International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 1772 –1779

Detection and Prevention of Tea Leaf Diseases by using K-NN Algorithm

Ms. Payal Mhaiske, Prof. K.V.Warkar
Department of Computer Engineering B.D.C.E Wardha

Abstract
All around the world tea is popular beverage and in India the cultivation of tea plays a vital role. The proper growth of tea leaves leading to its reduction of hindering for the production of tea because many diseases affect on the growth of tea leaves. But, if identified the disease at the early age then it would solve all the problems through the pruning of the disease leaves to prevent the spread further of the disease. An image processing is the best option to solve this problem to detect and diagnose the disease. The main goal of this research to developed an image processing system for identify and classify the six most widespread tea leaves diseases from healthy tea leaf. Disease identification is the first step in this step there are many methods are used for leaf disease identification. In this paper, K-NN classifier (K-NN) is used for disease recognition. During the classification analyzed the thirteen features. Thus features are used to find the most suitable match for the disease. An image is every time uploaded into a K-NN database. When the new picture is uploaded into a system the most suitable match is found and then recognized the disease. In this approach the number of features compared by the K-NN classifier, which retains an accuracy of more than 96%. This also speed up the identification process, with each tea leaf image taking time is 200ms less processing time compared to the previous research by using K-NN, thus ensuring in given time frame the greater number of leaves can be processed. Thus proposed system increase the efficiency of the detection, identification, classification and prevention process will enable the tea industry in India to become globally more competitive for reducing the losses which are suffered due to the leaf diseases and thus increasing the overall rate of tea production.

Keywords: *Image processing, Disease detection, Disease recognition, Feature Extraction, K-nearest neighbor(K-NN)*

1. Introduction

Tea is one of the essential beverages in India. Most of the Indian people start their day with a cup of tea. Assam 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 Tea production in India 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]. India 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

ISSN: 2233-7857 IJFGCN

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 .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 India.

RELATED WORK

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.

ISSN: 2233-7857 IJFGCN

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.Healthy leaf **Fig 2. Algal disease leaf** **Fig 3.Brown blight leaf**

System Architecture

The proposed work is planned to be carried out in the following manner.

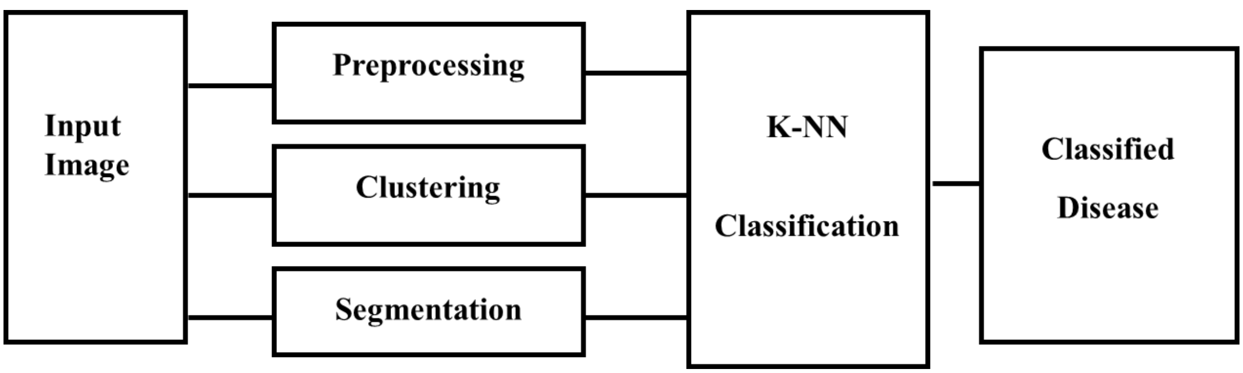

Fig 4. System Architecture

Image Pre-processing

Noise gets added during acquisition of leaf images. So we use different types of filtering techniques to remove noise. We create device independent color space transformation structure. Thus we create the color transformation structure that defines the color space conversion. The next step is that we apply device-independent color space transformation, which converts the color values in the image to color space specified in the color transformation structure. The color transformation structure specifies various parameters of transformation. A device independent 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 brightness and contrast on display device used. Thus the RGB system is a color space that is dependent. To improve the precision of the disease detection and classification process, a device independent color space is required. In device independent color space, the coordinates used to

ISSN: 2233-7857 IJFGCN

specify the color will produce the same color regardless of the device used to take the pictures. CIE L*a*b is a device independent color space in which a & b components carry color information.

Image segmentation

Image segmentation is the process used to simplify the representation of an image into something that is more meaningful and easier to analyse. Image will be segmented to fetch out the image edges and then detected all required parameters.

Masking green pixels

In this step, we identify the mostly green colored pixels. After that, based on specified threshold value that is computed for these pixels, the mostly green pixels are masked as follows: if the green component of the pixel intensity 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 the green colored pixels mostly represent the healthy areas of the leaf and they do not add any valuable weight to disease identification and furthermore this significantly reduces the processing time.

Removing the masked cells

The pixels with zeros red, green, blue components as well as pixels on the boundaries of infected cluster are completely removed. This is helpful as it gives more accurate disease classification and significantly reduces the processing time. Infected cluster is converted from RGB to HSI color format.

K-means Algorithm

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The algorithm has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means because of the k in the name. One can apply the 1-nearest neighbor classifier on the cluster centers obtained by k-means to classify new data into the existing clusters. This is known as nearest centroid classifier or Rocchio algorithm.

K-Nearest Neighbors(K-NN)

K- nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). K-NN has been used in statistical estimation and pattern recognition. In pattern recognition, the k-nearest neighbors algorithm is a non-parametric method used for classification and regression. In both cases, the input consist of k closest training examples in the future space. The output depends on whether K-NN is used for classification or regression. In K-NN classification , the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k-nearest neighbors.In K-NN regression , the output is the property value for the object. This value is the average of the values of k-nearest neighbors.

GLCM methodology

Gray level Co-occurrence matrix (GLCM) is generated for each pixel map

1. The graycomatrix function creates a gray level co-occurrence matrix by calculating how frequently a pixel with the particular intensity value i occurs in a specified spatial relationship to a pixel with the value j.
2. By default this spatial relationship is the pixel of interest and its immediate right pixel.
3. However we can specify some other spatial relationship between twos. To create multiple GLCMs, specify an array of offsets to the graycomatrix function. These offsets define pixel relationships of varying direction and distance.
4. Calculating statistics from GLCM matrix also known as SGDM

ISSN: 2233-7857 IJFGCN

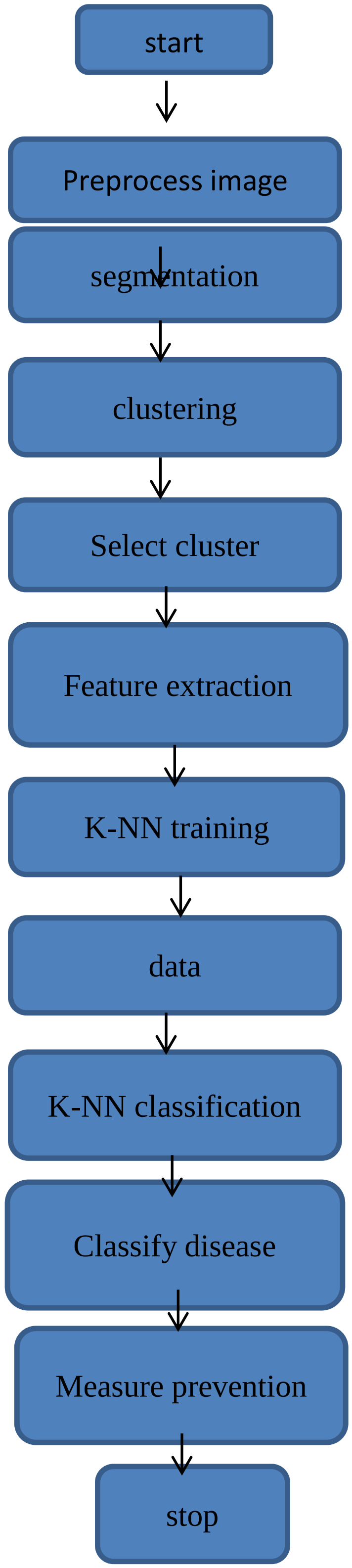

Fig 5. Flow diagram

ISSN: 2233-7857 IJFGCN

Algorithm for k-means(distance)

K-means(d)
Step 1: set no. of cluster=4
Step 2: for i=1 to no. of cluster
    extract rgb cluster
    image[]=extract(k,d)
    end for

Algorithm for K-NN(test, train[])

K-NN (test , train[])
Step 1: Fetch test features into an array test[]
Step 2: Call k match function to match features
        for i=1 to train.lenght
        match(test [], train[][i])
        end for
Step 3: Sort match[] in desending order to find class c
Step 4: return c

Algorithm for disease detection

Step 1: Read image
Step 2: Create done of original image for processing
Step 3: Convert image from RGB2LAB
        image= rgb2lab(image)
Step 4: Perform segmentation on scale of [256 256]
Step 5: Call K-means(image)
Step 6: Ask for ROI in image d
Step 7: Extract GLCM features of d
        t=[1 13]=GLCM(d)
Step 8: To find class
        c=KNN(t,'train_d')
Step 9:Display result

Conclusion

In this research, an automated system has been developed for detecting six different types of tea leaf's diseases using k-means and K-nearest neighbors(K-NN) algorithm with number of features. The proposed method is able to classify the disease more accurately compared to the other classifiers and neural network .This method can also be extended to reduce and extract the features that are required thereby reducing the required processing time. The proposed algorithm can process a leaf classification by 200ms quicker than previous research using K-NN.

ISSN: 2233-7857 IJFGCN

REFERENCES

[1] Md. Selim Hossain1 , Rokeya Mumtahana Mou1 , Mohammed Mahedi Hasan2 , Sajib Chakraborty1 , M. Abdur Razzak1 "Recognition and Detection of Tea Leaf's Diseases Using Support Vector Machine IEEE 2018.

[2] Prabira Kumar Sethy#1, Baishalee Negi*2, Santi Kumari Behera#3,Nalini Kanta Barpanda#4, Amiya Kumar Rath , An Image Processing Approach for Detection, Quantification, and Identification of Plant Leaf Diseases -A Review IJET UAE 2017..

[3] Vishal Mani Tiwari &Tarun Gupta , Plant Leaf Disease Analysis using Image Processing Technique with Modified SVM-CS Classifier IJEMT 2017.

[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 Engineering 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] 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.

[12] 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.

[13] 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.

[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] 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.

[16] 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.

[17] 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.

ISSN: 2233-7857 IJFGCN

[18] 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.

[19] 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.

[20] 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.

ISSN: 2233-7857 IJFGCN