The Classification Model for Plant Disease Detection with Segmentation and GLCM

Shalini Sharma[1], Rajesh E[2], Baskar Kasi[3], Sakthi Govindaraju[4]

sharmashalini414@gmail.com[1], rajsugan20@gmail.com[2], sakthihit@gmail.com[3]

School of Computing Science and Engineering, Galgotias University, India

Abstract. Diseases in plants may be detected automatically as the first symptoms show on the leaves. In this research, we provide a method based on algorithms for identifying plant diseases. The four steps that make up this method are as follows: picture pre-processing, image segmentation, feature extraction, and disease classification. A thorough literature study is done, including all of the many approaches that have been proposed before. The GLCM algorithm is used to assess the severity of illness based on the presence or absence of symptoms on diseased leaves, and a voting classifier is used to rank the symptoms. By integrating the Random Forest and Decision Tree methods with the Multi-Layer Perceptron, the voting classifier increases the reliability of early illness identification. The agriculture sector is crucial to the nation's gross domestic product. If farmers want to get the most out of their crops, they need to be able to spot illnesses early. Automation raises the bar for speed and accuracy in early disease diagnosis. Automatic detection of the illness occurs after symptoms show on the leaves. In this study, an automated system was developed to help with the diagnosis of plant illnesses. Its foundation is a four-step process: image preprocessing, image segmentation, feature extraction, and disease classification. This report also takes into consideration the results of a literature review based on the research methods that were previously outlined. The GLCM technique is used to assess a leaf's symptoms, and a vote classifier is offered to assist identify the condition. This classifier uses Random Forest, Decision Tree, and MLP methods to improve its ability to detect diseases at an early stage.

Keywords: Plant Disease, GLCM, MLP, Decision Tree, Random Forest, Voting.

1 Introduction

Many poor countries' economy rely heavily on agriculture. The agricultural sector is crucial to India's economy. Food and other necessities are provided by the agriculture sector of the Indian economy. Therefore, a considerable decline in output arises from any harm to the crop, which has a huge influence on the progress of the nation. The leaves of a plant are very fragile. Diseases in plants often initially show themselves as discoloration or wilting of the leaves. The size and

IACIDS 2023, November 23-25, Lavasa, India

DOI 10.4108/eai.23-11-2023.2343186

quality of a harvest depends on the health of the plants that produced it. Therefore, it is crucial to detect leaf diseases at an early stage [1]. The sooner a disease is contained and prevented from spreading on a farm, the better. Diseases in plants may be detected by the variety and condition of their leaves [2]. It takes more than average talent and effort to put this strategy into practise successfully. But it wouldn't be practical on such huge farms. Plant pathologists may trace the spread of contamination throughout a plant by looking for telltale indicators, such as unusual marks, coloration, or shading [3]. Many poor countries' economy rely heavily on agriculture. The agricultural sector is crucial to India's economy. Food and other necessities are provided by the agriculture sector of the Indian economy. Therefore, a considerable decline in output arises from any harm to the crop, which has a huge influence on the progress of the nation. The leaves of a plant are very fragile. Diseases in plants often initially show themselves as discoloration or wilting of the leaves. The size and quality of a harvest depends on the health of the plants that produced it. Therefore, early detection of plant diseases on leaves is crucial [4]. The sooner a disease is contained and prevented from spreading on a farm, the better. In instance, by examining the leaf type and condition, framers may discover plant illnesses. It takes more than average talent and effort to put this strategy into practise successfully. But it wouldn't be practical on such huge farms. Alterations to marks, colours, and other visual cues are scrutinised for signs of contamination. Several image processing methods are offered to detect the diseased leaves [5].

There are a number of image processing methods that might be used to identify sick leaves. The graphic displays a number of examples of how image processing may be used to identify leaf diseases. Here's how it goes step by step:

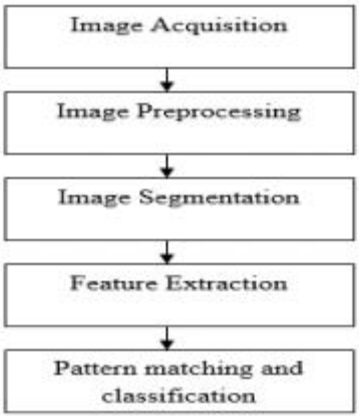

Fig. 1. Different Stages in Plant Disease Detection

These stages are explained as:

Taking a picture of anything in the actual world is called "image acquisition." The term is used here to describe the action of regaining access to a missing photo. The hardware is the first step in restoring the picture [7]. These actions are done to prepare for this potentially crucial source in the future. Since a picture must initially be taken, this is the starting point of the imaging system's workflow. All of the video that was compiled is real.

The term "pre-processing" refers to a stage in the image-processing pipeline in which extraneous image distortions are fixed and/or certain image qualities are enhanced. Image size and type can be changed on-the-fly, denoising can be done, the picture can be translated, the quality of the shot can be enhanced, and morphologic operations may be performed all during this phase [8].

To segment a picture is to divide it into smaller, more manageable sections. Information extraction and object recognition are at the heart of this approach. With this technique, the camera evaluates the resulting image. This technique is used to identify the edges and components of an image. In order to properly label a whole image, we make advantage of neighbouring pixels that share a similar label section. KMC, Otsu's algorithm, or thresholding methods may be used to segment the images.

When features are extracted, a "region of interest" (or "RoI" for short) is created. As a result, we will now take the RoI and parse it for its properties. This technique includes extracting a collection of characteristics called features from an image in order to generate data about the image and do subsequent processing. There is a wide range of signs that may be utilised to identify plants that have been tainted. Multiple methods exist for extracting attributes. A few examples are histogram-based feature extraction techniques [9] and graphical linear discriminant analysis (GLCM). Using the GLCM to statistically categorise textures.

Pattern Recognition Classification: This technique is used to analyse the given string of symbols in order to keep the essential features of a pattern while matching them. The strategy relies on contrasting your own attributes with those of the listed photographs in the repository. The patterns are organised according to the kind of inferences that may be made from them. To get things off, we will go over two distinct methods of classification: supervised and unsupervised learning algorithms. The sample pixels for the first-generation classifiers are selected by humans during training.

2 Literature Review

Abirami Devraj et al. (2019) investigated the role of agriculture in meeting the needs of a growing population. Seventy percent [11] of Asia's population worked in agriculture for their main source of income. However, the quality of crops was diminished by a number of other diseases. No damage could have been done to the agricultural industry when the sickness was correctly detected. The major objective was to create a piece of software that could automatically recognise and categorise illnesses. The illness was gradually discovered. Leaf images were utilised to help spot diseases in plants. Therefore, the image processing methods might be used in agriculture for the purpose of classifying plant illnesses.

R. Meena Prakash et al. (2017) employed image processing methods [12] to detect infections inside plant leaves. The major objective of this research was to classify leaf diseases using image analysis. The constructed system consists of four stages: image preprocessing, segmentation, feature extraction, and condition classification. We segmented images of the leaves using K-means clustering (KMC) to locate the specific sites of damage. We were able to extract additional texture attributes by using a Gray-Level Co-occurrence Matrix (GLCM) strategy. Leaf kinds on plants were finally classified using a Support Vector Machine (SVM).

Since plant diseases result in production losses, recognising them is an important issue in agriculture, as mentioned by Sharath DM, et al. (2019). This affected the quality of farm products [13]. Manually checking on the health of plants and looking for signs of insect infestations was a laborious and time-consuming task. Diagnosing plant diseases accurately needed specialised knowledge. The method outlined was also quite time-consuming. Therefore, image processing became relevant for diagnosing plant diseases. The process consisted of more than one phase. These tests allowed us to monitor the plant's recovery while it battled the disease. This research analysed digital images of damaged plants to find the most effective method for identifying plant disorders.

Gurleen Kaur Sandhu et al. (2019) investigated and discussed a number of methods for utilising photos to identify plant diseases. Over the last several decades, many researchers have been studying these mechanisms for their potential to identify plant diseases [14]. Some of the most important mechanisms were BPNN, SVM, KMC, etc. Algorithms were able to determine which plant leaves were healthy and which were not. However, the system did have a few holes in it. These vulnerabilities were the result of automation utilised in identifying frameworks, such as the usage of outside light and complex photographs captured in difficult climatic circumstances. The existing methods have been shown to be efficient and effective. Despite several limitations, these techniques may form the backbone of a comprehensive system for identifying plant diseases. Therefore, further research is required to advance the current body of work in this area.

The potential for early disease detection in plant leaves to dramatically minimise crop loss and boost agricultural productivity was first outlined by Sachin D. et al. [15] in 2015. Plant pathology was established on the study of plants with distinctive characteristics. For farming to be sustainable, it was essential to keep an eye on plant health and look for signs of illness. It took a lot of effort to physically monitor plants for signs of illness. Efficiency and diligence in the study of plant diseases were essential to the success of this method. The duration of this process was equally lengthy. That's why we needed better picture processing technologies. Identifying plant disease requires a number of steps, including capturing images, preprocessing them, segmenting them, extracting features from them, and classifying them. This study detailed the process of using photographs of leaves to identify plant infections. In addition, this publication outlined a few techniques for identifying plant infections.

R. Anand et al. (2016) proposed a new method for detecting illnesses that are transmitted by plant leaves. The cutting-edge method accurately identified illnesses [16]. Disease in the brinjal plant's leaves was identified using a combination of image processing and ANN algorithms. Brindle leaf rot was a serious issue. The production of brinjal fell dramatically for this reason. The brinjal plant was treated in this way simply for its leaves, rather than the complete plant. In around 85-95% of cases, the brinjal leaves served as the primary host for the infection. The photos were segmented using the KMC approach and classified using NN. The ANN framework successfully diagnosed leaf diseases in brinjal plants.

3 Research Methodology

Plant disease detection uses an image of a leaf to pinpoint the site of infection. This study proposes a number of methods to improve plant disease diagnostics.

Conditioned: -The initial stage in plant disease detection is image pre-processing. Plant Village is the dataset of choice since it is reliable and contains information on plants and the diseases that impact them. There are pictures of both healthy and blight-affected potato leaves. Before anything more can be done with the photographs, they need to be converted from RGB to grayscale.

Segmentation is the next stage in preparing plant pictures for object identification or data extraction. By inspecting the edges of pictures and objects, segmentation makes image analysis more manageable. Because pixels with the same label have distinctive characteristics, K-Means Clustering (KMC) may be used to classify objects. In order to segment images, a higher value of k (three) is employed once disease prediction from the input leaf has determined the required area.

The Feature Extraction process involves the removal of characteristics from the obtained ROI. Identifying the affected areas of plants may be aided by extracting features such as colour, texture, and more. One statistical method for determining textural characteristics is the correlation on a saturation scale matrix. When characteristics such as contrast, IDM, entropy, etc. are retrieved, early signs may be discovered. The following is a description of these characteristics:

Contrast is used to determine the quality of an image's finer differences.

$$Contrast = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^{G} \sum_{j=1}^{G} P(i,j) \right\}, |i-j| = n \qquad (4)$$

Supporting inputs from P(i,j) orthogonal to the diagonal (ij) is a primary function of this module.

Homogeneity: It's a measure of how close the GLCM's element distribution is to the diagonal.

$$\sum_i \sum_j \frac{P_d[i,j]}{1+|i-j|} \qquad (5)$$

Local Homogeneity, IDM:

$$IDM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} P(i,j) \dots (6)$$

Inhomogeneous regions (ij) are given less weight by the Inverse Difference Moment (IDM) because the weighting factor (1+(i-j)2)(-1)IDM. pictures with a lack of homogeneity have lower IDM values, whereas pictures with a greater sense of uniformity have higher IDM values.

In contrast, entropy quantifies the amount of information present in a picture. Intensity distribution unpredictability is measurable in this way. The 1st order entropy is lowest in inhomogeneous settings and highest non homogeneous ones.

$$-\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P(i,j) \times \log\big(P(i,j)\big) \dots . (7)$$

Correlation: This characteristic is used to determine the linear gray-level dependency between pixels at particular positions in relation to one another.

$$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P \frac{\{i \times j\} \times P(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y} \dots . (8)$$

Sum of Squares, Variance:

$$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} (i-\mu)^2 \, P(i,j) \qquad (9)$$

This feature underlines the need of allocating larger weights to components that can be distinguished from the average valueof$P(i,j)$.

Sorting Information By: - Finally, a mechanism for identifying plant diseases has to be developed. The whole dataset is split into a training set and a test set. About 60% of the information is used in the introduction, while the remaining 40% is used in the final disease classification. The decision is based on the results of a voting classifier. This classifier's superior accuracy is the result of a combination of decision trees, random forests, and Gradient boosting. It is feasible to utilise a decision tree to break down the whole set of possibilities into more manageable subsets. The specified decision-making criteria serve to differentiate each subgroup from the others. When studying data containing hierarchical and fixed-target variables, decision trees are a valuable tool. Both the selection criteria and the halting rule are crucial parts of the DT algorithm. The Gini index, the entropy of the data, or the misclassification error may all be used to narrow down the possibilities. When the aim is known in advance, one common criteria is to minimise the square of the error. First, DT examines the whole set of training data to locate each of the predictive factors across all conceivable gaps. This will carry on indefinitely until the withholding limitations are implemented. The DT does not need a functional technique for its predictors since it is non-parametric and does not presume a certain distribution. Because even a little change in the training information may significantly influence the tree structure, decision trees are fundamentally faulty. It is hypothesised that by pooling the results from several trees, one may benefit from volatility and uncover previously unnoticed signals in the data. Since RF is made up of several performance-enhancing designs, it is classified as an ensemble framework. When opposed to a single tree, RF's numerous trees are far more disorganised as they compete to show their superiority. The evolution of these trees was influenced by a wide range of factors. of linkage between procedures or between branches of a tree.

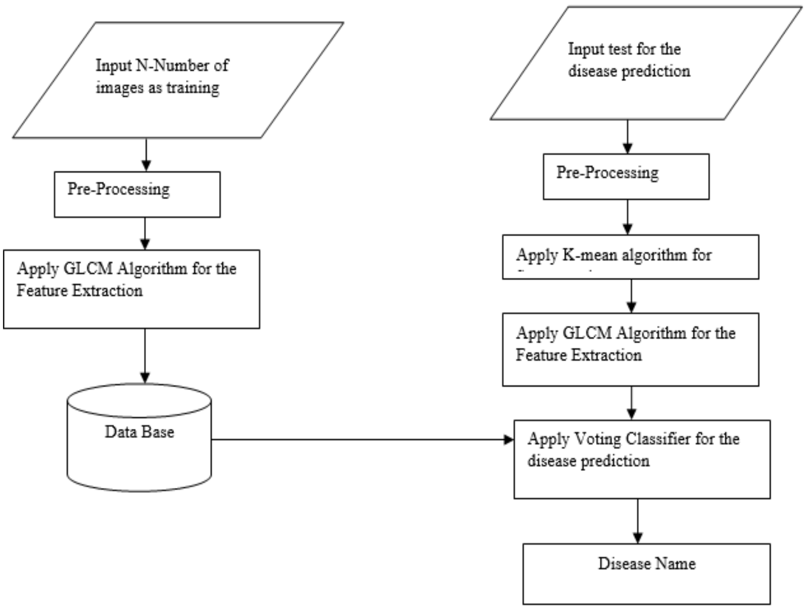

Fig. 2. Proposed Model

4 Result and Discussion

The method used makes advantage of the publically accessible Plant Village dataset, which is rich in information on many plant diseases. A short explanation of the corresponding medical issue is provided beside each picture in this set. In the developmental phase, we use images of potato leaves. Images may be sorted into three categories: those depicting healthy circumstances, those depicting the early stages of blight, and those depicting the late stages of the disease. There are a variety of measures used to assess performance, including as Memory, precision, and accuracy are all required. We'll discuss these defensive methods in further depth below:

A. The percentage of properly identified samples compared to the total number of samples in a particular programme is what is meant by "accuracy." The value of the parameter may be expressed numerically as:

$$A_i = \frac{t}{n} \cdot 100$$

To calculate the proportion of correct classifications, t, we need to know how many samples were used (n).

B. Accuracy: This metric shows how many correct positive results there were relative to the total number of positive results.

$$\text{Precision} = TP/TP+FP \ldots (11)$$

C. ***Recall:*** This metric is found by dividing the TPs with the total positive cases.

Table 1. Comparison Analysis

Classifiers	Accuracy	Precision	Recall
SVM	77.77%	60%	78%
KNN	62.96%	57%	63%
MLP	66.66%	58%	67%
Random Forest	81.48%	81%	81%
Proposed Model	85.18%	85%	85%

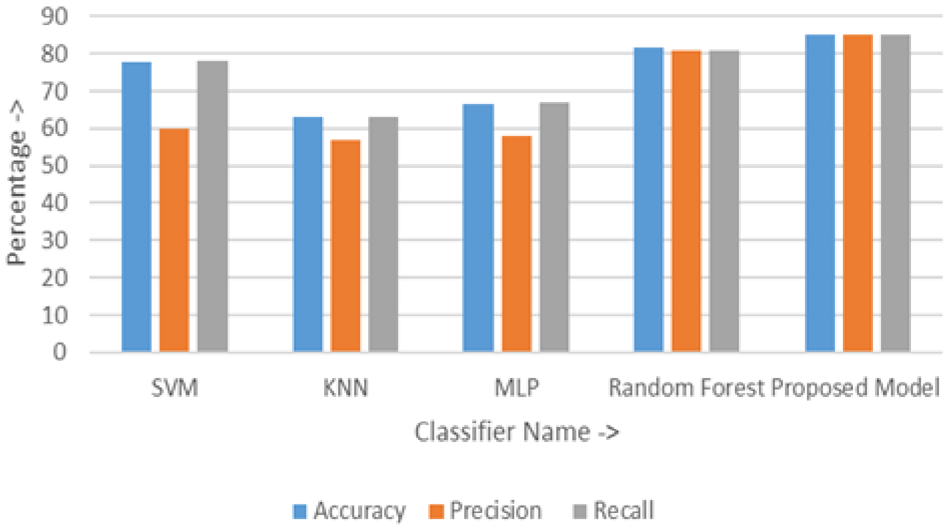

Fig. 3. Performance Analysis

The effectiveness of the suggested model is compared to that of the SVM, KNN, MLP, and Random Forest models in Figure 4. The proposed model outperforms state-of-the-art models by roughly 5% in terms of accuracy, precision, and recall.

5 Conclusion

The primary objective of this research is to create and validate an end-to-end system for the identification of plant diseases. All four phases of the system include picture pre-processing, attribute extraction, image segmentation, and illness classification. Once upon a time, microscopy was required for identifying plant ailments. The approach was formerly very successful, but it lost favour owing to the challenges of microscopic leaf examination. In order to extract features from images, we use the K-Means Clustering (KMC) method for segmentation and the Gray-Level Co-occurrence Matrix (GLCM) strategy. We're going to stop using the Support Vector Machine (SVM) technique in favour of this new one. The proposed procedure incorporates a voting classification system. Using three metrics, we find that the innovative method is about 10% more accurate and has a lower False Positive Rate (FPR) than the status quo technique.

References

[1] Rekha Balodi, Sunaina Bisht, Abhijeet Ghatak and K.H. Rao, “Plant disease diagnosis: technological advancements and challenges”, Indian Phytopath. 70 (3): 275-281 (2017).

[2] Monishanker Halder, Ananya Sarkar, Habibullah Bahar, “PLANT DISEASE DETECTION BY IMAGE PROCESSING: A LITERATURE REVIEW”, 2019, SDRP Journal of Food Science & Technology

[3] Surender Kumar, Rupinder Kaur, “Plant Disease Detection using Image Processing: A Review”, International Journal of Computer Applications (0975 – 8887) Volume 124 – No.16, August 2015

[4] Sabah Bashir, Navdeep Sharma, “Remote Area Plant Disease Detection Using Image Processing”, IOSR Journal of Electronics and Communication Engineering (IOSRJECE), Volume 2, Issue 6 (Sep-Oct 2012), PP 31-34

[5] Rajneet Kaur, Miss. Manjeet Kaur, “A Brief Review on Plant Disease Detection using in Image Processing”, International Journal of Computer Science and Mobile Computing, Vol. 6, Issue. 2, February 2017, pg.101 – 106

[6] Sushil R. Kamlapurkar, “Detection of Plant Leaf Disease Using Image Processing Approach”, International Journal of Scientific and Research Publications, Volume 6, Issue 2, February 2016

[7] Santhosh Kumar.S, B.K.Raghavendra, “Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review”, 2019, 5th International Conference on Advanced Computing & Communication Systems (ICACCS)

[8] Saradhambal.G, Dhivya.R, Latha.S, R.Rajesh, “Plant Disease Detection and its SOLUTION USING IMAGE CLASSIFICATION”, International Conference on Advanced Computing & Communication Systems (ICACCS), Vol. 119, No. 14 2018, pp. 879-884

[9] Sujatha R, Y Sravan Kumar and Garine Uma Akhil, “Leaf disease detection using image processing”, 2017, Journal of Chemical and Pharmaceutical Sciences, Volume 10 Issue 1

[10] Tejal Deshpande, Sharmila Sengupta, and K.S.Raghuvanshi, “Grading & Identification of Disease in Pomegranate Leaf and Fruit,” IJCSIT, vol. 5 (3), pp 4638-4645, 2014.

[11] A. Devaraj, K. Rathan, S. Jaahnavi and K. Indira, "Identification of Plant Disease using Image Processing Technique," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0749-0753

[12] R. M. Prakash, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017, pp. 1-4

[13] S. D.M., Akhilesh, S. A. Kumar, R. M.G. and P. C., "Image based Plant Disease Detection in Pomegranate Plant for Bacterial Blight," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0645-0649

[14] G. K. Sandhu and R. Kaur, "Plant Disease Detection Techniques: A Review," 2019 International Conference on Automation, Computational and Technology Management (ICACTM), 2019, pp. 34-38

[15] S. D. Khirade and A. B. Patil, "Plant Disease Detection Using Image Processing," 2015 International Conference on Computing Communication Control and Automation, 2015, pp. 768-771
[16] R. Anand, S. Veni and J. Aravinth, "An application of image processing techniques for detection of diseases on brinjal leaves using k-means clustering method," 2016 International Conference on Recent Trends in Information Technology (ICRTIT), 2016, pp. 1-6.