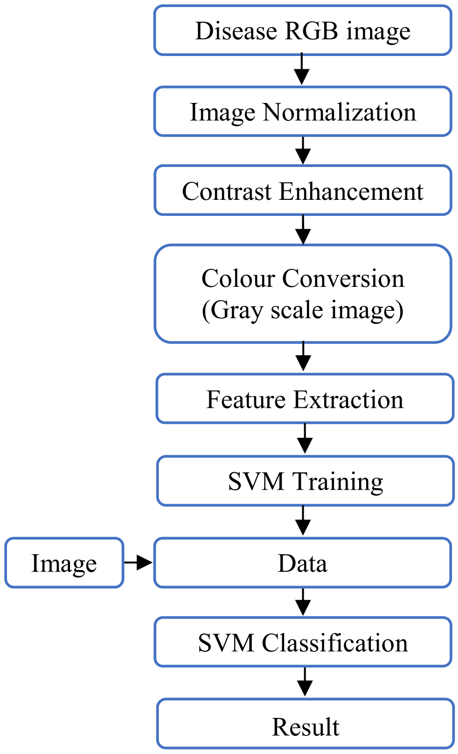


Fig.1. Research methodology

III. Research Methodology

This paper proposes a system which is based on Support Vector Machine (SVM) classifier, since all concepts for any vision related approach for image classification remains the same. A digital camera is used to capture a tea leaf, and image processing techniques are applied to these images to extract various features. The useful features are used to train the SVM which performs the classification as shown in Figure 1.

A. Image Acquiation

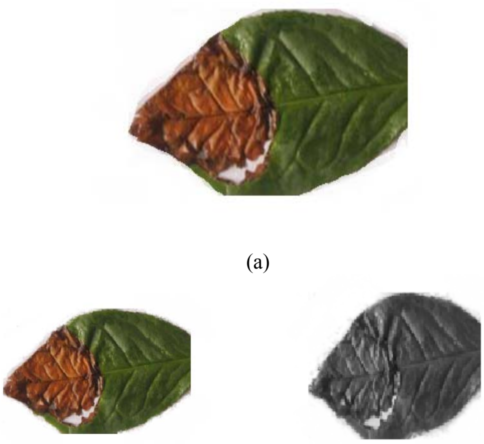
A Nikon D5600 was used to take pictures of many tea leaves from Bangladesh Tea Research Institute (BTRI)located in Sri Mongol. A large numbers of image samples were collected, some which were affected by diseases (brown blight, and algal) and others that were unaffected or otherwise healthy. Figure 2 shows a representation of these three types of leaves.

(a) (b) (c)

Fig.2. Acquired tea leaf images (a) Brown Blight(b) Algal(c) Healthy

(a) (b)

Fig.3. Contrast Enhancement



(a)

(b) (c)

Fig.4. (a) Original image (b) Normalized image (c) Grayscale image

B. Preprocessing

The aim of pre-processing is to improve the image data and suppresses unwanted distortion or enhances some image features which would be important for further processing. We enhance the image so that we can get the better features as shown in Figure 3.

C. Image processing

This consists of two steps: Image Normalization and Color space Conversion. In image normalization, the captured image has been converted into the normalized image (i.e. the image size have been made constant). Then the color space of the normalized image has been converted into grayscale image as shown in Figure 4.

IV. Feature Selection and Extraction

Feature selection is one of the most important segments in this research.Feature selection works by selecting the best features based on univariate statistical tests. The features are carefully selected based on their unique differences between the three types of leaves. The computational complexity of the process was reduced by discarding features that were not deemed useful, or features that had no visible variance between healthy and diseased leafs. In this research total 11features have been utilized, while three commonly used features in past researches have been eliminated. After selecting these features, the only thing left is to extract these feature values and use them for classification.

A. Classification

Image Classification is the process by which an image can be categorized into different predefined classes based on diverse properties of images. There are so many classifiers that exist in Artificial Intelligence (AI). Every classifier has some advantages and drawbacks. In this research Support Vector Machine (SVM) classifier has been used because it has some advantages over other classifiers including:

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.

B. Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm used for both classification and regression. In SVM, each data item is plotted as a point in n-dimensional space; the number of dimensions corresponding to the number of features being classified. The classification is obtained by discovering the hyper-plane that uniquely distinguishes between different groups of scattered data points. The following equation describes the SVM process [17-18]:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i \qquad (1)$$

Subject to the constraints:

$$y_i(w^T \phi(x_i)+b) \geq 1-\xi_i \, and \, \xi_i \geq 0, i=1,.....N \qquad (2)$$

where C is the capacity constant, w is the coefficient vector, and ξ is the parameter for handling non-separable data. The index i represents the current iteration out of maximum N iterations. y_i are the class labels, and x_i are the independent variables. The Kernel ϕ transforms independent input data into the feature space. Large values of C will tighten tolerance and thus exacts a higher penalty for errors. The SVM process of differentiating between two distinct clusters of data points is shown in Figure 5.

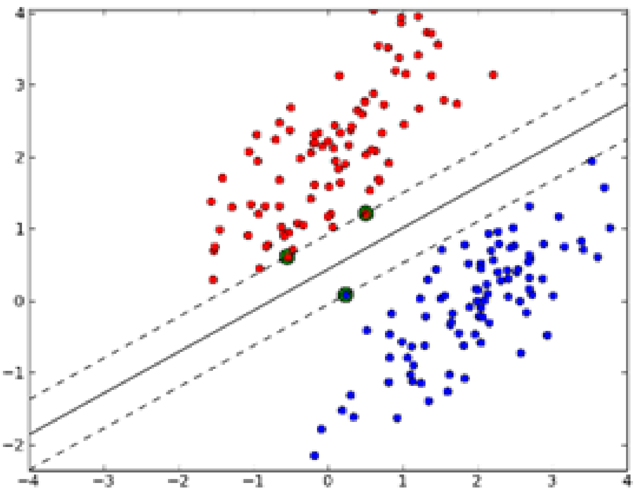


Fig.5. Support vector machine classifier

Previously, the system would give three clusters and from those three only one would be taken manually. But, in this proposed system no cluster is required. Therefore, when a new image is taken the system automatically extracts the features. Then the new image features and the training data are put into the classifier. Finally, the classifier would give the predictable result [17-22].

V. RESULTSAND DISCUSSIONS

In this experimental study, 150 samples for training and another 50 samples for testing has been used in the system. Furthermore, 10 features have been taken to train the classifier as shown in Table I. Non essential features have been eliminated. Previous researches used 13 features, however; three features including homogeneity, smoothness, and IDM were determined to not show any difference between healthy, brown blight disease, and algal disease using statistical analysis. Therefore these three features have been excluded from the SVM classification.However, as can be seen from Table II and III, after elimination of these features, the accuracy has not degraded very much but the system has become faster than before. Our algorithm has been tested by processing 300 tea leaves and has been found to be approximately 1.5 minutes faster than previous researches. That averages to speed up of 300ms per leaf!

There has also been the removal of manual clustering. The designing of the system is in such a way that it can select the best cluster automatically. From the graph in Figure 6, it is shown that as the number of training cycle's increases, the error rate decreases and after 300 training cycle the curve becomes steady, which implies that the classifier do not require more than 300 samples to become fully trained.

The average features of the system for Healthy leaf, Brown Blight leaf and Algal leaf and results after conducting the experiment are shown in Table I. It is seen that there is a significant difference in contrast between healthy and diseased leaf. It also seen that the correlation, standard division, skewness value is greater than healthy leaf value. But the kurtosis feature of healthy leaf value is less than diseased leaf. There is difference between healthy and diseased leaf in RMS value but there is only slight difference between the RMS values of the two types of diseased leafs. So from the table it is seen that when leaf is affected by disease then correlation, standard division, skewness value is increased and there is significant change in contrast. In the experiment the overall accuracy is 93.33% as shown in Table III.

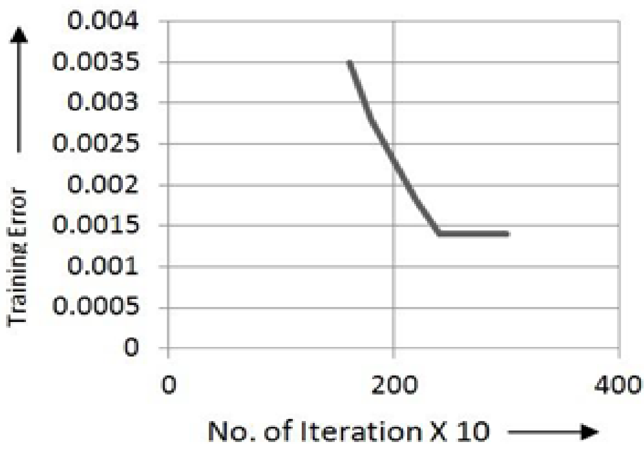


Fig.6. Training Error vs Number of Iteration

TABLE I. THE AVERAGE FEATURES OF THE SYSTEM FOR HEALTHY LEAF, BROWN BLIGHT LEAF AND ALGAL LEAF

Features	**Healthy**	**Brown Blight**	**Algal**
Contrast	0.0951	0.1850	0.1435
Correlation	0.9830	1.0217	1.0309
Energy	0.2618	0.2029	0.2248
Mean	101.3420	101.9871	88.0292
Std Deviation	62.3517	65.2077	67.8828
Entropy	7.2012	7.6262	7.4902
RMS	15.9469	16.7164	16.6964
Variance	2.1339e+03	2.6127e+03	2.4773e+03
Kurtosis	2.4668	2.2952	2.4351
Skewness	0.6693	0.7689	0.9398

TABLE II. THE SUCCESS RATE IN THIS EXPERIMENT

ClassifiedLeafCondition	Actual Leaf Condition		
	Healthy	Brown Blight	Algal
Healthy	49	2	3
Brown Blight	0	45	1
Algal	1	3	46

TABLE III. THE ACCURACY OF THIS EXPERIMENTAL SYSTEM:

Image Type	Success Rate(%)
Healthy	98
Brown Blight	90
Algal	92

VI. CONCLUSION

In this research, an automated system has been developed for detecting three different types of tea leaf's diseases using Support Vector Machine (SVM) with less number of features.The proposed method is able to classify the disease more accurately (93%) compared to the other classifiers and neural network (91%).This method can also be extended to reduce and extract the features that are required thereby reducing the required processing time. As shown through simulation results, the proposed algorithm can process a leaf classification by 300ms quicker than previous research using SVM.

VII. FUTURE WORK

The future scope of this work is to improve the segmentation process and to classify the disease using different classifier in order to carry out more accurate result.

VIII. ACKNOWLEDGEMENT

The authors gratefully acknowledge the research grants (FY 2016-2017) received from the Department of Electrical and Electronic Engineering, Independent University, Bangladesh.

REFERENCE

[1] H. L. Khatibl, F. Hawels, H. Hamdi, and N. L. Mowelhi. “Spectral Characteristics Curves of Rice Plants Infected by Blast,” IEEEGeoscience and Remote Sensing Symposium, vol. 2, pp. 526-528,1993.

[2] H. Panda , “The Complete Book on Cultivation and Manufacture of Tea” (2nd Revised Edition) Format: Paperback ISBN: 9788178331683 Code: NI242.

[3] Online:https://www.plantvillage.org/en/topics/tea/diseases_and_pests_description_uses_propagation, accessed on September 2016.

[4] M. Ahmed and M.S.A. Mamun,“Distrributional Pattern and SeaonalAbundance of Major PestsofTea in Bangladesh”, Tea J. Bangladesh, Vol. 41, pp. 1-10, 2012, ISSN: 0253-5483.

[5] A.F.M.B. Alam, “Profile of tea industry in Bangladesh” pp.1-22 (Jain NK ed. 1999), Global Advances in Tea Science. New Delhi: Aravali Books. 882pp. Chen Z and Chen X. 1989. An analysis of the world tea fauna. J. Tea Sci. Vol. 9,pp. 13–22.

[6] Tea Productions, Consumptions and Exports: Bangladesh Perspective Tasnuba Nasir and Mohammad Shamsuddoha Faculty of Business Administration (FBA), University of Science and Technology Chittagong (USTC), 2003.

[7] Harler, C.R. 1964. The Cultivation and Marketing of Tea. 3d. Ed. Oxford: Oxford University Press.

[8] Kholis Majid, Yeni Herdiyeni, Annu Rauf, “I-Pedia: Mobile Application For Paddy Disease Identification Using Fuzzy Entropy And Probabilistic Neural Network”, ICACSIS, 2013.

[9] Qing Yao;Zexin Guan;Yingfeng Zhou;Jian Tang; Yang Hu;Baojun Yang. Application of support vector machine for detecting rice diseases using shape and color texture features. International Conference on Engineer-ing Computation, 2009, pp.79-83.

[10] Elham Omrani, Benyamin Khoshnevisan, Shahaboddin Shamshirband, Hadi Saboohi, Nor Badrul Anuar, Mohd Hairul Nizam Md Nasir, “Potential of radial basis function-based support vector regression for apple disease detection”, Measurement, Vol.55, pp.512–519, Sept-2014, (Elsevier).

[11] Santanu Phadikar and Jaya Sil, “Rice Disease Identification using Pattern Recognition Techniques”, Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008), 25-27 December, 2008, Khulna, Bangladesh, pp. 420-423..

[12] Haiguang Wang, Guanlin Li, Zhanhong Ma and Xiaolong Li, “Image Recognition of Plant Diseases Based on Principal Component Analysis and Neural Networks”, 8th International Conference on Natural Computation (ICNC 2012), pp.246-251.

[13] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad, “A Framework for Detection and Classification of Plant leaf and Stem Diseases”, 2012 International conference on Signal and Image Processing, Chennai, India, pp.113-118.

[14] V. A. Gulhane, Dr. A. A. Gurjar, “Detection of Diseases on Cotton Leaves and Its Possible Diagnosis”. 2011 International Journal of Image Processing (IJIP), Vol.5, Issue 5, pp.590-598.

[15] P. Revathi, M. Hemalatha, “Cotton Leaf Spot Diseases Detection Utilizing Feature Selection with Skew Divergence Method”, International Journal of Scientific Engineering and Technology (ISSN : 2277-1581), Vol.3, Issue 1, pp. 22-30, Jan 2014.

[16] S. Arivazhagan, R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini, “Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features”, Agric Eng Int: CIGR Journal, Vol.15, pp.211-217, March 2013.

[17] V. Cherkassky, M. Yunqian,"Practical selection of SVM parameters and noise estimation for SVM regression," Neural Networks, Volume 17, Issue 1, Pages 113-126, 2004.

[18] E. G. Ortiz-García, S. Salcedo-Sanz, Á. M. Pérez-Bellido, J. A. Portilla-Figueras, “Improving the training time of support vector regression algorithms through novel hyper-parameters search space reductions,” Neurocomputing, Volume 72, Issues 16–18, Pages 3683-3691, 2009.

[19] Y. Yuan, M. Zhang, P. Luo, Z. Ghassemlooy, L. Lang, D. Wang, B. Zhang, D. Han, “SVM-based detection in visible light communications,” Optik - International Journal for Light and Electron Optics, Volume 151, Pages 55-64, 2017.

[20] K. Li, L. Wang, J. J. Wu, Q. Zhang, G. Liao, L. Su, “Using GA-SVM for defect inspection of flip chips based on vibration signals,” Microelectronics Reliability, Volume 81, Pages 159-166, 2018.