

# Tea Leaf Diseases Recognition using Neural Network Ensemble

Bikash Chandra Karmokar<sup>1</sup>, Mohammad Samawat Ullah<sup>2</sup>, Md. Kibria Siddiquee<sup>3</sup> and Kazi Md. Rokibul Alam<sup>4</sup>

Department of Computer Science and Engineering,  
<sup>1,3,4</sup> Khulna University of Engineering and Technology, Bangladesh  
<sup>2</sup>Durham University, United Kingdom

## ABSTRACT

This paper proposes a tea leaf diseases recognizer (TLDR), an initiative to recognize diseases of the tea leaf. In TLDR, at first the image of the tea leaf is cropped, resized and converted to its threshold value in the image processing. Then feature extraction method is applied. Neural Network Ensemble (NNE) was used for pattern recognition. The extracted features are passed to the ANN along with the disease type and the ANN is trained. When a new image is uploaded into the system the most suitable match is found and the disease is returned. After going through the testing process 91 % of accuracy was found. The proposed solution would support the tea industry of Bangladesh to grow in the global market and also increase its tea production by minimizing the effect of tea leaf diseases.

## Index Terms

Negative correlation learning, Feature Extraction, Image Processing, Tea leaf diseases.

## 1. INTRODUCTION

Tea leaf disease is one of the crucial causes that reduces quantity and degrades quality of the tea in Bangladesh. 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 good stage then the color of the leaf is different but as soon as the leaf is affected by some harming pathogens, the color changes automatically. The emergence of plant disease has become more common now a day as factors such as climate and environmental conditions are more unsettled than ever [1]. Tea plant diseases are usually caused by fungi, bacteria, alga and viruses. Also there are other diseases which are caused by adverse environmental conditions [2]. 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 earlier stage is an important task. Hence an intelligent system for recognition, prevention and control of tea leaf diseases is needed.

Technological advancement is gradually finding its applications in the field of agriculture [3]. The information and communication technology (ICT) application is going to be implemented as a solution in improving the status of the agricultural sector [4]. The idea of integrating ICT with agriculture sector motivates the development of an automated system for tea leaf disease classification and its recognition. Some works have been done on recognition of rice diseases, citrus diseases, crop diseases, wheat leaf diseases, Betel vine plant diseases etc. But in case of tea leaf diseases recognition

there is no significance improvement in the research area. So we tried to develop the TLDR which will help the tea industry of Bangladesh as well as of the world.

The outlined is as follows. Section II describes existing works on diseases recognition. Section III mentions our proposed intelligent approach for tea leaf disease recognition. Experimental results are presented and discussed in Section IV. Finally, conclusions are explained in Section V.

## 2. EXISTING WORKS

From many years leaf disease recognition is gradually increasing and in case of rice, wheat, betel vine there is a lot of examples out there. But in case of tea leaf disease recognition there is no such work specially in Bangladesh. Most of the research works done on leaf recognition used various types of approach for getting high accuracy. But none of them able to reach to the point.

The system proposed in [5] presents a SVM-based Multiple Classifier System (MCS) for pattern recognition of wheat leaf diseases. Based on the infected images of various rice plants the paper [6] describes a software prototype system for rice disease detection. Using an approach [7] that regularizes and extracts eigen feature from image and developing and decomposing scatter matrix cotton leaf diseases also identified. Design and implementation of an artificial vision system which extracts specific geometric and morphological features from plant leaves presents in paper [8]. The proposed system consists of an artificial vision system, a combination of image processing algorithms and feed forward neural network based classifier. A fuzzy surface selection technique for feature selection was used. Based on support vector machines a prediction approach for developing weather based prediction models of plant diseases is proposed in paper [9]. The performance of conventional multiple regression, artificial neural network (back propagation neural network, generalized regression neural network) and support vector machine (SVM) was compared. It was concluded that SVM based regression approach has led to a better description of the relationship between the environmental conditions and disease level which could be useful for disease management. Another paper [10] proposed Back propagation neural network for recognition of damaged leaves. It was proved that just a back propagation network and shape of leaf image is enough to specify the species of a leaf. Prewitt edge detection and thinning algorithm is used to find leaf tokens as input to back propagation algorithm. It was reported that there is a scope for enhancement of this work which involves more experimentation's with large training sets to recognize various damaged leaves of different diseases.

### 3. PROPOSED TEA LEAF DISEASE RECOGNIZER (TLDR)

This paper proposes a tea leaf diseases recognizer (TLDR) that recognizes affected leaf of the tea plant. The input of TLDR is the images of leaf of tea taken by the digital camera. The proposed TLDR combines feature extraction [11] and neural network ensemble (NNE) for training that yields a better performance. A NNE consists of a set of independent trained NN whose predictions are combined by various methods. Bagging, Boosting, Negative correlation learning (NCL) [12], DECORATE [13] etc are familiar techniques of NNE.

There are numerous reasons for using NNE; some of them are given below:

- Intuition evinces that in almost every cases, NNE offers better generalization ability when compared with a single monolithic NN in the same problem domain.
- In spite of under-fitting and over-fitting [14] problems, NNE works well in comparison with an individual NN, as NNE involves data re-sampling.
- In many cases, additional information is available that cannot be given to the same NN tackling the core information. The additional information can be trained separately and then combined [15].

In order to develop TLDR among various NNE techniques, NCL [12] has been employed. The NCL technique along with efficient feature extraction enables TLDR to attain faster training and recognition performance.



Fig. 1 Cephaleuros virescens affected tea leaf.

#### 3.1 Major Stages of TLDR

Major stages of the proposed scheme proceed as follows:

- 1) *Pre-processing*: This process starts through the detection of affected leaves of the tea tree. An affected area becomes grey or brown of the leaf. A captured image of tea leaf which is affected by grey blight has been shown in Fig. 1.

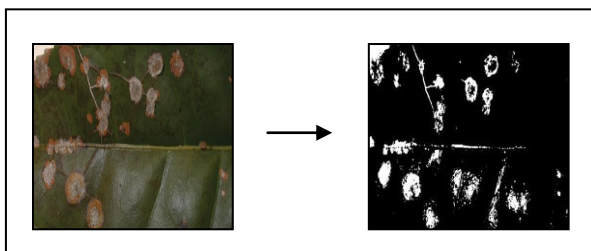
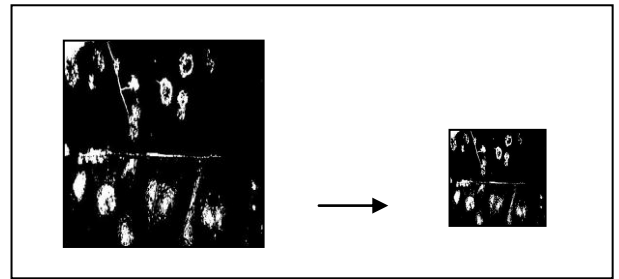


Fig. 2: Conversion into threshold image.



(a) Original size (b) Normalized size

Fig. 3 Normalization

- 2) *Image processing*: At first captured image has been converted into its threshold value where affected areas is shown in white pixels and whole leaf is shown in black pixels which has been shown in Fig. 2. Then the converted image has been normalized to 30x33 scale pixels which has been shown in Fig. 3. Then feature extraction method has been applied and NCL [12] algorithm has been used to train these images.

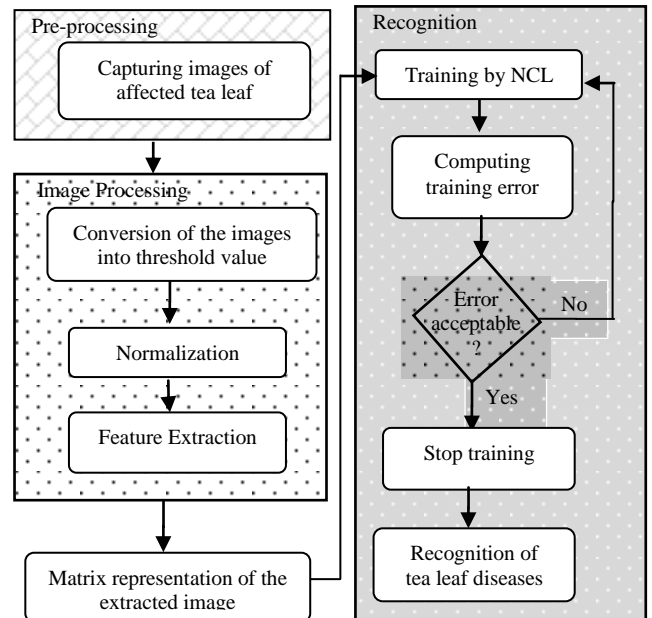


Fig. 4: Block diagram of the proposed TLDR.

- 3) *Operational Steps of TLDR*: The block diagram of the proposed TLDR has been shown in Fig. 4. Its operational steps are as follows:

- Capturing images of tea leaf by using a digital camera.
- Converting the images into threshold images.
- Applying normalization process to get the reduced size of pixels.
- Extracting the features of the images.
- Matrix representation of the extracted images.
- Training of these obtained data i.e. patterns by using NCL algorithm up to an expected error level.
- Testing i.e. recognizing leaf by taking inputs later on.

## 4. EXPERIMENTAL STUDIES

### 4.1 Experimental Setup

In the experiment, we have used 5 types of affected leaf of tea as input to train where each type has 10 samples that make  $5 \times 10 = 50$  samples of pre-defined images of same pixel size. Then TLDR can recognize any of 5 types of diseases. In the training of TLDR, the NCL [12] learns through adjusting its weights which is a supervised learning algorithm. We have sufficiently trained TLDR by NCL in an iterative process that adjust the weights in the mode of pattern-by-pattern updating considering its penalty function. In TLDR, the chosen architecture of individual NN involved in NCL has  $30 \times 33$  i.e. 990 nodes (after normalization) in input layer and one hidden layer with flexible no. of nodes. Its output layer contains 5 nodes and the training continues until the error reaches to a certain error level.

The output of NCL is a simple averaging of outputs of a set

$$F(n) = \frac{1}{M} \sum_{i=1}^M F_i(n)$$

of NNs which is given by  $F(n) = \frac{1}{M} \sum_{i=1}^M F_i(n)$ , where M is the number of the individual NNs in the NNE,  $F_i(n)$  is the output of NN i on the nth training pattern, and  $F(n)$  is the output of the NNE on the nth training pattern. NCL uses a correlation penalty term into the error function of each individual NN in the NNE so that all the NNs can be trained simultaneously and interactively on the same training data set. The error function  $E_i$  for NN i in NCL is defined by

$$E_i = \frac{1}{N} \sum_{n=1}^N E_i(n) = \frac{1}{N} \sum_{n=1}^N \frac{1}{2} (F_i(n) - d(n))^2 + \frac{1}{N} \sum_{n=1}^N \lambda p_i(n)$$

where  $E_i(n)$  is the value of the error function of NN i at presentation of the nth training pattern. Here, the first term in the right side of the above equation is the empirical risk function of NN i. The second term  $p_i$  is a correlation penalty function which has the form

$$p_i(n) = (F_i(n) - F(n)) \sum_{j \neq i} (F_j(n) - F(n)).$$

Here the parameter  $0 \leq \lambda \leq 1$  is used to adjust the strength of the penalty.

### 4.2 Performance of Recognition

As mentioned earlier, TLDR has been trained by 50 samples. Here, training has been continued until the error rate has reached to 0.02. As shown in Fig. 5, the error rate decreases as the number of training cycles increases and the curve becomes steady after 2500 training cycles. To gain the better performance, we have chosen 3000 training cycles for our experiment.

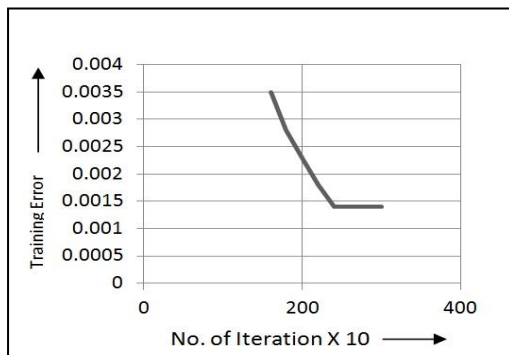


Fig. 5 Training error

Table I shows the performance of the proposed TLDR system. Considering  $\lambda =$  an arbitrary value and  $\lambda = 0$ , NCL has been used to develop TLDR. Here for  $\lambda = 0$ , individual NNs are trained independently which is as same as Scale Conjugate Gradient (SCG) algorithm. The table shows that NCL produces better recognition accuracy in comparison with SCG in all perspectives. Moreover, NCL learns faster than SCG.

Table 1 Recognition Performance of TLDR

No. of outputs = 5	Iterations	Training time (s)	Accuracy (%)	
			NCL	SCG
No. of samples $10 \times 5 = 50$	3000	NCL	Test	Test
		160	91	75
Input pixel size = $30 \times 33$	3000		90	79
			92	77
			91	77
	Average	160	91	77

### 4.3 Performance with different NNs involved in NCL

For the pursuit of better generalization, NCL has been used with different NN architectures to develop TLDR. In the domain of the same data space, NCL has been designed considering five and ten individual NNs. Table II shows the comparison of performance of TLDR among these different NCL architectures. As training data space is large, NCL consisting of ten NNs with feature extraction shows the best accuracy.

Table 2 Performance Comparison of TLDR

Network Type	Accuracy (%)
NCL with 05 NNs	85
NCL with 10 NNs	89
NCL with 10 NNs with feature extraction	91

## 5. CONCLUSION

According to our knowledge the proposed system of tea leaf disease recognition is the first initiative in Bangladesh. Above method can identify tea leaf which are infected and thus preventive action can be taken well in advance such that the entire production can be saved from damages. The methodology begins with image acquisition. Captured images are processed and after feature extraction the extracted features are used as input which appropriately identifies tea leaf diseases. Once the disease identified its treatment advisory can be provided. The method of detecting the disease is low cost effective and non-destructive as it only requires the digital photograph of the leaf samples in random.

Periodic inspection of the farm is required to prevent the disease. This method can also be extended to detect diseases of all kind of leaf without major changes to initiate early preventive action. A future plan of improvement is to employ TLDR to recognize tea leaf diseases in real time using mobile.

## **6. REFERENCES**

- [1] Camargo A., Smith J.S. (2009) Biosystems Engineering I02, 9-21.
- [2] Md. Saidur Rahman Khan, SUST, Disease scenery of Bangladesh in tea.
- [3] Narendra V.G., Hareesh K.S. (2010) International Journal of Computer Applications, 2, 1.
- [4] Nur Badariah Ahmad Mustafa, Syed Khaleel Ahmed, Zaipatimah Ali, Wong Bing Yit, Aidil Azwin Zainul Abidin, Zainul Abidin Md Sharrif (2009) IEEE International Conference on Signal and Image Processing Applications.
- [5] Yuan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo. "SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases" Proceedings of 2010 Conference on Dependable Computing (CDC'2010) November 20-22, 2010, Yichang, China
- [6] 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
- [7] Ajay A. Gurjar, Viraj A. Gulhane. Disease Detection On Cotton Leaves by Eigenfeature Regularization and Extraction Technique. International Journal of Electronics, Communication & Soft Computing Science and Engineering (IJECSSE) Volume 1, Issue 1
- [8] Panagiotis Tzionas, Stelios E. Papadakis and Dimitris Manolakis[2005] Plant leaves classification based on morphological features and fuzzy surface selection technique, 5th International Conference ON Technology and Automation ICTA'05, Thessaloniki, Greece, pp.365-370,15-16 .
- [9] Rakesh Kaundal, Amar S Kapoor and Gajendra PS Raghava[2006] Machine learning techniques in disease forecasting: a case study on rice blast prediction, BMC Bioinformatics.
- [10] M. S. Prasad Babu and B. Srinivasa Rao[2007] Leaves Recognition Using Back Propagation Neural Network-Advice For Pest and Disease Control On Crops, IndiaKisan.Net: Expert Advisory System.
- [11] Isabelle Guyon and Andr e Elisseeff, "An Introduction to Feature Extraction", Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer, 2006.
- [12] Y. Liu, X.Yao, "Ensemble learning via negative correlation", Neural Networks 12 (1999) 1399–1404.
- [13] Melville and Mooney, "Creating Diverse Ensemble Classifiers to Reduce Supervision", PhD Thesis, Department of Computer Sciences, University of Texas at Austin, November 2005.
- [14] Hafiz T. Hassan, Muhammad U. Khalid and Kashif Imran, "Intelligent Object and Pattern Recognition using Ensembles in Back Propagation Neural Network", International Journal of Electrical & Computer Sciences (IJECS-IJENS) Vol: 10 No: 06.
- [15] Robi Polikar, "Ensemble based systems in decision making", Article IEEE Circuits and Systems Magazines, 2006.