ABSTRACT

Plant maintenance is an important factor determining the quantity and quality of agricultural products. Crop diseases may lead to severe agricultural yield. Hence identification of crop diseases is essential to improve the agricultural yield. This disease identification process consists of several phases such as preprocessing, segmentation, feature extraction, and classification. Among these phases, feature extraction plays an important role because proper feature extraction leads to the highest classification accuracy. So this paper aims at developing a comparative model of feature extraction techniques that will facilitate crop production by accurately identifying diseases that affect productivity. This paper presents a comparison of feature extraction using texture, shape and color information separately as well as find out the best technique on each feature extraction.

Keywords: Leaf disease, ABC, GLCM, DWT, LBP, Gabor, Sobel Edge, Zernike Moment, Color Histogram, K-Nearest Neighbour Approach

I.INTRODUCTION

India is a grower and about 70% of the population depends on agriculture. Farmers have a variety of different ways of choosing different crops and finding insecticides that are suitable for plants. Plant diseases have resulted in a significant decrease in both the quality and quantity of agricultural produce. The study of plant diseases is related to the study of plants observed by the eye. Plant health and disease surveillance play an important role in successful farming on farms. Initially, the control and analysis of plant diseases were performed manually by experts in the arts. This requires a large number of works and requires too much processing time. Image processing techniques can be used to detect plant diseases. In most cases, the symptoms of the disease are observed on the leaves, leaves, and fruit. Plant leaves for disease detection are considered to show symptoms of the disease. Automatic plant detection diseases is a major research topic, as it can benefit from monitoring large crops and thus automatically detect symptoms of diseases as soon as they appear on the leaves of the plants. There are many image processing techniques for fast and accurate detection of plant diseases. The image needs to go through stages such as image acquisition, pre-processing, disease spot segmentation, feature extraction, and disease classification. The accuracy of the result depends on the method used for disease spot detection. The accreditation process is used to the category the input data provided in the number of classes and groups. It helps to category data based on selected features. It is very important to identify an early infection and propose a solution to reduce loss. In this paper, the comparative study of extracted techniques for image characteristics is used in the detection of plant diseases.

BACKGROUND OF THE RESEARCH
Different algorithms and techniques used to identify plant diseases. Sushil R. Kamlapurkar proposed a method that was used to identify plant diseases. In this paper extract, which is used for the Gabor filter, and after applying the extraction characteristics, plant plants, the leaves are considered using an artificial neural network. The main drawback of this article is the complicated calculation [1].

Surender Kumar and Rupinder Kaur proposed another approach to identify plant diseases. This article is used partly on a virus-based section using edge-based brightness based on similarities and areas of differences are made. After the excerpts were extracted, using the coarse coarse grayscale method. Classify plant diseases using linear and nonlinear filters [2].

Bernardes A. A. et al, [3] provided a method to automatically classify cotton disease by removing the symptoms of digital sheets. Wave change energy is used to derive traits while SVM is used for division. Images of alleged rotting leaves are classified as one of the sub-four: MA, RA, AS and NO. The system works well with the accuracy of 96.2% for the SA class, 97.1% for MA, 80% accuracy for RA and 71.4% for AS class.

Malvika Ranjan, Manasi Rajiv weginwar, Neha Joshi, Prof.A.B. Ingol has proposed methods for detecting and classifying leaf diseases. This paper extracts features by converting the RGB format to saturation of color. From the hue, saturation values draw from the future. On the basis of the characteristics that are classified as leaf diseases using the artificial neural network [4].

P.Revathi et al, [5] use three characteristics: variations, appearance, shape, and texture to identify palm leaf points. The variant function is calculated by histograms. The shape of the chamfer shaft is calculated by Sobel and Canny edge detection method. The biased texture function is calculated from the Gabor filter and texture descriptor. The study was based on six types of disease and used three attributes combining the Neural Network that accurately recommended 95%.

Xinhong Zhang et al. [6] recommend the machine-based visibility tool based on automatic verification techniques used in tobacco leaves. Machine vision techniques are used in this system to address issues related to extracts and tobacco leaf analysis, including color, texture, size, appearance, and texture. The experimental results suggest that the system can remove the traits of tobacco leaves and can be used to automatically classify tobacco leaves.

Jagadeesh Pu Pujari, Rajesh Yakkundimath Abdulmunaf S.Byadgi, proposed the method used for the extracted gray matrix derivatives and the limitations of plant diseases using the nearest neighborhood classification in this classification algorithm [7]. Describes the method for automatic disease detection systems found in sugar leaves. Several descriptors, such as aspect ratio, eccentricity, and circularity have been tested to identify disease spots on sugar cane leaves. Using this method has given a classification accuracy of 95.25% using a minimum distance classifier [8].

Amandep Singh Meydandar La Singh suggested the method used to convert the image to an HSI RGB image, and then a part of the virus infected with morphological operations. Comparison of training courses for the classification of plant diseases [9].

Garima Tripathi [10] focuses on automatic image processing methods and suggested the detection of plant diseases. The collaborative color method is applied to download a set of colors and texture specific characteristics to the type of leaf disease. A set of properties is used as a pathway for training networks and the subsequent discovery of leaf disease. Based on the proposed approach, a simple, affordable automated system for detecting and classifying plant diseases may be established.

S.Arivazhagan [11] focusing on the use of diagnostic texture scanning and plant pathology, is explained in this article. Proposed algorithms were tested on 10 species of plants - bananas, beans, mango, potatoes, tomatoes, and sapota. The experimental results suggest that the proposed approach
can recognize and classify leaf disease with little calculation effort. So plant diseases can be identified at an early stage and insect pests can be used to address pest problems while minimizing risks to humans and the environment.

The rest of the paper is organized as follows. In Section II we briefly explain the methodology of various feature extraction comparing the plant disease classification. Section III describes experimental results on the classification performance of plant disease using various feature extraction. Finally, concluding remarks and comparative analysis is given in Section IV.

II. METHODOLOGY

The workflow of the proposed method is shown in Fig. 1. In the first module, the disease affected part of the input image is segmented. In the second module, the texture features are extracted from the segmented part of the input image. In the last module, the classification approach is used for finding the affected disease in the input image. The further details of these modules are discussed below:

CROP DISEASE DETECTION FROM LEAF IMAGES

The proposed method has three modules. They are

1. Affected Part Segmentation
2. Feature Extraction
3. Classification

3.1 Affected Part Segmentation

In this module from the given input image the disease affected part is segmented. Then only find the type of disease and its severity can also be easily measured. To segment, the affected part of the Artificial Bee Colony approach is used [13].

Artificial Bee Colony Algorithm

1. Initialize the bee colony $X = x_i \{i = 1, 2, ..., n\}$, where $n$ denotes the population size, $x_i$ is the $i$th bee.
2. According to the fitness function, calculate the fitness $f_i$ of each employed bee $x$, and record the maximum nectar amount as well as the corresponding food source.
3. Each employed bee produces a new solution $v_i$ in the neighborhood of the solution in its memory by $V_i=x_i + (X_i - X_k) \times \Phi$, where $k$ is an integer near to $i$, $\neq i$, and $\Phi$ is a random real number in $[-1, 1]$. 

---

S. Vijayalakshmi, D. Murugan, T. Ganesh Kumar
FEATURE EXTRACTION APPROACHES FOR LEAF DISEASE CLASSIFICATION:
CONSEQUENCES EVALUATION

4. Use the greedy criterion to update $X_i$. Compute the fitness of. If $V_i$ is superior to $X_{i-1}$, $X_{i-1}$ is replaced with; $V_i$ otherwise $X_{i-1}$. Remained.
5. According to the fitness $f_i$ of $x_i$, get the probability value $P_i$ via formulas (1) and (2).
\[
P = \frac{\text{fit}_i}{\sum_{i=1}^{n} \text{fit}_i}
\]
\[
\text{fit}_i = \begin{cases} 
\frac{1}{1+f_i}, & \text{if } f_i \geq 0 \\
1 + \text{abs}(f_i) & \text{if } f_i < 0 
\end{cases}
\]
6. Depending on the probability $P_i$, onlookers choose food sources, search the neighborhood to generate candidate solutions, and calculate their fitness.
7. Use the greedy criteria to update the food sources.
8. Memorize the best food source and nectar amount achieved.
9. Check whether there are some abandoned solutions or not. If true, replace them with some new randomly-generated solutions by $x_i = \min + (\max - \min) \times \Phi$, where $\Phi$ is a random real number in $[0, 1]$, $\min$ and $\max$ stand for lower and upper bounds of possible solutions respectively.
10. Repeat steps (3)–(9), until the maximum number of iterations ($k_{max}$) is reached or stop conditions are satisfied.

3.2. Feature Extraction

After the segmenting process, the next step is to calculate the features of the disease affected part. The features are used to uniquely identify the disease name and its severity. To extract the features in this paper based on texture, shape, and color. S Gray Level Co-Occurrence Matrix (GLCM), Discrete Wavelet Transform (DWT) and Local Binary Pattern (LBP) methods are used in texture features. Shape-based feature extraction performed by Gabor Filter, Sobel edge and Zernike moment, the inherent values of a pixel in plant leave that is identified as meaningful shapes. Color based approaches considering specific colors about disease leaf through color histogram and color moment.

The GLCM is one of the feature extraction technique which extracts different combinations of pixel brightness values (grey levels) occur in leaf disease image. After changing the color image to gray-level image then extract the texture features using GLCM. To extract the 5 features such as energy, entropy, correlation, skewness and kurtosis of each leaf images. Our initial assumption in characterizing image texture in leaf image is that all the texture information is contained in the grayscale Co-occurrence matrices. Hence all the textural features are extracted from these gray-level Co-occurrence matrices.

Wavelet Filter is used to extracting the low level and high-level information from the image. The wavelet filter produces four bands such as LL, LH, HL, and HH after processing the input image. In these bands, L means Low Pass Filtered Image and H means High Pass Filtered Image. It is called a multi-resolution analysis technique because it produces the above bands in half of the size of the input image. By separating their bands can be used for processing. In this work, the Discrete Wavelet transform is used for texture features.

The local binary pattern (LBP) texture operator was first introduced as a complementary measure for local image contrast. The first incarnation of the operator worked with the eight-neighbors of a pixel, using the value of the center pixel as a threshold. An LBP code for a neighborhood was produced by multiplying the threshold values with weights given to the corresponding pixels and summing up the result. Since the LBP was, by definition, invariant to monotonic changes in grayscale, it was supplemented by an independent measure of local contrast.

Color is one of the most widely used features. Color features can be obtained by various methods like Color histogram and Color moment. The Color moment method has the lowest feature vector dimension and lowers computational complexity. Hence it can be considered as a suitable parameter to generate feature vectors which can be further used for leaf disease classification purpose. The color of the diseased leaf is used to generate the features of color histogram. The color feature extractions are applied to samples that are a healthy and unhealthy leaf of the plant. Plant disease is detected by using histogram matching is based on the color feature.
A color histogram of an image represents the distribution of the composition of colors in the image. It shows different types of colors appeared and the number of pixels in each type of colors appeared. The color histogram can be built for any kind of color space, although the term is more often used for three-dimensional spaces like RGB or HSV. The color histogram can also be represented and displayed as a smooth function defined over the color space that approximates the pixel counts.

Moments are a purely statistical measure of pixel distribution around the center of gravity of characters and allow capturing global character shapes information. Zernike moments are used as shape descriptors and identified as rotation invariant due to Orthogonality property. The Zernike moments, however, are only invariant to image rotation for them. To achieve translation and scale invariance, extra normalization processes are required. The translation normalization is achieved by moving the image centroid.

Gabor filters are directly related to Gabor wavelets since they can be designed for a number of dilations and rotations. However, in general, expansion is not applied for Gabor wavelets, since this requires computation of bi-orthogonal wavelets, which may be very time-consuming. Therefore, usually, a filter bank consisting of Gabor filters with various scales and rotations is created. The filters are convolved with the signal, resulting in a so-called Gabor space.

The Sobel operator is used in image processing and computer vision particularly within edge detection algorithms where it creates an image emphasizing edges. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. At each point in the image, the result of the Sobel edge operator is either the corresponding gradient vector or the norm of this vector. The Sobel edge operator is based on convolving the image with a small, separable, and integer-valued filter in the horizontal and vertical directions and is therefore relatively inexpensive in terms of computations. On the other hand, the gradient approximation that it produces is relatively crude, in particular for high-frequency variations in the image.

### 3.3. Classification

The final process is to classify the disease name and its severity. To do this process the K Nearest Neighbor Classifier is used. In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor [13].

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point. A commonly used distance metric for continuous variables is Euclidean distance. The Euclidean distance between point p and q is calculated by using the below formula

\[
d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}
\]

\[
= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}
\]

### IV. PERFORMANCE ANALYSIS

#### 4.1 Experimental Images

In this paper, the images taken from the real cameras are used. In total 200 images are taken. From that 100 images are used for training and the remaining 100 images are used for testing. The sample images are shown in below Fig 2. The image size is 512 x 512 color images. All of these images are which are affected by any one of disease. These images are used for experimental purposes.
FEATURE EXTRACTION APPROACHES FOR LEAF DISEASE CLASSIFICATION: CONSEQUENCES EVALUATION

Table 1 : comparative results on various feature extraction approaches

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>Crop Disease Detection Method for Feature Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Texture</td>
</tr>
<tr>
<td>GLCM</td>
<td>88</td>
</tr>
<tr>
<td>Wavelet</td>
<td>83</td>
</tr>
<tr>
<td>LBP</td>
<td>84</td>
</tr>
<tr>
<td>Sobel</td>
<td>17</td>
</tr>
<tr>
<td>Gabor</td>
<td>16</td>
</tr>
<tr>
<td>Zernike</td>
<td></td>
</tr>
<tr>
<td>Color Histogram</td>
<td></td>
</tr>
<tr>
<td>Color Moment</td>
<td></td>
</tr>
</tbody>
</table>

The above table shows the classification accuracy and error rate obtained using the various feature extraction techniques. This experiment result takes texture, Shape, Color features to extract the disease affected part. Here uses the KNN approach as the classifier which gives the best result as shown in Table 1. In this table shows the detection accuracy analysis of LBP, Zernike moment and color moment methods are giving the best results to extracting the affected parts in leaves.

4.2 Performance Analysis

To evaluate the performance of the crop disease detection techniques several performance metrics are available. This paper uses the Classification Accuracy and Error Rate to analyses the performance.

Classification Accuracy

Accuracy is the measurement system, which measures the degree of closeness of measurement between the original disease and the detected disease.

\[ \text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \]

Where, TP – True Positive (equivalent with hit)
FN – False Negative (equivalent with the miss)
TN – True Negative (equivalent with correct rejection)
FP – False Positive (equivalent with false alarm)

Error Rate

Error Rate is the measurement system, which measures no of falsely identified diseases name forms the given input images.

\[ \text{Error Rate} = \frac{\text{Number of Images of Falsely identified diseases}}{\text{Total Number of Images}} \]
Fig 3. Accuracy of Texture

Fig 3 shows the accuracy value with the GLCM, wavelet, and LBP. LBP method obtained the best results for texture features than other existing methods.

Fig 4. The error rate of texture

Fig 4 shows the results on error rate analysis for texture feature extraction method. As seen in this figure as plant leaves of error rate in the LBP method is lower than existing approaches as GLCM and Wavelet.
Fig 5. Accuracy of Shape

Fig 5 shows the accuracy value of Zernike moment is better than the other existing approaches such as Gabor and Sobel.

Fig 6. The error rate of shape

Fig 6 shows an error rate value on the Zernike moment feature method which is lower than the Sobel and Gabor approach.

Fig 7. Accuracy for color

Fig. 7 shows an accuracy analysis value of the color moment which is higher than the color histogram approach.
Fig 8 computes the error rate value on crops using the color histogram and color moment based on color features. As a result, the color moment obtained the best results than a color histogram.

### IV. CONCLUSION

In this paper, a comparative study on feature extraction techniques was detected to identifying the diseases in plant leaves. The features extraction methods are used to identifying the disease name and its severity. To extract the features in this paper based on features such as texture, shape, and color. In texture, features are applied were Gray Level Co-Occurrence Matrix (GLCM), Discrete Wavelet Transform (DWT) and Local Binary Pattern (LBP). In shape, features have used the method as Gabor Filter, Sobel Edge and Zernike Moment. In color, the feature is estimated based on Color Histogram and Color Moment. From the experimental results: (1) LBP performs well than texture feature such as GLCM and DWT (2) Zernike moment feature produces the best result on accuracy with reducing the error rate. (3) Color moment approach performs better than the color histogram approach. Therefore, these approaches are considered as the best method for the feature extraction process.
REFERENCES