Plant Disease Detection Based on Region-Based Segmentation and KNN Classifier

Chapter · January 2019

DOI: 10.1007/978-3-030-00665-5_154

Harpreet Kaur
Chandigarh University

Plant Disease Detection Based on Region-Based Segmentation and KNN Classifier

Jaskaran Singh and Harpreet Kaur

Abstract The plant disease detection is the technique which can detect disease from the plant leaves. The plant disease detection has various steps which are textural feature analysis, segmentation, and classification. This research paper is based on the plant disease detection using the KNN classifier with GLCM algorithm. In the proposed method, the image is taken as input which is preprocessed, GLCM algorithm is applied for the textural feature analysis, k-means clustering is applied for the region-based segmentation, and KNN classifier is applied for the disease prediction. The proposed technique is implemented in MATLAB and simulation results show up to 97% accuracy.

1 Introduction

Agricultural area is the only area through which the food requirements of complete human race are being served. In India, around 70% of the total population relies on agriculture and grow several kinds of fruits and vegetable crops within their fields. However, there is a need for high technicality while cultivating crops that are of optimum yield and quality. It is thus important to diagnose the diseases within plants [1]. The identification of diseases is a difficult process due to which the farmers face lots of issues. A research study in which the contagious diseases of the plant are studied is known as potato plant pathology. In order to upgrade the agricultural areas, the disease detection application is applied. Accurate treatment advices are provided through disease management. A computer, digital camera as well as application software are the three components of machine vision system today. Within the application software, numerous algorithms are integrated. In biology, there are several applications that include digital image processing and image analysis technology

J. Singh
ECE Department, Chandigarh University, Gharuan, India

H. Kaur (✉)
CSE Department, Chandigarh University, Gharuan, India
e-mail: harpreet8307@gmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_154

in order to propose advances within the microelectronics as well as computer systems [2]. The various issues that are relevant to traditional photography are also resolved here. The images that are collected from the microscopic to telescopic ranges are enhanced and analyzed with the help of this tool. Within biology, there are several applications involved. For less production of potatoes, there are numerous factors that are responsible such as early and late blights, insect damage, and roll viral diseases. Within image analysis and pattern recognition, image processing is the initial step and the most important one as well [3]. The final results of the analysis are determined through this step. The process through which an image can be partitioned into disjoint regions is known as image segmentation process.

2 Literature Review

Prakash et al. (2017) studied several image processing techniques in order to identify diseases in plants [4]. In order to identify the leaf diseases and classify them, the image analysis and classification techniques are to be implemented as an aim in this paper. There are four different parts of the proposed framework. They include image processing, segmentation of leaf using k-means clustering in order to identify the diseased areas, feature extraction, and classification of diseases. Using statistical Gray-Level Co-Occurrence Matrix (GLCM) features and classification, texture features are extracted in this approach. The Support Vector Machine (SVM) classifier is used within this proposed mechanism in order to provide better extraction of features.

Kaur et al. (2017) presented a comprehensive study related to several diseases identified within the fruits [5]. In order to minimize the time required to identify the diseases within the fruits, an automated approach is generated for identifying the diseases within fruits. The image is distorted due to the noise. Within this case, there is elaboration of denoising mechanism. As per the analysis of this study, most of the fruit crops are being affected by blight disease. The image of distorted lead is used in order to perform analysis such that the disease can be identified within these fruit crops. In order to identify the diseases within the initial stage, the singular valued analysis is used within the image processing techniques.

Dhaware et al. (2017) proposed a method which can identify the automatic leaf unhealthiness classification from leaves by using image processing [6]. Practical requisition is used here for applying the system because the images are forwarded directly as the farmers make the least efforts. Thus, the farmers that do not make much efforts are provided with advises through this technique. The plant leaves image is seized by the farmer by using the mobile camera. Without making any additional inputs, this image is forwarded to the DSS approach.

Padol et al. (2016) proposed the method that utilized SVM classification technique for identification and classification of diseases present within grape leaves [7]. Within image processing, methods such as resizing, thresholding as well as Gaussian filtering are applied. In order to execute segmentation on the leaf, k-means clustering technique is utilized in which texture and color features are utilized. Further, in order to identify the type of leaf disease, SVM classification technique is utilized. Downy mildew and Powdery mildew are the two different classes of grape leaves that were utilized within this experiment. For both of these categories, an average of 88.89% of accuracy is provided by this proposed system.

Rajan et al. (2016) proposed a mechanism that utilized image processing in order to identify pests within the crops [8]. Upon different images gathered from various sources, tests were performed using this system. At an early stage, it is possible to identify the pests within crops through this proposed method. Thus, the usage of pesticides in agricultural fields can be minimized here through which costs and environment both can be saved. As per the evaluations made, this system is termed as simple and efficient for various applications. In comparison to other manual systems, the time and accuracy level achieved through it is also enhanced.

3 Support Vector Machine

The set of relevant supervised learning mechanisms which are utilized in order to perform classification and regression are known as Support Vector Machines (SVMs). The empirical classification error can be minimized and geometric margin can be maximized simultaneously with the help of SVM. Due to this property, this is also known as Maximum Margin Classifier and it is also based on the Structural Risk Minimization (SRM). Toward the higher dimensional space in which maximal separating hyperplane is generated, the input vector is mapped by SVM. In order to partition the data, two parallel hyperplanes are generated on each side of the hyperplane. The hyperplane that maximizes the distance amongst two parallel hyperplanes is known to be the separating hyperplane [9]. It is assumed here that the larger is the margin or distance amongst these parallel hyperplanes, the better is the generalization error of the classifier. The data points that are considered here are in the form as given in Eq. (1).

$$\{(x1, y1), (x2, y2), (x3, y3), \ldots, (xn, yn)\} \tag{1}$$

Here, $yn = 1/-1$ is a constant that is used to represent the class in which point xn belongs. The numbers of samples involved are represented by n. The p-dimensional real vector is each xn. In order to guard against the variables that have larger variance, scaling is important. With the help of partitioning the hyperplane, the training data is viewed here which includes

$$w.x + b = 0 \tag{2}$$

Here, "b" is a scalar values and the p-dimensional vector is represented by "w". Perpendicular to the separating plane, the vector "w" is indicated. The margin is increased by adding the offset parameters "b". The hyperplane is forced to pass the origin in case "*b*" is not present due to which the solution can be restricted. There is an involvement of parallel hyperplanes here as well which can be described by the equations below

$$w.x + b = 1 \tag{3}$$

$$w.x + b = -1 \tag{4}$$

4 Proposed Methodology

The technique is proposed for the plant disease detection. The plant disease detection techniques consist of the following phases:

1. Preprocessing phase: In the first phase, the image is taken as input and the input image is converted to the gray scale image.
2. Textural feature analysis: The textural features of the input image are analyzed using GLCM algorithm. This algorithm, statistical texture analysis is performed on the observed combinations of intensities for calculating texture features. This calculation is done using the intensities at particular positions that are relevant to each other within an image. There are first-order, second-order and higher order statistics classified on the basis of the number of intensity points or pixels present within each combination. The method through which the second-order statistical texture features can be extracted is known as Gray-Level Co-occurrence Matrix (GLCM) method [10]. In numerous applications, this technique has been applied. Within an image that has number of rows and columns that are equal to the number of gray levels is known as GLCM matrix. The relative frequency through which two pixels are partitioned through a pixel distance (Δx, Δy) and their intensities are i and j for an element is defined as $P(i, j|\Delta x, \Delta y)$. Amongst the gray levels i and j at "d" displacement distance and (θ) angle, the matrix element $P(i, j \mid d, \theta)$ can be considered to have second-order statistical probability values. For an input image that has G gray levels from 0 to G–1 and $M \times N$ neighborhood, the intensity at sample m is considered as $f(m, n)$ for neighborhood line n. Hence,

$$P(i, j|\Delta x, \Delta y) = WQ(i, j|\Delta x, \Delta y) \tag{5}$$

Here, (Table 1)

$$W = \frac{1}{(M - \Delta x)(N - \Delta y)} \tag{6}$$

$$Q(i, j|\Delta x, \Delta y) = \sum_{n=1}^{N-\Delta y} \sum_{m=1}^{M-\Delta x} A \tag{7}$$

and

Table 1 Mathematical formulation of features [12]

Feature	Mathematical formulation
Contrast	$\sum_i \sum_j (i-j)^2 g_{ij}$
Correlation	$\frac{\sum_i \sum_j (ij) g_{ij} - u_x u_y}{\sigma_x \sigma_y}$
Energy	$\sum_i \sum_j g_{ij}^2$
Homogeneity	$\sum_i \sum_j \frac{1}{1+(i-j)^2} g_{ij}$
Mean, M	$\sum_{i=0}^{L-1} g(i) P(g(i))$
Standard Deviation, S	$\sqrt{\sum_{i=0}^{L-1} (g(i)-M)^2 P(g(i))}$
Skewness	$\frac{1}{S^3} \sum_{i=0}^{L-1} (g(i)-M)^3 P(g(i))$
Entropy	$\sum_{i=0}^{L-1} P(g(i)) \log_2 P(g(i))$
Kurtosis	$\frac{1}{S^k} \sum_{i=0}^{L-1} (g(i)-M)^k P(g(i))$
RMS	$\sqrt{\frac{1}{L*L} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (g(i,j)-I)^2}$

$$A = \begin{cases} 1\, if\ (m,n) = 1 \text{ and } f(m+\Delta x), n+\Delta y) = j \\ \quad 0 \quad \text{elsewhere} \end{cases} \tag{8}$$

5 Region-Based Segmentation

In this phase, the k-means clustering is applied for the region-based segmentation. The k-means clustering algorithm consists of the following phases:

1. Input data: The image which is taken as input is considered as the data on which segmentation needs to implemented.
2. Calculate arithmetic mean: The arithmetic mean of the input data is calculated which defines centroid point of the data.
3. Formation of segments: The Euclidean distance is calculated for the central point and points which have similar distance is clustered in one cluster and other in the second.
4. Selection of ROI: The segmented image is selected as ROI which is given input to next phase of classification.

6 Classification of Disease

The step of classification will classify the input image into defined disease. The ROI which is selected in the last step is taken as input. The KNN classifier is a technique that is used within the pattern recognition process to classify the objects on the basis of closest training examples present within the feature space is known as K-Nearest Neighbor algorithm (KNN). In this instance-based learning algorithm, there is only local approximation of the function and until classification, all the computation is held upon. Amongst all the machine learning algorithms, KNN is the simplest one. Through the majority vote of neighbors, an object is classified [11]. The class that is most common amongst its k-nearest neighbors is assigned that object. The object is directly assigned to the class of its nearest neighbor in the case when $k = 1$. For both classification and regression predictive issues, NN is utilized. The ease to interpret output; compute the time; and the predictive power available are the three important aspects that are considered while evaluating any technique. By default, Euclidean distance is used by $knn()$ function and the following equation is used to calculate it.

$$D(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 \cdots + (p_n - q_n)^2} \tag{9}$$

Here, the subjects that have "n" properties and are to be compared are represented as "p" and "q".

KNN is also computed by deciding the number of neighbors that can be selected here which is represented by parameter "k". The performance of KNN algorithm is affected on large scale by the choice of value of k. The variance generated due to random error can be minimized by maximizing k. The balance amongst overfitting and underfitting can be maintained on the basis of the selection of appropriate value of k. The output of the KNN classifier will give the disease name.

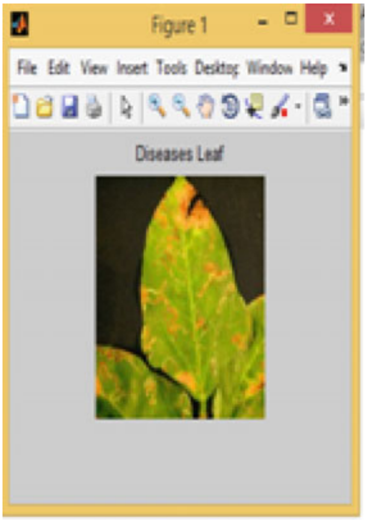
7 Results and Evaluation

From the publically accessible dataset "Plant Village", a database of around 40 images of potato leaves is gathered and used in order to evaluate the proposed algorithm. Image of 10 healthy leaves and 30 diseased leaves are present within the database used for conducting experiments. The complete database is partitioned into 2 different sets during the experiments. There are 24 images (60%) of images present within the training set and 16 images (40%) of images present within the testing set. The KNN that includes linear Kernel is utilized in order to perform classification. Various performance parameters such as accuracy, sensitivity, recall, and F1 score are utilized in order to evaluate the performance of this classification model. There is around 97% of accuracy achieved for classification.

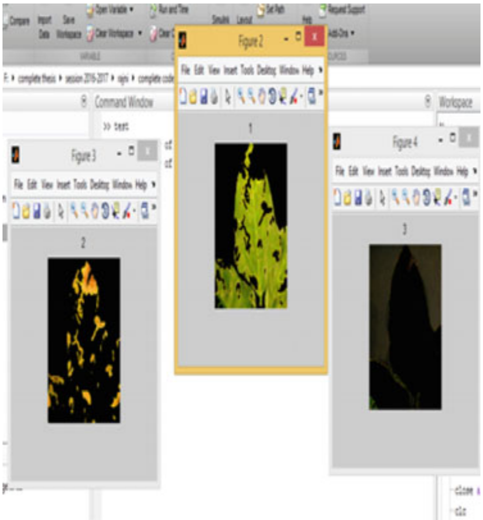
Table 2 shows the performance measures applied. Comparisons are made amongst potato and other species only since there is not much work proposed on disease detection of potatoes. The performance of the KNN classification technique is presented in Table 2.

Table 2 Performance measures of classification

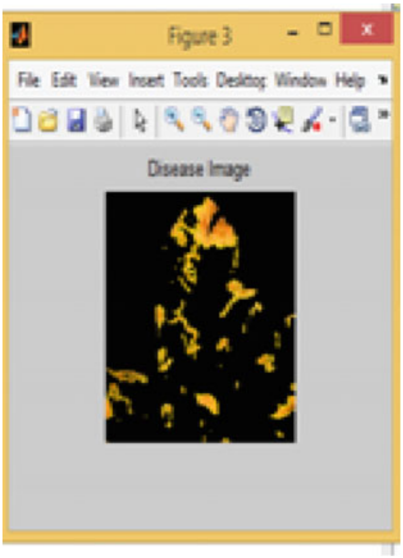
Class	Precision (%)	Recall (%)	F1 score (%)
0: Leaf Minar	89	94	92
1: Mosnic virus	97	93	95
2: White fly	98	98	98
Average/total	97	97	97



(a) Input Image



(b) Segmented Image



(c) Diseased Image

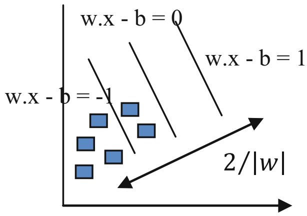


Fig. 1 Maximum margin hyperplanes for an SVM trained with samples from two classes

As shown in Fig. 1, a is the input image for plant disease detection. b Apply k-mean region-based segmentation. c Diseased image.

8 Conclusion

In this research paper, it is concluded that plant disease detection is the approach to detect diseases from the plants. In this work, the GLCM algorithm is applied for the textural feature analysis, k-means clustering is applied for the region-based segmentation, and KNN classifier is applied for the disease prediction. The simulation of the proposed modal is done in MATLAB and results are shown in form of figures and tables. The simulation results illustrated that accuracy is achieved up to 97% of disease prediction.

References

1. Rastogi A, Arora R, Sharma S, Leaf disease detection and grading using computer vision technology and fuzzy logic. In: 2015 2nd international conference on signal processing and integrated networks (SPIN)
2. Khirade SD, Patil AB (2015) Plant disease detection using image processing. In: 2015 international conference on computing communication control and automation
3. Waghmare H, Kokare R, Dandawate Y (2016) Detection and classification of diseases of grape plant using opposite colour local binary pattern feature and machine learning for automated decision support system. In: 2016 3rd international conference on signal processing and integrated networks (SPIN)
4. Prakash RM, Saraswathy GP, Ramalakshmi G (2017) Detection of leaf diseases and classification using digital image processing. In: 2017 international conference on innovations in information, embedded and communication systems (ICIIECS)
5. Kaur K, Marwaha C (2017) Analysis of diseases in fruits using image processing techniques. In: International conference on trends in electronics and informatics ICEI 2017
6. Dhaware CG, Wanjale KH (2017) A modern approach for plant leaf disease classification which depends on leaf image processing. In: 2017 international conference on computer communication and informatics (ICCCI—2017)
7. Padol PB, Yadav AA (2016) SVM classifier based grape leaf disease detection. In: 2016 Conference on advances in signal processing (CASP)

8. Rajan P, Radhakrishnan B, Suresh LP (2016) Detection and classification of pests from crop images using support vector machine. In: 2016 international conference on emerging technological trends (ICETT)
9. Duan X, Zhao T, Li T, Liu J, Zou L, Zhang L (2017) Method for diagnosis of on-load tap changer based on wavelet theory and support vector machine. In: 2017. In: The 6th international conference on renewable power generation (RPG)
10. Parvez A, Phadke AC (2017) Efficient implementation of GLCM based texture feature computation using CUDA platform. In: International conference on trends in electronics and informatics, ICEI
11. Zhang Zhongheng (2016) Introduction to machine learning: *k*-nearest neighbors. Ann Transl Med 4(11):218
12. Islam M, Dinh A, Wahid K (2017) Detection of potato diseases using image segmentation and multiclass support vector machine. In: 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE)