Grape Leaf Disease Identification using Machine Learning Techniques

S.M. Jaisakthi
SCOPE
Vellore Institute of Technology
Vellore, India
jaisakthi.murugaiyan@vit.ac.in

P. Mirunalini
Department of CSE
SSN College of Engineering
Chennai, India
miruna@ssn.edu.in

D. Thenmozhi
Department of CSE
SSN College of Engineering
Chennai, India
theni_d@ssn.edu.in

Vatsala
Barclays
Eon IT Park, Kharadi
Pune, India
vatsala3012@gmail.com

***Abstract*—Having diseases is quite natural in crops due to changing climatic and environmental conditions. Diseases affect the growth and produce of the crops and often difficult to control. To ensure good quality and high production, it is necessary to have accurate disease diagnosis and control actions to prevent them in time. Grape which is widely grown crop in India and it may be affected by different types of diseases on leaf, stem and fruit. Leaf diseases which are the early symptoms caused due to fungi, bacteria and virus. So, there is a need to have an automatic system that can be used to detect the type of diseases and to take appropriate actions. We have proposed an automatic system for detecting the diseases in the grape vines using image processing and machine learning technique. The system segments the leaf (Region of Interest) from the background image using grab cut segmentation method. From the segmented leaf part the diseased region is fruther segmented based on two different methods such as global thresholding and using semi-supervised technique. The features are extracted from the segmented diseased part and it has been classified as healthy, rot, esca,and leaf blight using different machine learning techniques such as Support Vector Machine (SVM), adaboost and Random Forest tree. Using SVM we have obtained a better testing accuracy of 93%.**

***Index Terms*—Grape Leaves, Disease Identification, Machine learning, SVM**

I. Introduction

Indian Economy is highly dependent on agricultural productivity of the country. Grape is very commercial fruit of India. It can easily be grown in all tropical, sub-tropical and temperate climatic regions. India has got different types of climate and soil in different parts of the country. This makes grapevines a major vegetative propagated crop with high socio-economic importance. The grape plant will cause poor yield and growth when affected by diseases. The diseases are due to the viral, bacteria and fungi infections which are caused by insects,rust and nematodes etc., These diseases are judged by the farmers through their experience or with the help of experts through naked eye observation which is not accurate and time consuming process. Early detection of disease is then very much needed in the agriculture and horticulture field to increase the yield of the crops. We have proposed a system that can detect and identify diseases in the leaves of the grape plants.

II. Literature Survey

Web enabled disease detection system have been proposed in [8]. The system proposed a segmentation method which has used mean based strategy for computing threshold and textual features were extracted and classification was done by SVM. The survey proposed by Vijai et al. in [12], discuses about different disease classification techniques used for plant leaf disease and used genetic algorithm for image segmentation. An integrated approach of particle swarm optimization and SVM for plant leaf disease detection and classification was proposed in [10]. Disease detection system for pomegranate leaves was proposed in [5] which used colour-based segmentation and features like color, morphology and texture for classifying the leaves. Agrawal et al. proposed a leaf detection and climatic parametric monitoring of plants using IOT in [1]. Neural Network based classification was proposed in [9] for detecting plant leaf diseases based on the texture features extracted using GLCM matrix. Mokhtar et al. proposed SVM based classification by extracting the texture based features in [6]. SVM classifier with different kernel functions including Cauchy kernel, Invmult Kernel and Laplacian Kernel were employed to evaluate the ability of the approach to detect and identify the infected tomato leaf. Leaf detection system for pomegranate leaves was proposed in [13] which uses K-means for segmentation and statistical features for classification using SVM. Sabrol et al. have proposed a system for leaf disease classification using decision tree by extracting different features after segmenting the leaf using ostu thresholding [4]. A system for two type of disease classification such as Downy mildew and Powdery mildew in grape leaves was proposed in [11] using Back propogation Neural Network. A fast system was proposed for disease detection and classification using Neural Network after extracting the texture features using gray level co-occurrence methodology in [2]. A smartphone based system was developed by Mwebaze et al. in [7] using machine learing technique to detect the state of the disease of the plant and also the severity levels of each diseases. Machine learning based techniques such as decision tree, Navie Bayes theorem, Neural Network, K-Means and Random forest algorithms were proposed for leaf disease classification in [3] using the features such as size, shape, dryness, wilting. Most of the work in the

978-1-5386-9471-8/19/$31.00 2019 IEEE

literature uses K-means segmentation for segmenting the leaf and extract low level features of the image to classify the plant leaf diseases. We have proposed a system which uses global features to classify the plant diseases and segmented the region of interest using graph cut method. We have also compared the results obtained using different machine learning techniques.

III. Proposed Methodology

We have proposed an automated disease detection and classification system for grape leaves using traditional image processing and machine learning techniques. The proposed system first segments the ROI from the back ground using grab cut algorithm and classify the segmented leaves as healthy, balck-rot, esca and leaf blight.Figure. 1 depicts different types of disease in grape leaves.

These diseases are caused due to fungi infection on the leaves. Each disease have different characteristics where black rot appears to be circular in shape and has dark margins, esca appears as dark red stripes and leaf blight appears to be solid reddish-purple spots. The proposed system consists of five different process such as image preprocessing, image segmentation, feature extraction, disease detection and identification. The overall flow of the proposed system is depicted in the Figure. 2

A. Image Preprocessing

The images are acquired from the web and are from different sources and sizes. The images also contains noise due to bad lightening condition, weather occlusion etc. To reduce the computational complexity the images are scaled down to a standard width and height. This scaled image are then processed to filter the noise using Gaussian filter. The Gaussian blur is a low pass filter that reduces the high frequency components, we have used 5*5 kernel size to filter the noise

B. Image Segmentation

From the preprocessed image, the leaf part of the image is segmented from the background image Grabcut segmentation algorithm. This algorithm label a pixel as foreground or background using Gaussian Mixture Model (GMM) and also takes initial rectangle which is a rough segmentation between background and foreground. We have used a rectangle of dimension (10, 10, w-30 and h-20) as the bounding box where w and h are width and height of the image. The results of the Grab-cut method is depicted in Figure. 3.

From extracted foreground i.e. the leaf part, the diseased parts are extracted. The disease part contains lesions, coloured spots and some yellowish part of the leaf. For extracting diseased region from the leaves we have proposed two different methods.

1) Diseased Part Identification- Global Thresholding: In this method, the RGB image is converted into grey scale image and then global thresholding is applied to convert the image into a binary image. On the thresholded image, connected component labelling is applied to find the contours. The contour with the largest area is then identified and morpological operations such as dilation and erosion is applied. The original image is converted to an HSV image and in the h channel, thresholding is applied. Binary AND operator is then applied to contour detected image and the HSV image. The resultant image was again thresholded using binary invert thresholding. The output of this method is shown in Figure.4

2) Diseased Part Identification - Semisupervised Learning: The diseased part of the leaves generally appears in blue color in BGR image. To segment the diseased part blue color pixels are filtered out by converting the RGB image into BGR image. To filter blue color pixels we have used the training image to find the lower and upper boundary of blue color pixels. The pixels which lies within lower and upper boundary is then filtered as blue pixels from the input image. In the filtered image thresholding is applied and finally the diseased areas are identified. The output of this method is depicted in Figure.5

3) Feature Extraction: Image features provide rich information about the content of the image. These features represent certain distinctive characteristics that can be used for differentiating among the categories of input patterns. In this work, we have used texture and colour features of the images for classification.

The texture of an image is usually expressed by contrast, uniformity, entropy etc. A statistical method of examining the texture of image is Gray-level Co-occurrence matrix (GLCM). GLCM extracts second order statistical texture features from the training images. GLCM an (NXN) matrix where N is the number of grey levels in the image. Before extracting GLCM features, wavelet decomposition of an image is done. The Discrete Wavelet Transform decomposes an image into different sub-band images such as LL (low-low), LH (low-high), HL (high-low), and HH (high-high). LL image is used for texture analysis as it contains the highest quantity of information. In a GLCM matrix M, if i represents a row and j represents a column, the element M (i, j) represents the count of a pair of two pixels having i and j as their intensities that are separated by a distance d and lies along a direction specified by an angle theta. The algorithm takes d and theta values as input by the user. In this system, we have extracted six texture features namely, energy (randomness of intensity distribution), homogeneity (closeness of distribution of GLCM element to the diagonal of GLCM), contrast (amount of local variations), dissimilarity, correlation (image linearity) and angular second moment in combination with different values of d and theta. The values of d and theta are chosen as (1, 2, 3) and (0, 45, 90, 135) respectively. So the total number of features are (6 x 3 x 4) = 72 texture features for an image.

After segmentation of disease affected portion of the leaf, colour features were examined to find whether the given input leaf is healthy or not. If the input leaf image is healthy leaf then the segmented image will not have any region and it contains only black pixels.For healthy leaves image colour feature vector contains only 0s whereas for non-healthy leaves feature vector contains some color information.

978-1-5386-9471-8/19/$31.00 2019 IEEE

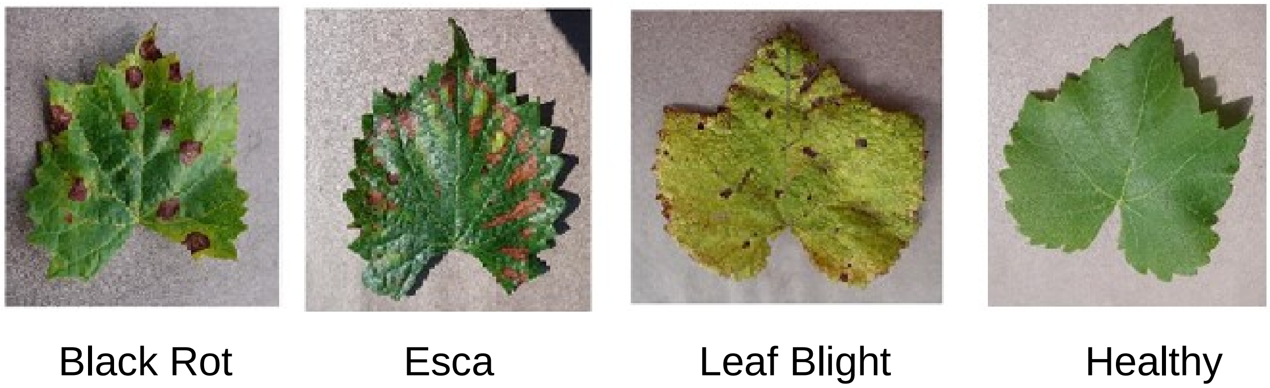

Fig. 1. Different types of disease in Grape Leaves

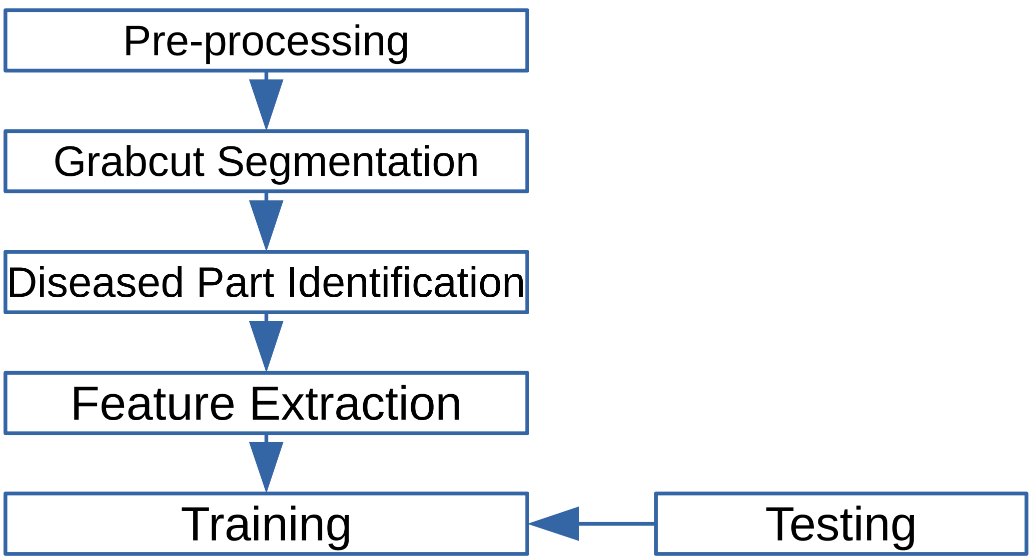

Fig. 2. Architecture of the Proposed System

4) Classification using Different Classifiers: The extracted feature vectors are then used to train different classifiers and the results were analysed.

Support Vector Machine: Given a labelled training data, SVM outputs an optimal separating hyperplane. This hyperplane categorizes new data point into classes. In order to improve the accuracy of SVM, some parameters of the SVM classifier needs to be tuned. One of the parameters is kernel which defines whether separation should be linear or nonlinear. Another parameter is regularization which defines the extent to which misclassification of a training sample needs to be avoided. Linear kernel and regularization parameter with value 1000 is used in this system. A larger value of regularization chooses small margin of hyperplane if it ensures minimum misclassification of training examples.

Random Forest: Random Forest is an ensemble learning method and also a supervised learning algorithm. It builds a forest of decision trees. Many trees fit into a random forest classifier. The extracted feature vector is passed as input vector to each tree of the forest and a decision rule is obtained. In other words, the trees vote for a class. The class having the majority votes by the trees is chosen by the forest.

AdaBoost: AdaBoost is used to boost the performance of a machine learning algorithm. It is used with weak learners. Weak models are trained using weighted training data where each instance is weighted. After training the model, the misclassification rate is calculated. This error rate is modified further and used to update the weights of training instances. The purpose of weight updating is to give more weight to misclassified instances. This process continues until there is no

978-1-5386-9471-8/19/$31.00 2019 IEEE

Fig. 3. Result of applying Grabcut segmentation algorithm for a sample image

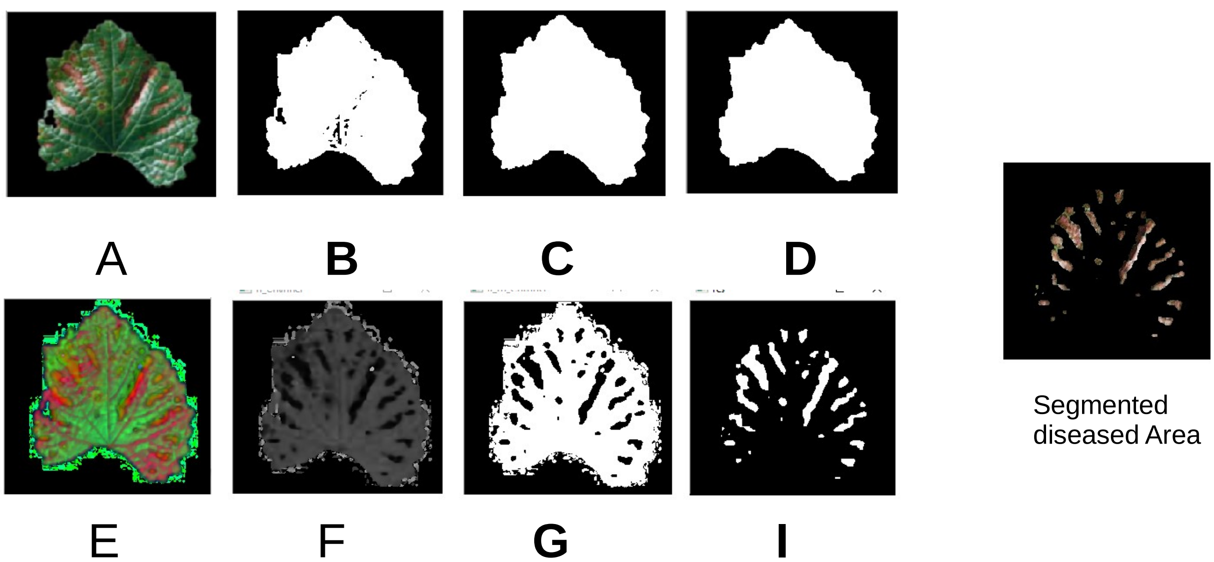

Fig. 4. Disease Part Identification using Global Thresholding. A. Input Image B. Thresholded Image C. Dilation D. Erosion E.HSV Image F. h channel Image G. Thresholded Image F. Binary AND Between D and G

scope for improvement. In this way, the AdaBoost algorithm improves the learning of weak learners.

C. Result And Discussion

We have evaluated the proposed system using 5675 grape leaves which have been downloaded from the plant village website and also from web. We have used 80%of the images for training and others for testing.

The global thresholding method used for segmenting leaf disease part was found to be more suitable for training the model as it segments the precise diseased part of the leaves which leads to improved classification results. Training accuracy obtained using different machine learning techniques are summarized in Table. I.

From the results it is evident that good training accuracy is achieved when the features are extracted from the diseased part of the images using global thresholding and trained using SVM classifier with tuned parameters since SVM performs good when the data is highly non-linear. The overall accuracy obtained is 93.035% for 1135 test images.

D. Conclusion

In this paper, we propose an automatic leaf recognition system that identify diseases in grape leaves using machine

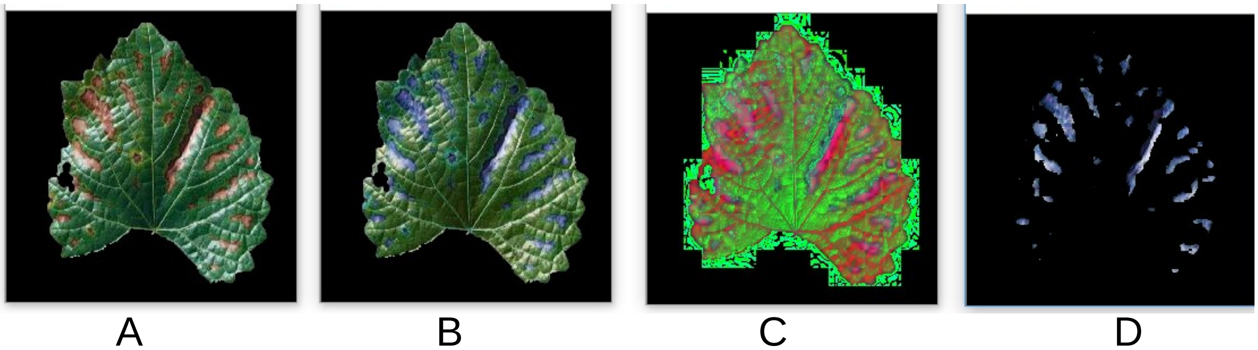

Fig. 5. Disease Part Identification- Semisupervised learning. A. Input Image B. BGR Image C. HSV Image D. Filtered Blue pixels

TABLE I
TRAINING ACCURACY OBTAINED USING PROPOSED METHODS

Classifier	Global Thresholding	Semisupervised Method
SVM	**91%**	86.57%
Adaboost (Decision Tree)	83%	76%
Random Forest Tree	74.79%	74%

TABLE II
CLASSIFICATION RESULTS FOR TEST IMAGES USING SVM

Class	No. Of Images	Healthy	Black Rot	Esca	Leaf Blight	Accuracy
Healthy	423	407	16	-	-	96.2%
Black Rot	230	-	208	14	8	90.4%
Esca	262	-	24	236	2	90.08 %
Leaf Blight	220	-	9	1	210	95.45 %
					Average	**93.035 %**

learning technique. The proposed system first segments the leaf part from the background using grab cut segmentation technique.From the segmented leaves diseased region are identified using two different methods. The first method uses globlal thresholding technique whereas the second method using semisupervised learning technique. From the identified diseased part texture and color features are extracted and trained using different classifiers and the results are compared. We have used SVM, random forest and Adaboost algorithms for classification. We have achieved a better result of 93.035% as testing accuracy by using globlal thresholding and SVM.

ACKNOWLEDGMENT

The authors would like to thank the management of Vellore Institute of Technology, Vellore, India and SSN College of Engineering, Chennai India for funding the respective research labs where the research work is being carried out.

REFERENCES

[1] Agrawal, G.H., Galande, S.G., Londhe, S.R.: Leaf disease detection and climatic parameter monitoring of plants using iot **4**, 9927 – 9932 (2015)
[2] Al-hiary, H., Bani-ahmad, S., Reyalat, M., Braik, M., Alrahamneh, Z.: Fast and accurate detection and classification of plant diseases. International Journal of Computer Applications **17**(1) (2011)
[3] G.PremRishiKranth, HemaLalitha, LaharikaBasava, AnjaliMathurh: Plant disease prediction using machine learning algorithms. International Journal of Computer Applications **18**(2) (2018)
[4] H.Sabrol, K.Satish: Tomato plant disease classification in digital images using classification tree. In: 2016 International Conference on Communication and Signal Processing (ICCSP), pp. 1242–1246 (2016)
[5] Khot.S.T, Supriya, P., Gitanjali, M., Vidya, L.: Pomegranate disease detection using image processing techniques **5**(1), 2248 – 2251 (2016)
[6] Mokhtar, U., Ali, M.A.S., Hassenian, A.E., Hefny, H.: Tomato leaves diseases detection approach based on support vector machines. In: 2015 11th International Computer Engineering Conference (ICENCO), pp. 246–250 (2015)
[7] Mwebaze, E., Owomugisha, G.: Machine learning for plant disease incidence and severity measurements from leaf images. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 158–163 (2016)

978-1-5386-9471-8/19/$31.00 2019 IEEE

[8] Nandhini, A., Hemalatha, Radha, Indumathi: Web enabled plant disease detection system for agricultural applications using wmsn. Wireless Personal Communications **102**(2), 725–740 (2018)
[9] Nivedita.R.Kakade, Dnyaneswar.D.Ahire: Real time grape leaf disease detection **1**, 598 – 610 (2015)
[10] P, K., S, S., S, S.: Detection and classification of leaf diseases using integrated approach of support vector machine and particle swarm optimization. **4**(1), 79 – 83 (2017)
[11] Sannakki, S.S., Rajpurohit, V.S., Nargund, V.B., Kulkarni, P.: diagnosis and classification of grape leaf diseases using neural networks. In: 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1–5 (2013)
[12] Singh, V., Misra, A.: Detection of plant leaf diseases using image segmentation and soft computing techniques. Information Processing in Agriculture **4**(1), 41 – 49 (2017)
[13] Sonar, R., S.Panchal, S.: Pomegranate leaf disease detection using support vector machine. International Journal of Engineering and Computer Science **5**(6) (2016)