DETECTION OF SEPTORIA SPOT ON BLUEBERRY LEAF IMAGES USING SVM CLASSIFIER

M. Latha and S. Jaya

Department of Computer Science, Sri Sarada College for Women, India

Abstract

Identification and classification of the plant leaf is efficient way to preventing loss occurred in agricultural field. The Septoria leaf spot is mainly affect the leaves which caused by a fungus, flu, bacteria. The Production of blueberry fruit is decreasing due to the disease affected on its stem and leaf. Small brown spots are frequently visible on blueberry leaves at specific period in the year. The spots, generally surrounded by bright yellow halos, start on the lower leaves and slowly appear on upper leaves over time. Image processing technology has been proved to be an efficient analysis to identify and detect the disease on a leaf. This proposed paper intends to focus to detect and classify a Septoria leaf spot on blueberry using Image Processing techniques such as, k-means clustering (k-nearest neighbor) for Segmentation, Gray-Level Co-occurrence Matrix for feature extraction and Support Vector Machine classifier to detect the leaf stage whether it is affected by Septoria spot or not. Totally 13 features have been extracted from each Blueberry leaf images where dataset of 40 images were taken for training and testing process partially and obtained the accuracy level was 96.77% using F-measure.

Keywords:

Septoria Leaf Spot, K-means Clustering, Segmentation, Feature Extraction, GLCM, SVM

1. INTRODUCTION

Blueberries have incredible nutrition and health benefits. It is used for more medicinal purpose and having one of the highest antioxidant properties in nature. The native of this fruit is North America. In India blueberry farming is very less and it has a huge future potential in agriculture due to excellent health benefits. The benefits of blueberries are high antioxidant foods, high in nutrients, treatment for cancer, preventing heart disease, frequent brain function. The Septoria leaf spot is one of the diseases which affect the plants gradually from top to bottom. It is very difficult to identify a diseased plant to the farmer.

1.1 ROLE OF IMAGE PROCESSING

In India Blueberry farming will have bright future in agriculture. In such a case, the Image processing techniques has vital role to protect the plant from these diseases. Image processing techniques, has found in different fields such as industrial inspection, medical imaging, remote sensing, and agricultural field etc. The growth of automatic detection system using advanced computer technology such as image processing help to support the farmers in the identification of diseases at an initial stage when they scan targeted leaf and provide useful information about solution for its control. Detection of leaf spot disease using following techniques such as image acquisition, feature extraction, disease spot segmentation, image pre-processing and disease classification were carried out by various workers.

1.2 PRE-PROCESSING

The paper consists of three level of processing. They are Low level processing, Intermediate-level processing and High-level processing. Datasets of 40 healthy and infected blueberry leaf images were used for both training and testing stages. The paper followed by preprocessing, segmentation, feature extraction, classification and identifies the leaf state. In preprocessing Contrast stretching is an image enhancement technique that improves the contrast in an image by expanding the dynamic range of intensity values it contains. Resize the image with $800 \times 500 \times 3$ resolutions. K-means clustering algorithm is used for segment the diseased part of the leaf images and obtained segments are passed through a trained neural network. The final result indicates that the proposed approach precisely and automatically detects the leaf disease. The classifier is implemented which is based on statistical classification had performed well and can successfully classify the leaves. The segmented region will be converted into gray scale image as well as binary image.

1.3 FEATURE EXTRACTION

The purpose of Gray Level Co-occurrence Matrix algorithm is used for texture analysis and it is a tabulation of how often different combinations of pixel brightness values gray level occur in an image which expands the features in an image, which are Contrast, energy, homogeneity, mean, standard deviation, entropy, RMS, variance, smoothness, Skewness and kurtosis.

1.4 SVM CLASSIFIER

The Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. It is mostly used in classification problems. The Support Vector Machines performs the task of classification of diseases, results show that the methodology place forward in this paper provides considerably better results than the previously used disease detection techniques. The proposed work of this paper identify the blueberry leaf is affected or not used by SVM. SVM is a frontier which best segregates the two classes are hyper-plane and line which leads to classify a blueberry plant leaf is diseased by Septoria leaf spot or not.

2. LITERATURE SURVEY

The author refers about diseased leaf of grape using SVM classifier followed by K-means clustering. Grape constitutes one of the most widely grown fruit crops in the India. The accuracy level had achieved 88.89% where SVM is the best classifier to classify a data [1]. In this research work, Plant leaf diseases can be detected based on the disease symptoms of disease affected

leaves is considered for experimentation. This dataset contains some healthy leaf images. For this analysis, the concept of SVM cuckoo search is considered to optimize the classification parameters [2]. Advantages of image processing techniques to detect and classify diseases in agricultural applications are helpful. The early proposal of the paper refers on tomato quality estimate with image processing. Presented an appraisal of the main attributes of tomato fruit quality for final consumers and overview the methods currently used to evaluate these attributes [3]. This article introduces an efficient approach to detect and identify unhealthy tomato leaves using Gray-Level Co-occurrence Matrix (GLCM) for detecting and identifying tomato leaf state is it healthy or infected. Support Vector Machine (SVM) algorithm with different kernel functions is used for classification phase. N-fold cross-validation technique is used to evaluate the performance of the presented approach [4]. The studies of the pomegranate plant diseases mean the studies of visually observable patterns seen on the plant. It is very difficult to monitor the pomegranate plant diseases manually. K-means clustering algorithm is used for segmentation and support vector machine is used for classification of disease [5]. In [6], the author introduced a technique to find out disease related to both leaf and fruit. In early proposed work, developed k-means clustering algorithm with multi SVM algorithm in MATLAB. The present study deals with automatic disease detection of plant leaf of Phaseolus vulgaris and Camellia assamica using image processing techniques. It involves image acquisition, image preprocessing, image segmentation, feature extraction and classification [7]. The paper mainly focused on plant disease in different methods in image processing. SVM classifier is given the better result to detect the diseased plant [8]. The earlier paper performed k-means clustering technique, segmentation, multiple SVM classifiers and CLBP shows more accuracy on diseased apple [9]. In the first phase the author create a color transformation structure for the RGB leaf image and then, we apply color space transformation for the color transformation structure. Next, in the second phase, the images are segmented using the K-means clustering technique. In the third phase, calculate the texture features for the segmented infected objects. Finally, in the fourth phase the extracted features are passed through a pre-trained neural network [10].

3. PROPOSED METHODOLOGY

The proposed work contains three stages to detect the diseased region of the plant. In this paper had improved the accuracy level using F-Measure. SVM is one of the traditional machine learning algorithms used for the purpose of classification.

- Low-level processing: pre-processing steps such as reading an image, contrast enhancement.
- Inter-mediate level processing: K-means clustering used for Segmentation of diseased region state of the blueberry leaf.
- High level processing: Feature extraction and SVM classifier for detection and classification of leaf state. The Architecture of proposed system,
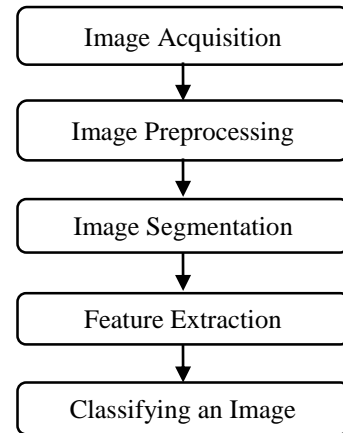

Fig.1. Architecture of proposed system

3.1 IMAGE ACQUISITION

Image Acquisition is retrieving an image from some source by using hardware devices like digital camera, mobile phone, etc. The first stage of any vision system is the image acquisition stage. After the image has been captured, various methods of processing can be applied to the image to perform some task on the image. In this step the sample images are collected, which are required to train the system. Blueberry leaf images have been taken by using digital camera and are used for both training and testing the system. The standard jpg format is used for classification.

3.2 IMAGE PREPROCESSING

Preprocessing is the step to move forward to improve the quality of the image and perform some operation on query image of blueberry leaf. Read the image from required path using choice menu and resize the image with 800×500 pixel.

3.2.1 Image Enhancement:

The min-max contrast stretched algorithm is used for better visualization. The linear contrast stretched transformation given below,

$$DN_{st} = 255 \times [(DN\text{-}DN_{min}) / (DN_{max} - DN_{min})] \tag{1}$$

where,

DN is the Digital Number of the pixel

DN_{st} is the corresponding DN in the enhanced output image

DN_{max} and DN_{min} are the maximum and minimum DN values in the original image.

3.2.2 Segmentation:

Image Segmentation is the process of classifying a image into different groups. K-means clustering is used to group the objects in an image. The result of k-means clustering algorithm is used for segmentation which gives the diseased region (septoria leaf spot) alone in total area of the blueberry leaf. The very general representation of the Euclidian distance formula is as follows:

$$L^2 = \sum_{i=1}^{n} (x_i - y_i)^2 \tag{2}$$

where,

L is the value of Euclidian distance $(x_1;x_2)$ and $(y_1;y_2)$,

(x,y) - is the two pixel points.

$$L = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \qquad (3)$$

If the data points are closest to the cluster it will stop and move to next data points if it is not nearest to the cluster it will continue the process until it pass through entire data points.

3.2.3 *Conversion of Gray Scale and Binary Image:*

The segmented image is converted into gray scale as well as binary image for feature extraction. Gray scale is the image will occupy the colors with shade of gray. The grayscale intensity value is stored as an 8-bit integer, 256 possible different shades of gray from black to white. Binary image is only having two possible values of every pixel either 0 or 1, Black or White. That pixel will be store in a single bit.

3.3 FEATURE EXTRACTION

Gray Level Co-Occurrence Matrix (GLCM) has proved to be a best statistical method of extracting feature from images. It is otherwise called as 'gray-level spatial dependence matrix'. The important types of feature extractions are shape, color, texture. Based on these types the statistical value will be change for every input image. Texture feature measuring the pixel values using statistical methods of query image as follows,

- *Contrast* - Evaluating local variations in the gray-level co-occurrence matrix.
- *Correlation* - Evaluating joint probability occurrence of the specified pixel pairs.
- *Energy* - Sum of squared elements in the GLCM
- *Homogeneity* - Measures nearest distribution of elements in the GLCM.

$$\text{Energy} = \sum_{i,j=0}^{N-1} \left(P_{ij}\right)^2$$

$$\text{Contrast} = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2 \qquad (4)$$

$$\text{Correlation} = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2} \qquad (5)$$

$$\text{Entropy} = \sum_{i,j=0}^{N-1} -\ln\left(P_{ij}\right) P_{ij} \qquad (6)$$

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2} \qquad (7)$$

$P_{ij} = i,j$ is the elements of GLCM

N = number of gray levels in the image

μ = Mean of GLCM

σ^2 = variance of intensities of reference pixel.

Table.1. GLCM value of segmented Septoria leaf spot of Blueberry

GLCM Feature Extraction	Value
Contrast	0.0427
Correlation	0.8072
Energy	0.9649
Entropy	0.2146
Homogeneity	0.9955
IDM	255
Kurtosis	84.9643
Mean	1.4509
RMS	1.3526
Skewness	8.8574
Smoothness	1
Standard Deviation	12.1444
Variance	1.437583556998332e+02

3.4 CLASSIFICATION PHASE

The last step of this work is the Classification Phase. In this approach SVM Classifier had been used on Septoria leaf spot of Blueberry and also in Healthy leaf with the statement of 'Healthy' or 'Diseased'. The dataset has been divided into two phase for training and testing process to evaluate the septoria spot on blueberry leaf. The process of this stage is training the dataset feature vectors and their corresponding classes, whereas the output is the decision that will determine the type of input image whether it's healthy or infected. To achieve good results, SVM was trained and tested using RBF kernel functions in matlab.

In this paper the simplest algorithm has been implemented for classification and identification on Blueberry leaf images. The below methodologies are the stepwise process which used to identify a leaf state whether it is healthy or infected. There are many earlier proposed papers have been explained a different algorithms and methods to identify a diseased leaf. The algorithm shown below.

Algorithm 1: Classification of Blueberry Leaf Images

Start

Step 1: Read the input image of blueberry leaf using Matlab R 2016a from the folder where it is located.

Step 2: *Preprocessing*: improve the contrast level from original image by using contrast stretched min-max algorithm

Step 3: Convert it into grayscale and binary image.

Step 4: Segmenting the image using k-means clustering algorithm.

Step 5: Features are extracted from segmented image using GLCM,

stats = graycoprops(glcms, 'Contrast Correlation Energy Homogeneity');

Step 6: Dataset of 40 images are taken for evaluation of training and testing equally.

Step 7: Used SVM classifier for classify and identify a leaf state.

Step 8: Match the original image with database whether it is healthy or infected.

Step 9: Set label as 0 and 1, then

    a. Apply state

    b. if (result == 1)

        i. helpdlg (' Septoria Spot Occurred On Leaf ')

        ii. disp (' Disease Detect ')

    c. else

        i. helpdlg ('Healthy Leaf ')

    d. disp ('Disease not Detect')

    e. end

    f. if (choice==3)

        i. close all

    g. return

    h. end

Step 10: The output result will be displayed.

Stop

The classification is done under the followed steps,

Step 1: Load all the features- load ('Training_Data.mat')

Step 2: Insert the test features (feat_disease) into a variable 'test'

Step 3: Used svm.m file for Train_Feat. Train_Label and test to declare the result

Step 4: Visualize the result if the result is 1 it is 'defected' else it's 'healthy'.

4. RESULTS AND DISCUSSION

The aim of this proposed work is to detect and identify if the Septoria spot is affected the blueberry leaf or not. In this article, number of 40 blueberry leaf images are used which are 20 healthy and 20 infected leaves. The segmentation will be done using input image RGB of where the user selects the path of the file in a system. GLCM (Gray Level Co-occurrence Matrix) is calculating the statistical methods of feature extracted region of segmented image. A Support Vector Machine (SVM) is a discriminative classifier that defined by a separating hyperplane. The first step is choosing a query image where it is located in a system. K-means clustering is used for segmentation to detect the diseased region alone. Segmented image converted into gray scale and binary image for calculating GLCM. This proposed article is refers the Septoria spot of the leaf and healthy leaf where the output shown below. The paper can successfully detect and classify the examined disease with accuracy of 96.77%.

Table.2. Details about Dataset of Blueberry Leaves used in proposed work

Leaf Type	No of Images	Resolution	Pixel value	Image Type
Healthy Leaf	20	800×500×3	uint8	JPG
Disease Leaf	20	800×500×3	uint8	JPG

Table.3. Support Vector Machine

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Tested_Negative	0.962	0.456	0.805	0.901	0.918	0.890
Tested_Positive	0.786	0.089	0.75	0.921	0.891	0.890
Weighted Avg.	0.979	0.332	0.951	0.903	0.949	0.890

The Table.3 referred about the result of SVM values. It took 0.41sec to build and classified with 96.77%.

- *Precision*: Precision is that what fraction of those predicted positive are actually positive.

$$Precision = TP / (TP + FN) \qquad (8)$$

where TP is True Positive and FN is False Negative

- *Recall*: Recall is the True Positive rate and it is also referred as sensitivity.

$$Recall = TP / (TP + FN) \qquad (9)$$

- *F-Measure*: It is a measure that the combination of precision and recall. It is called as measure because of weight is calculated equally in both precision and recall.

$$F\text{-}measure = 2 \times [(precision \times recall) / (precision + recall)] \quad (10)$$
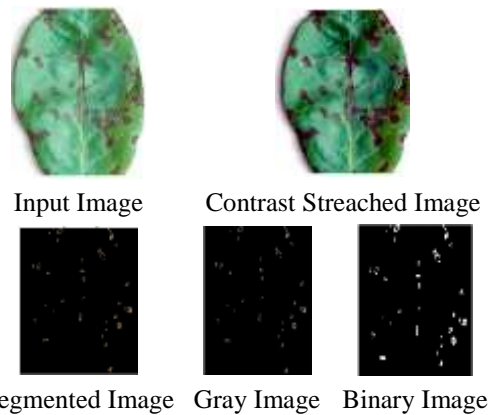
Select as input blueberry leaf

Input Image   Contrast Streached Image

Segmented Image   Gray Image   Binary Image

Fig.2. Experimental Result of Septoria spot on Affected Leaf

Input Image   Contrast streached image   Binary Image

Fig.3. Output of Healthy Leaf

5. CONCLUSIONS

This paper refers identifying and detecting the diseased plant of blueberry leaf where it is very difficult to monitor often. If the farmer gets delay to detect the plant it gradually ruin the whole plant from top to bottom. Blueberry fruit has contained the high nutrition properties. These plants mainly affected by various disease especially Septoria spot is the fast growing flu which affects the stem and leaf of the plant. Image processing has vital role to save the plant at initial stage. GLCM-Grey Level Co-occurrence Matrix has been used for the purpose of feature extraction on segmented leaf and SVM-Support Vector Machine Classifier is a machine learning process to classify the leaf is infected or not. F-measures are calculated using Precision, Recall and f1 score for calculating accuracy. The average accuracy of detect the Septoria spot on blueberry leaf is 96.77%.

REFERENCES

[1] S.R. Dubey and A.S. Jalal, "Detection and Classification of Apple Fruit Disease using Complete Local Binary Patterens", *Proceedings of 3rd International Conference on Computer and Communication Technology*, pp. 1-5, 2012.

[2] Abraham Gastelum Barrios, Rafael A. Borquez-Lopez, Enrique Rico-Garcia, Manuel Toledano-Ayala and Genaro M. Soto-Zarazua, "Tomato Quality Evaluation with Image Processing: A Review", *African Journal of Agricultural Research*, Vol. 6, No. 14, pp. 1-5, 2011.

[3] M.K. Chetia and S. Chetia, "Detection and Classification of Plant Leaf Diseases in Image Processing using Matlab", *International Journal of Life Sciences Research*, Vol. 5, No. 4, pp. 120-124, 2017.

[4] V.M. Gupta, "Plant Leaf Disease Analysis using Image Processing Techniques with Modified SVM-Classifier", *Indian Journal of Endocrinology and Metabolism*, Vol. 5, No. 1, pp. 1-8, 2017.

[5] Usama Mokhtar, Nashwa El Bendary, Aboul Ella Hassenian, E. Emary, Mahmoud A. Mahmoud, Hesham Hefny and Mohamed F. Tolba, "SVM-based Detection of Tomato Leaves Diseases", *Proceedings of International Conference on Advances in Intelligent Systems and Computing*, pp. 641-652, 2015.

[6] Niket Amoda, Bharath Jadhav and Smeetha Nailkwadi, "Detection and Classification of Plant Disease by Image Processing", *International Journal of Innovative Science, Engineering and Technology*, Vol. 1, No. 2, pp. 70-74, 2014.

[7] P. Padol and A.A. Yadhav, "SVM Classifier based Grape Leaf Disease Detection", *Proceedings of International Conference on Advances in Signal Processing*, pp. 1-10, 2016.

[8] V. Pooja, R. Das and V. Kanchana, "Identification of Plant Leaf Disease using Image Processing Techniques", *Proceedings of International Conference on Technological Innovations in ICT for Agriculture and Rural Development*, pp. 1-5, 2017.

[9] Sandesh Raut and Amit Fuldsunge, "Plant Disease Detection in Image Processing using Matlab", *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 6, No. 6, pp. 10373-10381, 2017.

[10] Shivaputra S.Panchal and R. Sonar, "Pomegranate Leaf Disease Detection using Support Vector Machine", *International Journal of Engineering and Computer Science*, Vol. 5, No. 6, pp. 1-4, 2016.

[11] S.R. Dubey and A.S. Jalal, "Detection and Classification of Apple Fruit Diseases using Complete Local Binary Patterns", *Proceedings of 3rd International Conference on Computer and Communication Technology*, pp. 1-6, 2012.

[12] R. Swarna Lakshmi, B. Kanchanadevi, "A Review on Fruit Grading Systems for Quality Inspection", *International Journal of Computer Science and Mobile Computing*, Vol. 3, No. 7, pp. 12-18, 2014.

[13] Dameshwari Sahu and Chitesh Dewangan, "Identification and Classification of Mango Fruits using Image Processing", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Vol. 2, No. 2, pp. 203-210, 2016.