ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2022.0403

IDENTIFICATION AND RECOGNITION OF LEAF DISEASE USING ENHANCED SEGMENTATION TECHNIQUES

Brijendra Gupta[1], V. Elanangai[2], G.N. Naveen Kumar[3] and P.T. Kalaivaani[4]

[1]*Department of Information Technology, Siddhant College of Engineering, India*
[2]*Department of Electrical and Electronics Engineering, St. Peter's Institute of Higher Education and Research, India*
[3]*Department of Electronics and Communication Engineering, CMR Institute of Technology, India*
[4]*Department of Electronics and Communication Engineering, Vivekanandha College of Engineering for Women, India*

Abstract

Segmenting refers to the technique of breaking up an image into its component parts one by one. When it comes to the process of segmenting photos, there is a plethora of choice available at current point in time. These options range from the easy thresholding approach to the complicated color image segmentation techniques. The bulk of the time, the parts that go into making up these sub-assemblies are items that individuals are able to easily identify and categorize as being distinct from one another. As a result of the limitation of computer lack of intelligence to differentiate between distinct items, a wide variety of techniques have been devised and utilized in the process of segmenting photographs. In order to complete its tasks, the image segmentation algorithm requires a wide range of image characteristics to be provided as input. This could be referring to the colors that are contained within an image, the borders that are included within the image, or a particular region that is contained within the image. In order to break down color images into their component elements, we make use of an algorithm that is inspired by natural selection. The research uses enhanced segmentation techniques to identify and recognize the leaf disease in plants. The study conducts extensive simulation to test the efficacy of the model. The results show that the proposed method achieves higher segmentation accuracy than other methods.

Keywords:

Identification, Recognition, Leaf Disease, Segmentation

1. INTRODUCTION

In many parts of the world, agricultural land is put to use for a wide range of activities in addition to the production of food. Agriculture has a significant role in India entire economy and is hence of utmost significance. As a direct result of this, the process of identifying illnesses that affect plants is of the utmost significance in the agricultural sector. Utilizing a strategy that is based on automatic disease detection is advantageous when it comes to recognizing plant diseases in the early phases in which they manifest. For instance, in the United States, pine trees are subject to the small leaf disease, which is caused by a fungus that has the ability to do significant damage and is named after the sickness. The tree does not survive for more than six years after it was planted since it is unable to reach its full potential. The states of Alabama and Georgia, which are located in the southeastern part of the United States, are beginning to experience the consequences of it. In instances such as this one, the detection of the issue sooner may have led to the good outcomes being achieved [1].

The well-trained eyes of specialists who carry out visual inspections are currently the sole approach that is used to diagnose and identify plant illnesses. This method is fully dependent on the fact that it is the only way that is currently employed. Because it requires the hiring of a large number of professionals and the continual inspection of the plant, accomplishing this aim with large farms is quite expensive. This is because it is necessary for the goal to be accomplished. In the meanwhile, farmers in some countries may not have access to the necessary resources or knowledge about where they can receive assistance from skilled professionals. This may prevent them from receiving the assistance they require. Receiving professional counseling requires an excessive amount of effort, as well as a significant amount of both time and money. This is a direct result of the problem. The technique that was described for monitoring broad swaths of agricultural land functions pretty well when employed in this situation. Simply paying attention to the signs that manifest themselves on a plant leaves is all that is required in order to successfully diagnose a disease. This contributes to a reduction in the amount of work that is required as well as the expenditures that are linked with it. Machine vision, which is used for automatic process control, inspection, and robot guidance based on pictures [2], is also aided by this, which is helpful. Machine vision is utilized for automatic process control, inspection, and robot guidance.

The process of visually diagnosing plant diseases is one that is labor-intensive, lacks precision, and can only be carried out in situations that are relatively contained. On the other hand, a detection method that is automated would cut down on the amount of time and effort required while also improving precision. Fungi, viruses, and bacteria can regularly cause a wide variety of ailments in plants, many of which are fatal. A few examples of these flaws are early and late burning, brown and yellow patches, and others. Image processing enables the measurement of color differences between healthy and sick areas [5], as well as the estimation of the total area that has been influenced by the disease [6]. Image processing also enables the measurement of the total area that has been damaged by the disease.

2. RELATED WORKS

The author [3] carry out an in-depth analysis of the several different classification methods for plant leaf diseases. It would appear that the k-nearest-neighbor approach is the class prediction algorithm that is both the most accurate and the easiest to understand, at least based on the sample data that was provided for testing. This conclusion is based on the fact that the k-nearest-neighbor approach yielded the best results when it was put through its paces. Establishing appropriate parameters for the SVM can be difficult if the training data are not linearly separable. This is considered to be one of the SVM apparent limitations, as it is considered to be one of the SVM apparent weaknesses.

The authors detail the four basic steps of the established processing method in their published work [4], which can be found here. The first thing that has to be done is to generate a color transformation structure for the RGB image that was provided as input. This RGB is what is utilized to generate colors, while the image of RGB that has been altered or converted is what is referred to as HSI. Both of these are used to talk about colors. In the second step, a threshold value will be applied to the mask in order to remove the green pixels. This will take place in the phase after the first. After it has been decided which parts of the image can be used, the following step is to mask those parts and eliminate any green pixels that may be in the way. This should be done before moving on to the next stage. The process of segmentation is complete once the participant has completed the fourth and final primary phase.

The authors in [6] provide a method that can be implemented in order to categorize plant diseases and establish whether or not they are present. It has been established that a recognition system that is based on machine learning is capable of saving time, money, and human resources, all of which are significant in the Indian economy. In addition, a system that is based on machine learning may also improve accuracy. An approach known as the color co-occurrence method is used in this demonstration of the process of extracting the feature set. Neural networks are utilized in order to carry out the process of automatic diagnosis of leaf diseases. Although the proposed method requires less work to be done on the computer, it has the potential to make a substantial contribution to the correct identification of diseases that affect the leaf, stem, and root.

As described in the publication [7] the following is an explanation of each of the technique four fundamental steps, which are used to identify diseases: After determining a color transformation structure for the RGB image that is being input, the green pixels in the image are masked and deleted by utilizing a certain threshold value. This is done after the initial step of determining the color transformation structure. After that, the method for segmentation is carried out, and after that, the texture statistics are computed in order to create meaningful segments. Afterwards, the segmentation results are examined. At long last, the retrieved features are entered into a classifier so that a determination may be made regarding the disease category. The robustness of the strategy is proved by the utilization of experimental data from around five hundred plant leaves that are stored in a database.

Kulkarni et al. have developed a method for the early and accurate identification of plant illnesses. This method takes use of an artificial neural network (ANN) in conjunction with a variety of other image processing techniques. Because it employs an artificial neural network (ANN) classifier for the purpose of classification and a Gabor filter for the purpose of feature extraction, the approach that has been described enables the achievement of a larger recognition rate, one that can reach up to 91%. The ability to recognize and categorize plant illnesses is made possible by ANN-based classifiers, which take into account a wide range of characteristics, such as color, texture, and features [8].

The authors of this paper [9] provide a method for the diagnosis of diseases in Malus domestica by making use of advanced methodologies such as K-means clustering, texture analysis, and color analysis. This method was developed in order to improve the accuracy of disease diagnosis. In order to categorize and identify the many kinds of agriculture, it makes use of the texture and color characteristics that are frequently found in normal and damaged areas. These characteristics can be found in both healthy and unhealthy places. Three different classification strategies, namely the K-means clustering, the Bayes classifier, and the principle component classifier, are on their way to becoming acceptable substitutes.

According to [10], histogram matching can be applied as a tool to assist in the diagnostic process for diseases that have an effect on plant life. Histogram matching in plants is based on a method of edge detection in addition to a color feature. This is due to the fact that diseases first reveal themselves on the leaves of the plant. A method known as layer separation is used when training these samples in order to achieve the desired results. The RGB image is broken down into its component colors using this method, which begins with an RGB representation of the image. After that, an edge detection method is applied in order to find out where the beginning and ending points of each color are within the layered image.

A support vector machine is a promising technique to artificial intelligence that has extensive applicability for solving categorization issues, as stated in the study that was published by [11]. In the process of solving regression issues, a special kind of support vector machine (SVM) called support vector regression (SVR) is applied. using a strategy based on artificial intelligence that has a wide range of potential applications for dealing with classification difficulties. In order to solve regression issues, an application of the support vector machine (SVM) known as support vector regression (SVR) is applied. As a result of the fact that SVR makes it possible to generalize the solution model, it has attained a great deal of popularity in the academic community.

There is a significant link between the prevalence of diseases on plantations and the failure of a great deal of agricultural production. When it comes to the process of identifying diseases that can affect cotton plants, Bernardes et al. [12] present a method that is dependent on the automatic categorization that can be carried out by feature extraction of foliar symptoms from digital images. This method can be utilized in the process of identifying diseases that can affect cotton plants. For the purpose of classification, this methodology makes use of a support vector machine (SVM), and in order to extract features, it makes advantage of the capabilities offered by a wavelet transform.

The present a detailed explanation on the many different methods that can be used to segment medical images. There are three basic kinds of algorithms that can be categorized based on the concepts that lie behind them. These categories can be categorized as sequential, parallel, and distributed. These algorithms include pattern recognition techniques, threshold-based algorithms, and deformable model algorithms, among others.

Because of the research done on the topic, the third class of algorithms has, within the past several years, shifted its attention to deformable models as the primary object of its study. The division of the organs and tissues that are located in the pelvic cavity is one of the most essential applications for these algorithms. A few easy investigations are used to both introduce and investigate these concepts.

3. PROPOSED METHOD

The different methods of picture identification and partitioning that have been proposed:

- The quality of the image that is being received can be improved by the application of some preprocessing, and any undesirable deformations can be removed through this process as well. After isolating the portion of the image that requires smoothing, the image of the leaf is cut down so that only the affected portion is left. It is important to place as much emphasis as possible on the distinction between the two. In addition to that, there is image editing that takes place.
- In this stage of the process, we are going to make a mask that is going to cover the overwhelming majority of the green pixels on the screen. The generation of a threshold value for these pixels came first, followed by the incorporation of that value into our study.
- The vast majority of green pixels are masked by using the following method: if the intensity of the pixel green component is lower than a threshold value that was computed in advance, then the pixel red, green, and blue components are all assigned the value zero. This happens if the intensity of the pixel green component is lower than a threshold value that was computed in advance.
- It is necessary to eliminate any disguised cells that may be present within the bounds of the diseased clusters.
- It is important to compile any relevant information that can be used to aid in the classification of leaf diseases. In order to deconstruct the components, one may use a genetic algorithm if they so choose.

An examination into the efficiency of the segmentation approach should be carried out in the context of the blight disease that damages the leaves of agricultural crops. Cotton plants are susceptible to a proteobacterium called Xanthomonas axonopodis, which causes a disease known as blight. Blight is a disease that can be very destructive. In a manner that is analogous, the blight disease that affects the Solanum nigrum plant results in a loss in production that falls somewhere in the range of 27.57 to 57.36 percent.

In the previous section, it was mentioned that photographs of cotton and Solanum nigrum leaves are used as the input image (I), and the influenced region of interest (ROI).

The term disjoint property is the official designation given to this classification of real estate. A brand-new technique known as the indices-based histogram intensity segmentation algorithm has been introduced in order to enhance the outcomes of the segmentation and classification performed by the algorithm that was described earlier. If a farmer makes a mistake in diagnosing a disease or in calculating the course of action it will take, he or she may be tempted to use an excessive amount of fertilizer, which may have negative effects on human health. This can be avoided by avoiding making these types of errors in the first place.

It is absolutely necessary to have the skill of precisely predicting the start of a sickness at an early stage in order to avoid this risky situation and keep the quality of the food intact.

It is crucial for agriculturalists to locate the disease as quickly and properly as is humanly possible if they wish to avoid financial losses. After that, we will proceed to discuss each stage of the method for identifying the blight disease, which comprises the preprocessing, segmentation, and validation of images, respectively.

3.1 PREPROCESSING DISEASE AFFECTED LEAF

The research talks about how to filter out input noise. The term noise refers to the existence of signals that could be considered distracting, while filtering refers to the process of removing these signals from the input. Denoising is a method for cleaning up images by removing unnecessary noise from the image while maintaining any important features of the picture. Denoising a picture is an important operation that takes place during the preprocessing phase and assists to increase the accuracy of the end product. The sounds were cleaned up by the median filter, which was accountable for it. This is a representation of an input matrix that was used to feed data into a median filter. This matrix was used to feed the data into the median filter. It is a matrix with three rows and three columns.

123	118	114
124	114	113
123	118	113

Input

→ 113,113,114,114,[118],118,123,123,124

Fig.1. Pre-processing

Applying a sorting method to the matrix M, which serves as the method input, enables one to obtain the matrix median value. The example that was just shown illustrates how the picture may be made to look better by exchanging the values of the surrounding area for the median value, which is shown down below. The contrast of the shot was altered in some way. It is possible to get rid of any background noise by adhering to the approaches that have been outlined above and putting them into practice throughout the process of pixel representation of the input image. Because of this, it is now possible to extract the precise damaged area from the image of the leaf.

The images of diseased leaves that have been damaged by three distinct conditions—foiler, rot, and rust—are given a median filter treatment. After the photos have been cleaned, they are processed using a method called region segmentation, which is going to be further detailed in the following paragraphs.

3.2 SEGMENTATION OF LEAF IMAGE

The process known as thresholding is yet another technique that may be applied in order to extract the ROI from an image. The region of interest (ROI) could refer to a very particular portion of an image or the entire thing, depending on the requirements of the particular project that is currently being worked on. Determine a cutoff point for the histogram, which will be designated by the letter T, by using the distribution curve. There are times when the local threshold, the global threshold, and the optimal threshold methods can each be useful [21]. In order to get a better look at the jujube leaf, we cut it off using a technique called OTSU thresholding in conjunction with a skilled edge detection operator. This allowed us to inspect the leaf in greater detail.

It is probable that the method will not yield trustworthy findings when applied to other scenarios, such as the diagnosis of leaf blight, even though it functions exceptionally well for one specific application. Using the f function found in the computer program MATLAB, a picture of a leaf from the plant Solanum nigrum can be converted into a digital image for representational purposes (x, y).

Finding a threshold value, denoted by the letter T, enables the image to be cut up into its individual components, which are referred to as segments. If the intensity of a pixel is greater than T, then the value of that pixel is set to 1, and if it is less than T, then the value of that pixel is set to 0. If T is not greater than the intensity of a pixel, then the value of that pixel is always 0. The illustration of the thresholding formula for the leaf disease induced by Solanum nigrum that follows is as follows:

It is essential to determine the threshold value applicable to each specific application. You can use any one of numerous well-established functions to calculate the T value; however, the result will be different depending on the domain you choose to operate in. This is because the T value is dependent on the environment in which the calculation is performed. If you tried to calculate ROI for several different types of input images using the same T value, you would not get reliable results. Utilization of the image segmentation approach that distinguishes between the unhealthy area and the tissue that surrounds it in a healthy state When it comes to separating photographs into their individual components, region growth is a helpful tool to have at your disposal. Pick the part of the photo that you want to be the first thing you see when you look at it. The seed point should be used to make a clean cut through the infected area. An incorrect selection of seed points is one of the many challenges that are inherently present in the process of generating regions [12], and as a result, it can lead to over-segmentation of the data. This is just one of the many challenges that are inherently present in the process of generating regions. In the case of the cucumber downy mildew illness, the segmentation results obtained through the use of the region growth technique are more accurate than those obtained through the use of the OTSU and K-means algorithms. This is because the region growth technique uses a growing region to determine where the segments should be placed. The region-growth method, which was described, performed significantly better than previous methods that had been utilized in the past for the purpose of detecting foliar diseases in leaves. In this piece of study, we describe a novel approach that utilizes indices that are obtained from histograms. This method was developed by us. As its input, the analysis method makes use of photos of Solanum nigrum leaves of 256×256.

Establishing the image value threshold at which it needs to be set is the first step that needs to be taken before the diseased part of the leaf can be eliminated. Your ability to determine the value of the minimal difference between each pixel in the image and the minimal value will determine how well you can compute the absolute value of the difference between the minimal value and each pixel in the image.

A graphical depiction of the occurrences of each individual pixel over the course of time is provided by the histogram. Find out where each value sits inside the histogram that you are currently creating. The value that the variable L stores should be equivalent to the total length of all index values. Following the completion of that stage, the histogram length (HL) can be expressed in meters. Figure out where the centroid falls inside the value K multiplied by itself.

The requirements of the application, the ideal value for K may change from time to time. Calculate, utilizing Euclidean distance measurements, the distance between each pair of values of the indices, and then save the results of this calculation in the Histogram Clustering (HC) variable. This calculation should be done after each pair of values has been evaluated.

Keep the pixel in question if its value is greater than zero, and move the data from the control channel to the history channel. It is vital to carry out Steps 3–9 as many times as required until there is no discernible difference in the outcome. Following this, you need to make a mask for yourself and put it to use to conceal any information that you do not want other people to view. It is necessary to make a copy of the most recent data set and paste it into a new matrix before it can be saved in that location. The process of extracting the necessary data from the input image, followed by the display of that data The information that is collected from a ROI can be stored for future use and kept for future use if necessary for the identification procedure.

The indices-based intensity histogram method is applied to images of damaged areas on leaves that have been affected by one of three distinct diseases: foiler, rot, or rust. The goal of this method is to differentiate between the damaged areas caused by each of these illnesses. Following the completion of the segmentation process in the previous step, the next essential step is to hold a conversation on the significance of validity metrics. Finding the same region in the segmentation image requires the use of similarity measures as the guiding tool. The use of similarity measures is essential. To evaluate the degree of resemblance between the segmented image and the source image, a Sorensen-Dice formula was applied. When doing an analysis of findings based on dice measurements, it is best to use the value that can be achieved by extracting the maximum possible benefit from a segmented image. This is because this value offers the greatest potential for improvement. The verification is done out using the following methodology:

The procedure that came before it validates the segmentation components by using the equation that was stated earlier. Using a similarity index is a beneficial metric to employ when determining whether or not a segmentation is accurate. In order to ensure that the damaged portion of the leaf is successfully removed, this operation is carried out multiple times. Following that, we evaluate the outcomes of the experiments to figure out how effective the system is.

4. RESULTS AND DISCUSSIONS

MATLAB was used in each experiment that was conducted. Input data are assumed to be diseased leaf samples, which may contain things like burnt rose leaves, bacterially infected bean leaves, early charred banana leaves, and fungally infected bean leaves. Similarly, early charred banana leaves may also be present. Banana leaves that have been sunburned are another type of example of sick leaf samples.

To diagnose a range of plant diseases, it is possible to make use of images that have been segmented. Table.2 shows the input, which is a picture of a banana leaf with early scorch disease. Table.2 also shows the output, which is an image displaying the categorization of disease using feature extraction.

Table.1. Comparative Results

Disease samples	Training	No. of images used for testing	Detection accuracy/%		
			MDC with K mean	MDC with proposed algorithm	SVM with proposed algorithm
Banana	15	10	80.00	90.00	90.00
Beans	15	14	92.85	92.85	92.85
Lemon	15	10	90.00	100.00	100
Rose	15	12	83.33	91.66	100

The computation of the co-occurrence features can start as soon as the red, green, and blue components of the input image have been mapped to the thresholded images. After the leaf co-occurrence features have been retrieved, the feature values that correspond to those features in the feature library are compared to the retrieved feature feature values. The first phases in the process of classifying the data are the K-Means clustering and the minimal distance criterion. With an accuracy of 86.54%, the findings of this method demonstrate that it is extremely successful. The accuracy of detection is improved by a percentage that is equivalent to 93.63 percent using the strategy that was proposed.

Table 2 displays the total number of leaf disease samples that were correctly classified into one of the five unique leaf disease categories. This analysis was conducted out in accordance with the procedure that was recommended. Only a small portion of the samples under scrutiny for frog eye leaf spot and bacterial leaf spot were found to have inaccurate identifications. One frog eye leaf spot and one bacterial leaf spot were discovered on a grand total of just two leaves. The suggested method obtains a classification accuracy of 97.6%, which is much higher than the industry benchmark for accuracy, which is 92.7%.

Table.2. Classification results per class for proposed method

Leaf disease	Bacterial leaf spot	Frog eye leaf spot	Sun burn	Fungal	Early scorch	Accuracy
Bacterial leaf spot	23	2	0	0	0	92
Frog eye leaf spot	1	24	0	0	0	96
Sun burn	0	0	25	0	0	100
Fungal disease	0	0	0	25	0	100
Early scorch	0	0	0	0	25	100

5. CONCLUSION

The representative samples of each of the several plant diseases were collected for the sake of further investigation. The fact that the best results could be achieved with a minimal amount of computing labor was a demonstration of the effectiveness of the suggested method for the recognition and categorization of leaf diseases. This proved that the method is efficient. Using this technology has a number of benefits, one of the most important being that it enables the early diagnosis of plant diseases at an early stage, often known as a primary stage.

REFERENCES

[1] R.K. Nayak, R. Tripathy and D.K. Anguraj, "A Novel Strategy for Prediction of Cellular Cholesterol Signature Motif from G Protein-Coupled Receptors based on Rough Set and FCM Algorithm", *Proceedings of 4th International Conference on Computing Methodologies and Communication*, pp. 285-289, 2020.

[2] B. Subramanian, V. Saravanan and S. Hariprasath, "Diabetic Retinopathy-Feature Extraction and Classification using Adaptive Super Pixel Algorithm", *International Journal of Engineering and Advanced Technology*, Vol. 9, pp. 618-627, 2019.

[3] R.D. Aruna and B. Debtera, "An Enhancement on Convolutional Artificial Intelligent Based Diagnosis for Skin Disease Using Nanotechnology Sensors", *Computational Intelligence and Neuroscience*, Vol. 2022, pp. 1-8, 2022.

[4] V. Pooja and V. Kanchana, "Identification of Plant Leaf Diseases using Image Processing Techniques", *Proceedings of IEEE Technological Innovations in ICT for Agriculture and Rural Development*, pp. 130-133, 2017.

[5] N. Krithika and A.G. Selvarani, "An Individual Grape Leaf Disease Identification using Leaf Skeletons and KNN Classification", *Proceedings of International Conference on Innovations in Information, Embedded and Communication Systems*, pp. 1-5, 2017.

[6] P. Kantale and S. Thakare, "A Review on Pomegranate Disease Classification using Machine Learning and Image Segmentation Techniques", *Proceedings of International Conference on Intelligent Computing and Control Systems*, pp. 455-460, 2020.

[7] T. Fang and B. Wang, "Crop Leaf Disease Grade Identification based on an Improved Convolutional Neural Network", *Journal of Electronic* Imaging, Vol. 29, No. 1, pp. 1-14, 2020.

[8] P. Kaur, S. Bhatia and A.M. Alabdali, "Recognition of Leaf Disease using Hybrid Convolutional Neural Network by Applying Feature Reduction", *Sensors*, Vol. 22, No. 2, pp. 575-583, 2022.

[9] S. Bashir and N. Sharma, "Remote Area Plant Disease Detection using Image Processing", *IOSR Journal of*

Electronics and Communication Engineering, Vol. 2, No. 6, pp. 31-34, 2012.

[10] V. Singh and A.K. Misra, "Detection of Unhealthy Region of Plant Leaves using Image Processing and Genetic Algorithm", *Proceedings of International Conference on Advances in Computer Engineering and Applications*, pp. 1028-1032, 2015.

[11] P. Revathi and M. Hemalatha, "Classification of Cotton Leaf Spot Diseases using Image Processing Edge Detection Techniques", *Proceedings of International Conference on Emerging Trends in Science, Engineering and Technology*, pp. 169-173, 2012.

[12] A.S. Tulshan and N. Raul, "Plant Leaf Disease Detection using Machine Learning", *Proceedings of International Conference on Computing, Communication and Networking Technologies*, pp. 1-6, 2019.