

Machine Learning Techniques For Fruit-Leaf Disease Detection By Image

¹Yalla Ramana Murty, ²S.Srinivas

^{1,2}Dept. of Computer Science & Engineering, KIET, Kakinada, AP, India

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 further 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.

Keywords

Leaf diseases, SVM, segmentation, morphological processing, features extraction, clustering, fuzzy logic. Back Propagation Neural Network, CCV, K-means Clustering, LBP, SVM

I. Introduction

Disease identification is critical to crop cultivation India is an agricultural nation, and agriculture employs the majority of the people. Farmers have a broad variety of Fruit and Vegetable crops to choose from. Technological assistance may help to enhance farming. Pathogens induce disease in plants under any environmental situation. In most instances, illnesses are visible on the plant's leaves, fruits, and stems; therefore, success. Pathogens, microorganisms, fungus, bacteria, viruses, and other microorganisms cause the majority of plant illnesses. Plant diseases are sometimes caused by an unfavorable environment, which includes soil and water.

There are many methods for detecting various kinds of illnesses in plants in their early stages. The traditional technique of plant disease detection is naked eye inspection, which is ineffective for big crops. Disease diagnosis in plants is more efficient, less time consuming, and accurate when using digital image processing methods. This method saves time, effort, labor, and insecticides. Various publications suggest various methods for accurate plant disease detection using digital image processing. Several image processing methods have been created by various academics. This article provides an overview of several kinds of Image processing methods are used to identify and classify different leaf illnesses as well as fruit diseases. Experts were formerly responsible for illness detection and identification. Farmers take additional measures to manage the illness based on professional observations. Because consulting specialists is expensive and time consuming in poor nations, automated illness detection methods were created. The agricultural sector is using technology solutions

to improve productivity. Productivity and quality of fruit Farmers' economic losses may be prevented by adopting this method. Automatic diagnosis of fruit disease is essential because it identifies signs on fruit at a very early stage, resulting in a lower economic loss per fruit. This would also assist growers in taking preventative steps for next year by monitoring current fruit conditions. Some diseases can affect other sections of the tree, such as the shoots and branches. Mango illnesses that are common include pectobacterium, agrobacterium, rhinocladium, anthracnose, powdery mildew, and red rust. Mango bacterial diseases include pectobacterium and agrobacterium. Rhinocladium is a fungus that causes mango sickness. Anthracnose is induced by humidity and changes in climatic circumstances, such as temperatures between 24 and 30 degrees Celsius and excessive rain. During blooming, powdery mildew is produced by rain or mist combined with a colder night. Alga is the cause of red rust. Because mangoes are uneven in form, a visual examination is insufficient to assess its conditions.

II. Literature Review

The authors of article [1] propose an image processing method for identifying rice illnesses and focus on the two most prevalent diseases in north east India, namely Leaf Blast (*Magnaporthe Grisea*) and Brown Spot (*Cochiobolus Miyabeanus*). The first step is to acquire an image, and then the author uses segmentation, boundary detection, and spot detection methods to extract features from the diseased regions of the leaf. The author of this article proposes a zooming method in which a SOM (Self Organising Map) neural network is utilized to classify sick rice pictures. In SOM, there are two ways to create an input vector. The first technique is zero padding, while the second way is the Missing point interpolation. The interpolation technique is used for fractional zooming to normalize the spot size. Image modification in the frequency domain does not result in improved categorization. Four distinct kinds of pictures are used for testing; the zooming algorithm produces acceptable categorization results for test images. The authors of article [2] propose an image-processing method for detecting leaf and stem illness. The author made use of a collection of leaf pictures from Jordan's Al-Ghor region. The five plants are The image processing method is used to evaluate illnesses such as early scorch, ashen mold, late scorch, cottony mold, and tiny whiteness. This approach begins with picture capture and then uses the K-Means clustering algorithm for segmentation. Following that, CCM (Colour Co-occurrence Method) is utilized in feature extraction for texture analysis of diseased leaf and stem. Finally, the article discusses Back.

In the categorization of plant diseases, a neural network propagation method is used. The results of this image processing method indicate that plant diseases may be detected and classified accurately with a high degree of accuracy (about 93 percent). The authors of article [3] utilized image processing tools LABVIEW and MATLAB to identify chili plant illness. This combination method identifies illness by looking at the leaves. Early stage inspection The image was taken using IMAQ LABVIEW Further image processing procedures are carried out using Vision and MATLAB. Fourier filtering, edge detection, and morphological procedures are examples of image pro-

processing activities. Color clustering is utilized in feature extraction to differentiate between chili and non-chili leaves. The healthiness of each chili plant is then determined using picture recognition and classification. This method minimizes the use of hazardous chemicals in the chili plant, lowering manufacturing costs and increasing chili quality. The authors of article [4] propose an image processing method for identifying *Malus Domestica* leaf illness. Histograms are used to calculate the intensity values of rayscale picturestechnique of equalization The Co-occurrence matrix technique algorithm is utilized for texture analysis in image segmentation, while the Kmeans clustering algorithm is used for color analysis. Texture analysis is the process of characterizing areas of a picture based on texture information. The term “color analysis” refers to the process of reducing the total of colorof distance squares between items and class centroid or matching cluster Individual pixel values are compared to threshold values in the threshold matching process; if the value is higher than the threshold, the pixel is designated as an object pixel. For the identification of plant diseases, the texture and color analysis pictures are compared to the prior photographs. In the future, the author plans to utilize Bayes and K-means clustering. The authors of article [5] offer image processing methodsfor detecting Bacterial infection in plants. Bacterial leaf scorch is an infection observed on plants, and early identification of this aids in the enhancement of plant development. Picture processing begins with image acquisition, which includes fundamental processes such as image capture and conversion to a computer-readable format. The K-means clustering technique in image segmentation is then used to separate the foreground and background images.Clustering is based on intensity mapping, and emphasizing the leaf region is accomplished by subtracting the clustered leaf pictures from the base images. In comparison to Fuzzy logic, the K-means clustering method is simple and efficient in identifying contaminated areas while requiring less human cluster selection. Further implementation is feasible using ADSP target boards and FPGA tools. The authors of article [6] propose an image processing method for detecting diseased regions of Citrus leaf. There are four of them.Citrus illnesses include (i) Citrus canker, (ii) Citrus wilt, and (iii) Citrus wilt (ii)Anthracnose, (iii) overwatering, and (iv) citrus greening are all symptoms of anthracnose. The author suggested a technique in which picture acquisition is the initial stage in creating a database by collecting images with a digital camera in high quality. Picture pre-processing is where color space conversion and image enhancement take place. For color picture improvement, the discrete cosine transform domain is employed. For color space conversion, the YCbCr color system and the $L^*a^*b^*$ color space are used. In the case of feature extraction. The author presents a statistical technique for calculating statistics such as contrast, energy, homogeneity, and entropy using the Gray-Level CoOccurrence Matrix (GLCM) and the graycoprops function. Citrus leaf diseases are differentiated using two kinds of support vector machine (SVM) classifiers: SVMRBF and SVMPOLY. The authors of article [7] propose an image processing method for detecting Orchid leaf disease. Sun sunburn and black leaf spotare the most common orchid leaf diseases. The first stage in image processing is image acquisition, which involves collecting pictures and storing them in a computer for later use. Picture pre-processing includes histogram equalization, intensity modification, and filtering in order to enhance or alter the image. In the boundary segmentation method, three morphological processes are employed to eliminate tiny objects while preserving big objects in the picture. Limiting inSegmentation is used to trace edges by defining the start and end points of a line. The ROI (area of interest) has been added to the GUI by the author. Following the boundary segmentation procedure, an image categorization is performed by calculating white

pixels. This method produces results with excellent precision and a low proportion of error. The authors of article [8] propose an image processing method for detecting illnesses in tomato leaves. In the acquisition of imagesDuring this phase, digital pictures of diseased tomato leaves are taken, including two kinds of tomato diseases: early blight and powdery mildew. Some methods are used in the pre-processing phase for picture improvement, smoothness, noise removal, image scaling, image isolation, and background removal. The author introduced the Gabor wavelet technique and the Support vector machine for tomato identification and classificationillnesses. During the feature extraction step, feature vectors are generated using the Gabor wavelet transform for the subsequent classification phase. During the classification step, a support vector machine (SVM) is trained to detect the kind of tomato disease. SVM’s inputs are feature vectors and associated classes, while its outputs are decision trees.it detects tomato leaf disease The Invmult Kernel, Cauchy Kernel, and Laplacian Kernel functions are used in SVM. Grid search and N-fold cross-validation methods are employed to evaluate performance. The authors of article [9] presented illness detection, in which image processing is the first step for getting an image in digital form and pre-processing to eliminate noise and other objects from the picture. In addition, pre-processing converts RGB pictures to greyscale images using the equation $f(x) = 0.2989*R + 0.5870*G + 0.114*B$ is used to equalize histograms. ImageFor identifying diseased parts of the leaf, segmentation is performed utilizing boundary and spot detection techniques. The K-means clustering technique is used to classify items. For thresholding, the Otsu threshold method is employed, which generates binary pictures from greyscale images. Color, texture, morphology, and edges are utilized in plant disease diagnosis with the aid of feature extraction. H & B extraction of leaf colorIn image processing, feature extraction techniques include components and the color co-occurrence method. Artificial neural networks (ANN) and back propagation networks are used to classify illnesses.

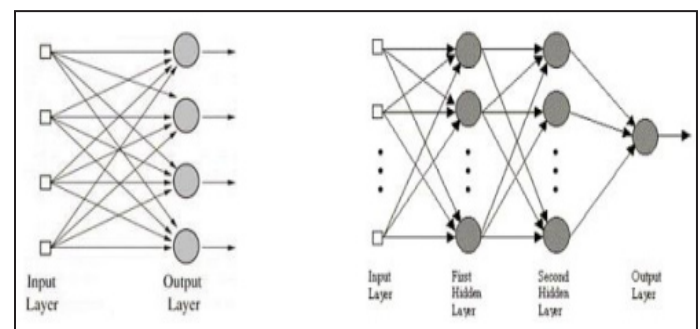


Figure 1: Basic Artificial Neural Network architectures

(Single layer on the left and multi-layer on the right)[11] The authors of article [10] propose an image processing method for detecting Scorch and Spot plant diseases. The first step is to capture an RGB picture of the plant. The color transformation structure is then constructed in pre-processing, and RGB color values are transformed to space. Green-pixel masking is performed after K-means clustering. This eliminates masked cells from inside infected cluster borders. Picture segmentation is used to extract usable parts from an image. Color, texture, and edge characteristics are extracted using color co-occurrence technique. Neural Networks are set up to recognize and classify illnesses. Future study will involve evaluating the disease status of citrus plants in an outdoor setting. The authors of article [11] propose an image processing method for detecting groundnut plant disease. The groundnut plant suffers from two main diseases: early leaf spot (*Cercospora*) and late leaf spot (*Cercosporidium personatum*). After collecting RGB pictures of leaves, they are transformed to HSV color images. Green-colored

pixels in a picture are discovered in order to decrease processing time. The cooccurrence matrices method is utilized in color and texture feature extraction analysis. There are two methods for analyzing texture pictures in texture feature extraction. The first technique is a structured approach, while the second is a statistical approach. In this work, the author used a statistical method. The back propagation method is used to classify and identify groundnut illnesses. There are two types of phases in back propagation: 1) weight update and propagation The authors diagnosed four distinct illnesses with 97 percent accuracy. The authors of the article [12] presented a plant disease identification method in which the first step is to build a color transformation structure for the RGB leaf picture and convert color values from RGB to the space defined in that structure. The picture is then segmented using the K-means method after applying color space modification. The unneeded portion is removed during the second step, which is known as green pixel masking. For example, the green region inside the leaf area is eliminated. The texture characteristics for the segmented infected item are calculated in the third step, and masked cells inside the borders of the infected cluster are removed. Infected clusters are transformed from RGB to HSI, and an SGDM matrix for H and S is produced. The GLCM function is used in the fourth phase to determine the features and texture statistics. Finally, for illness identification, the retrieved characteristics are fed into a pretrained neural network. The authors of article [13] propose an image processing method for identifying sugarcane leaf disease. Six authors were chosen. Brown Spot, Downy mildew, Sugarcane Mosaic, Red stripe, Red rot, and Downy Fungal are the diseases being tested. Image acquisition captures pictures at higher quality resolutions using formats such as TIF, PNG, JPEG, and BMP for image analysis. RGB pictures are used in pre-processing. The pictures are transformed to grayscale, and any undesirable data from the photographs is deleted. Segmentation identifies the healthy region of a given picture that includes green pixels as well as the possibly contaminated area. In feature extraction, three methods are used: linear SVM, nonlinear SVM, and multiclass SVM for the identification of illness.

III. System Design

- For detecting disease the image should go through the following steps: 1) Pre-processing
- 2) Segmentation
- 3) Feature extraction
- 4) Classification.

Image Pre-processing:

The goal of this phase is to remove noise and improve picture quality. The super resolution method may be used to transform a low-resolution picture into a high-resolution one. Noise may be removed using morphological techniques. Shadow removal and picture correction require pre-processing. Shadow reduction is critical because shadows may interfere with segmentation and feature extraction.

Image Segmentation:

Image partitioning between foreground and background, or identifying the area of interest using certain algorithms, or creating clusters of regions by analyzing the similarities between neighboring pixels.

Thresholding

When the contrast between background and foreground is clearly discernible, this method is employed. In global threshold, a benchmark is set below which all pixels are marked '0' and considered to be in the background, and beyond which all pixels are marked '1' and considered to be in the forefront. The 'Otsu threshold' is a thresholding method in which the threshold value is determined by the infraclass variance pixels.

Region Growing:

In the area growth method, a seed is selected depending on the characteristics of the neighboring pixels, after which the expansion of the region of interest begins. Seed selection is an important procedure in this method. The expansion of R.O.I is dependent on the similarity of neighboring pixels. k-mean Clustering:

In this method, k denotes the number of clusters. It is based on Real-World Observations. K-mean considers each observation to be an object. This k-mean clustering finds a partition in which distinct items in the cluster are closer to one other while objects in other clusters are further apart. The iterative method is used to reduce the sum of distances between each item and its clustered centroid.

Fuzzy C-mean Clustering:

Iterative algorithms are employed in c-mean clustering. Initially, fuzzy cluster centers are calculated, and a fuzzy partition matrix is created. The goal function is decreased in the following iteration to determine the best location for the cluster.

When the maximum number of iterations is reached, the iteration process is automatically terminated.

Feature extraction

It extracts relevant information from the input image in order to minimize the extent of sources required to define a dataset.

Color: A color feature is one of the visual features as it is very stable. RGB and HIS colour system is mainly used. Infected part of a fruit can be easily identified by colour feature. Color image processing is categorized into three principle areas: Color transformation, Spatial processing of individual colour planes, Colour vector processing.

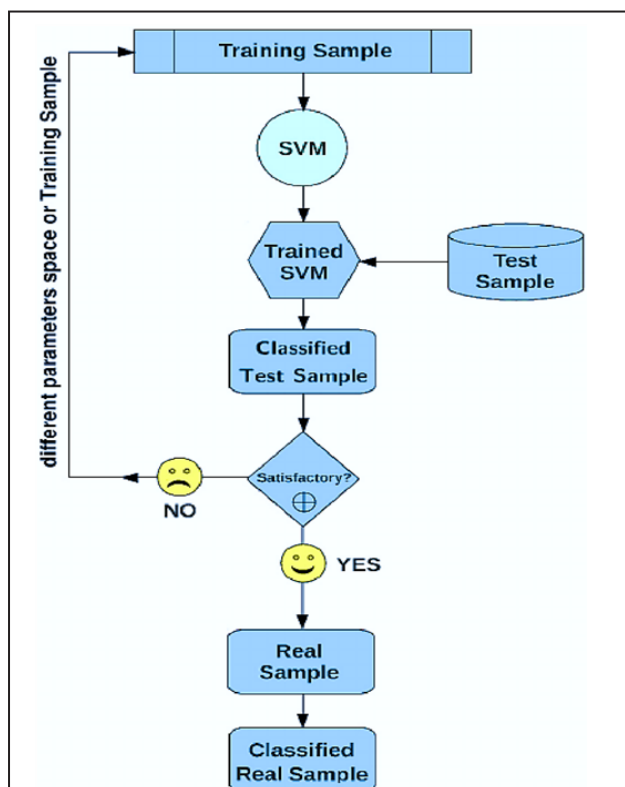


Figure 2. Schematic representation of the SVM algorithm classification process.

Texture:

Image texture provides information about spatial arrangements of colour of an image. When fruit suffer from any disease its texture feature change. So, we can analyze disease type by using texture feature.

Shape:

Morphology is tool used for extracting image components. By using morphological operations, infected part shape can be extracted from healthy fruit and leaf. Erosion operation can be used for obtaining boundaries of the image. Four major characteristics:

1. Geometric characteristics – it includes perimeter, area, axis, orientation angle and so on.
2. Area description features – depending on the target area its feature is described using a set of characteristics.
3. Moment invariants: In these, geometric characteristics such as Hu invariant moments, orthogonal moments, etc. are described by using its feature.
4. Fourier shape descriptor.

Classification

Support Vector Machine (SVM) making an N-dimensional hyper plane which is optimally partitions the data into different parts. Support vector models are related closely to neural networks. SVM evaluates more relevant information in a convenient way.

IV. Conclusions

This article provides an overview of image processing methods for detecting and classifying leaf diseases. For accurate illness diagnosis, various authors utilized different methods. The use of image processing methods has the advantage of detecting leaf diseases at an early stage. Most researchers used artificial neural networks and classifiers such as ANN, SVM, and others to improve recognition rates. All of the techniques in this article save time and provide effective results.

We examined the suggested method based on fuzzification in Fruit Detection System, in which fuzzy curves and fuzzy surfaces quickly identify the feature for pattern recognition system. Fuzzy surfaces eliminate features that are dependent on other significant features. It had shown to be effective in feature extraction. The K-mean clustering method has shown to be one of the most effective segmentation algorithms. SVM has proven to be the best classification algorithm because it maps input data with high accuracy. Linear or non-linear mapping methods in dimensional feature space. When the picture quality or resolution is poor, the Intent Search Technique has shown to be the greatest method for increasing image quality.

References

- [1] K. A. Garrett, S. P. Dendy, E. E. Frank, M. N. Rouse, and S. E. Travers, "Climate change effects on plant disease: genomes to ecosystems," *Annual Review of Phytopathology*, vol. 44, pp. 489–509, 2006.
- [2] S. M. Coakley, H. Scherm, and S. Chakraborty, "Climate change and plant disease management," *Annual Review of Phytopathology*, vol. 37, no. 1, pp. 399–426, 1999.
- [3] S. Chakraborty, A. V. Tiedemann, and P. S. Teng, "Climate change: potential impact on plant diseases," *Environmental Pollution*, vol. 108, no. 3, pp. 317–326, 2000.
- [4] A. J. Tatem, D. J. Rogers, and S. I. Hay, "Global transport networks and infectious disease spread," *Advances in Parasitology*, vol. 62, pp. 293–343, 2006.

- [5] J. R. Rohr, T. R. Raffel, J. M. Romansic, H. McCallum, and P. J. Hudson, "Evaluating the links between climate, disease spread, and amphibian declines," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 45, pp. 17436–17441, 2008.
- [6] T. Van der Zwet, "Present worldwide distribution of fire blight," in *Proceedings of the 9th International Workshop on Fire Blight*, vol. 590, Napier, New Zealand, October 2001.
- [7] S. A. Miller, F. D. Beed, and C. L. Harmon, "Plant disease diagnostic capabilities and networks," *Annual Review of Phytopathology*, vol. 47, pp. 15–38, 2009.
- [8] M. B. Riley, M. R. Williamson, and O. Maloy, "Plant disease diagnosis. The Plant Health Instructor," 2002.
- [9] J. G. Arnal Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," *SpringerPlus*, vol. 2, article 660, pp. 1–12, 2013.
- [10] H. Cartwright, Ed., *Artificial Neural Networks*, Humana Press, 2015.
- [11] I. Steinwart and A. Christmann, *Support Vector Machines*, Springer Science & Business Media, New York, NY, USA, 2008.
- [12] S. Sankaran, A. Mishra, R. Ehsani, and C. Davis, "A review of advanced techniques for detecting plant diseases," *Computers and Electronics in Agriculture*, vol. 72, no. 1, pp. 1–13, 2010.
- [13] P. R. Reddy, S. N. Divya, and R. Vijayalakshmi, "Plant disease detection technique tool—a theoretical approach," *International Journal of Innovative Technology and Research*, pp. 91–93, 2015.