Grenze International Journal of Engineering and Technology, June Issue

Recognition and Classification of the Paddy Leaf Diseases using K-Nearest Neighbor Classifier

A S Raju
Professor and HOD Department of BioMedical Engineering, Sri Siddhartha institute of technology, tumakuru
Email: asraju74@gmail.com

***Abstract*—The majority of individuals in rural regions work in the agricultural field industry. The primary component of a plant that produces food is a leaf. By conflict the diseases on the plant's leaves, stems, and nodes, the main factor is to reduce the yield. Therefore, early illness detection is essential to preventing losses. The diseases like a blast, brown spots, and bacterial blight are affected by the paddy leaf. These are impacts to reduce the growth of the plant and produce less yield. The image processing technique is used to detect the diseases on paddy leaves.**

***Index Terms*— KNN, Leaf Diseases, Image Processing.**

I. INTRODUCTION

Agriculture forms the bedrock of our nation, with approximately 70% of the population relying on it for their livelihood. Farmers have a vast array of crop options to choose from, allowing them to select the most suitable ones for their needs. The leaves are mainly damaged by weather, chemicals, mechanical, and nutritional problems. The leaf can be affected by pathogens like fungi, bacteria, viruses, phytoplasma, nematodes, and insects. Plant diseases significantly reduce agricultural production in terms of both quality and quantity. Fungi have primarily occurred on leaves but sometimes also may occur on stems or/and fruit. Usually, fungi produce spores in the leaf. This leads to the disease called spots. But sometimes leaf spots can be caused by bacteria. Leaf spots are definite spots of varying shapes, colors, and sizes. Bacteria can transfer the disease from one plant to another plant by rain, wind, and insects. The fungus Cochliobolous miyabeanus causes the brown spot on the leaf. It can observe the damages on leaves in big spots. This can able to kill the whole leaf. The bacterial blight is caused by bacteria. The name of the bacteria is Xanthomonous oryzea. Bacterial blight is mostly observed when there is strong wind and heavy rain. A bacterial leaf streak is a bacterial disease. Leaf streak can transfer the diseases from one leaf to another leaf. Bacterial leaf streak is also similar to the narrow brown spot. The blast is a fungal disease caused by Magnaporthe oryzae. Blast can affect the leaf, node, collar. It occurs where there is a cool temperature and low soil moisture.

The proposed system implements the technique used to identify the diseases on aleaf. Farmers are facing a low yield on the crop because it's difficult to identify diseases on the plant. Consequently, by using image processing methods to identify leaf disease, the following goals are pursued:

- Identifying diseases in plant images.
- Extracting the characteristics of a diseased leaf.

Grenze ID: 01.GIJET.10.2.752_1

II. LITERATURE SURVEY

Pavithra S et.al[1] have proposed the detection of an paddy leaf diseases using an SVM classifier. The input image is in RGB format. Removing a noise of the image and smoothening filter is applied at the pre-processing stage. Utilizing the K-means algorithm to perform image clustering. GLCM is used for extracting the feature of the input image. To classify the distinct kind of leaf disease, the SVM classifier is used.

R P Narmadha et.al[2] has proposed Analyzing symptoms of paddy leaf diseases through image processing begins with image preprocessing, which includes enhancing the image. The initial leaf image measures 109 by 310 pixels. To enhance accuracy, the RGB image was converted into a grayscale. The provided image is then used to create an occurrence matrix while accounting for probability distribution. The classification of leaf diseases is accomplished using the KNN algorithm.

M P Vaishnavi et.al[3] Within groundnut leaves, it is possible to identify five distinct disease types. The process of acquiring an image involves capturing a leaf image using a camera. Enhancements are made to improve the image quality and eliminate any unwanted noise from the groundnut image. The binary masked image is then transformed into an HSV image. The classification of the leaf disease was accomplished through an utilization of a KNN classifier. In this proposed system, a dataset comprises the total of 250 images, with 45 designated for a training and 105 for a testing. Sachin D et.al[4] Utilizing image processing techniques, paddy leaf crops are identified. In this approach, the leaf crop's color is matched against a leaf color chart for comparison. A robust method is employed to predict the color of paddy leaves. Detecting diseases in plant leaves is pivotal for minimizing agricultural product yield loss.

S M Mohidul Islam et.al[5] Predicting remedies for paddy leaf diseases involves employing Support Vector Machine (SVM). A leaf image was first enhanced by applying a median filter to make it smoother. Based on a color variations, the image is divided into a three clusters of images using a K- means algorithm for segmentation. Grayscale leaf occurrences serve as the basis for feature extraction. The SVM classifier is then applied for a classification of the leaf diseases and to recommend suitable pesticides. Rashedul Islam et.al[6] To determine a percentage of the paddy leaf damage through image processing, we utilize thresholding methods like Otsu's method for segmenting the leaf image. The proposed system accurately identifies the diseased region within the leaf.

N.Kavya, et.al[7] Identifying diseases on rice leaves employs deep learning techniques. Convolutional Neural Network (CNN) was utilized for the classification of the leaf diseases, with the specific use of an AlexNet Neural network model featuring five convolutional layers, three max-pooling layers, and three fully connected layers. The output image from each layer is transmitted as input to the subsequent layer, and the Rectified Linear Unit (ReLU) is employed within the fully connected layer. Melike Sardogan et,al[8] Leaf disease classification is achieved through the LVQ (Learning Vector Quantization) Algorithm. The Kohonen layer of the LVQ algorithm comprises 50 neurons, with 10 neurons assigned to each class. In this proposed system, the detection includes four distinct diseases, along with the classification of healthy leaves.

Amandeep Singh et.al[9] A color segmentation method is applied to a paddy plant leaf, involving color slicing to isolate the colors of interest and use them to create a mask for the targeted region. The approach employs Color Cube Component Selection. In this proposed technique, accuracy, sensitivity, and specificity metrics are computed.

III. METHODOLOGY

The system architecture was illustrated in Fig 1. The existing system is focused on a recognition of diseases on paddy leaf which benefits the farmers to increases the yield of paddy. The process of detecting paddy leaf diseases involves five key phases:

1. Image Acquisition
2. Image Preprocessing
3. Image Segmentation
4. Feature Extraction
5. Image Classification

A. Image Acquisition

The initial phase involves Image Capture, where an RGB color image of the paddy leaf was chosen as input to identify diseases. An input image is in the format of .jpg or .png. Each image is of 300×400 pixels.

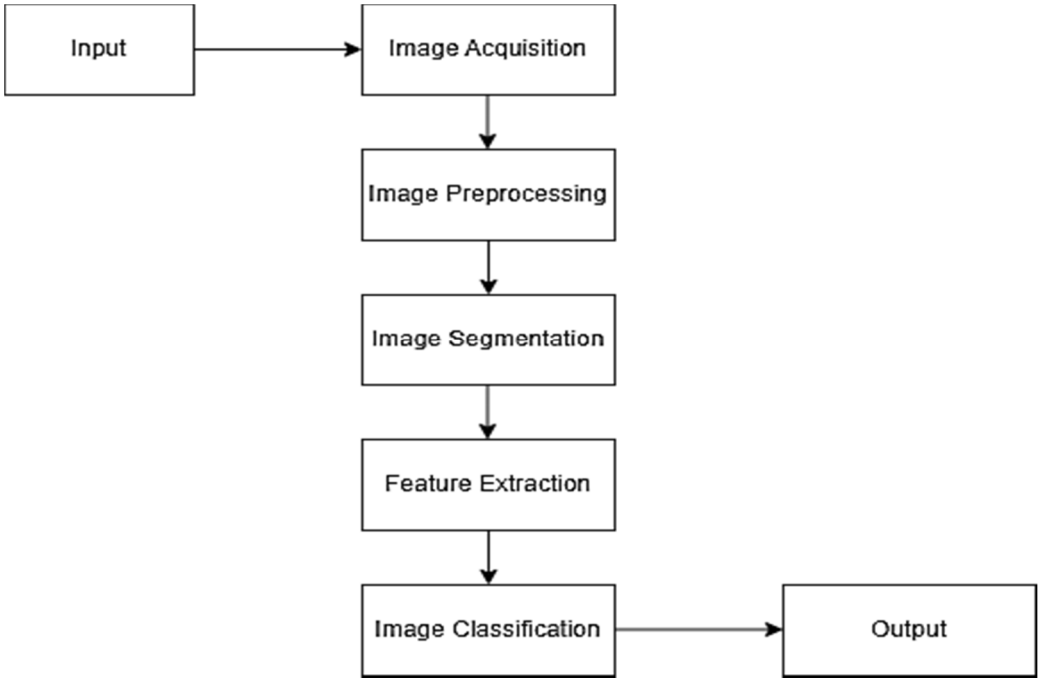

Fig 1: System Architecture

B. Image Preprocessing

During this stage, the input image undergoes processing for subsequent analysis, enhancing its overall image quality. The background of the leaf image is removed which is not necessary for the process. Theedge detection is used for identifying the boundaries within images and also findsthe discontinuity in the boundaries of an images. Edge detection was performed using the Sobel edge detection technique. The Sobel edge detection was calculated using the following equations 1 & 2;

Gm = (b6 +2b5+ b4) – (b4+2b1+ b2) Gn = (b2 +2b3+ b4) – (b4+2b7+ b6)

Here, Using kernels, Gm reflects a gradient with respect to X-direction and Gn reflects a gradient with respect to X-direction, and b1, b2, b3, b4, b5, b6, b7 denote the values within matrix a.

Image Segmentation

Image segmentation entails dividing image into numerous segments, and for this purpose, K-means clustering is employed, utilizing squared Euclidean distances. The K-means algorithm can be outlined in a following steps;

Step 1: Calculate a K, the required number of a clusters.

Step 2: Select a K centroids at random from the original positions

Step 3: Create a K clusters by connecting each data point to the closest centroid.

Step 4: Determine and place each cluster's new centroid. Reassign each data point to a nearest fresh centroid

Step 5: Return to step 4 if there is any reassignment; otherwise, the procedure is unaltered.

C. Feature Extraction

Features were essential in a field of image processing for dimensionality reduction. This transformation of input data into a set of features is particularly valuable when dealing with large input data sizes that may not contain a significant amount of information. In this feature extraction, a GLCM (Gray- level Co-occurrence matrix) was used for extracting the features from a image. In this proposed system, mean, variance, standard deviation, kurtosis, contrast, energy, entropy, and correlation are extracted as features.

Mean: The mean is evaluated by dividing the total number of pixels by the sum of the pixel values.

Mean: The mean is evaluated by dividing the total number of pixels by the sum of the pixel values.

$$m = \sum_{i=1}^{L-1} ri \; pri$$

The terms "m" stand for mean, "ri" for image intensity, and "pri" for intensity value probability.

RMS: By calculating a square root of mean of the squared data, root mean square (RMS) is obtained.

$$\frac{1}{\sqrt{mN}} \sqrt{\sum_{i=1}^{K} \sum_{i=1}^{K} (Iij - \mathrm{I})}$$

Iij represents the element at the ith row and jth column of a two-dimensional image with a size specified. I denote the average intensity calculated from all the pixel values within the image.

Variance: A measure of a contrast of an image.

$$\mu 2\,(r) = \sum_{i=1}^{L-1} (ri - m)2\mathrm{P}(ri\,)$$

Standard Deviation: It was the square root of a variance.

$$\sqrt{\sum_{i=1}^{L-1} (ri - m)2\mathrm{P}(ri\,)}$$

Kurtosis: The fourth moment serves as an indicator of the histogram's relative flatness.

$$\mu 4(r) = \sum_{i=1}^{L-1} (ri - m)3\mathrm{P}(ri\,) - 3$$

Contrast: It measures the overall intensity contrast between the pixel and its surrounding pixels.

$$\mathrm{Con} = \sum_{j=1}^{K} \sum_{j=1}^{K} (i - j)2 \qquad Pij$$

Energy: Uniformity is another name for energy, a metric of homogeneity in the [0, 1] range.

$$\mathrm{E} = \sum_{i=1}^{K} \sum_{i=1}^{K} P_{ij}^{2}$$

Correlation: A measurement of the degree of perfect positive and imperfect negative correlation between the pixel and its neighbours throughout the entire image.
Smoothing: Smoothing is performed by spatial and frequency filters.
The query picture is delivered to the classifier to identify paddy leaf diseases once the characteristics from the image have been extracted.

D. Image Classification

This is the final stage of the process is image classification. KNN (K-Nearest-Neighbor) algorithm was used for the classification of the disease on queryimages. It identifies the diseases like Bacterial blight, Brown spot, Blast. Here afollowed in the KNN algorithm.
Step 1 Begin by reading the input RGB image.
Step 2 Create an HSV picture from the RGB image.
Step 3 Obtain the value of K.
Step 4 Determine the type of distance metric (V) and the training data.
Step 5 Determine the distance between the training samples and the input sample.
Step 6 Make predictions regarding the disease name.

IV. Experimental Result

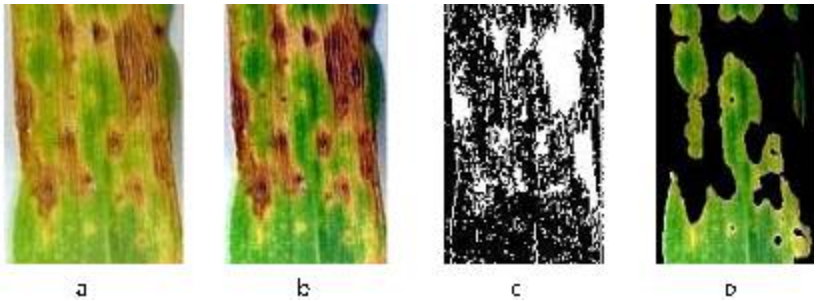

Fig 2: a) Original image b) Preprocessed image c) Edge detection for input image d) Segmented image

Fig 3 shows the cluster of three images in which one cluster image is selected to extract the features of the leaf. Fig 3a shows cluster1 image. Fig 3b shows image of cluster2. Fig 3c shows image of cluster 3.

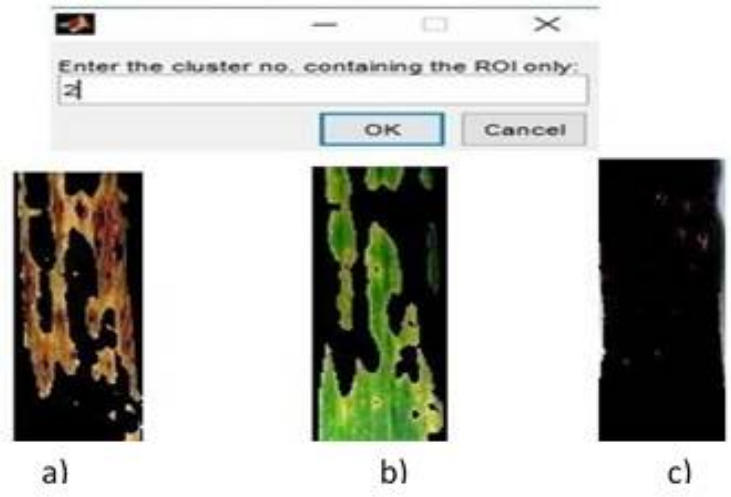

Fig 3: Cluster of images
a) cluster1 b) cluster2 c) cluster3

The extracted features in the form of mean, variance, standard deviation, kurtosis, contrast, energy, entropy, and correlation is shown in Fig 4.

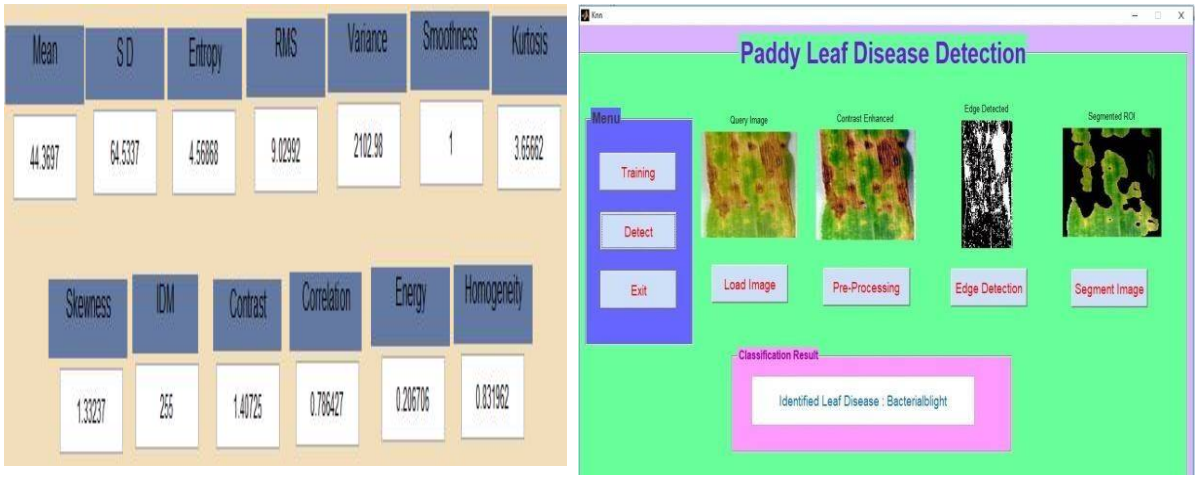

Fig 4: Feature Extraction of leaf image

V. Conclusion

The KNN classifier is used for classification. The classified disease is shown in Fig 5. For experimentationprocess 20 images are used for training, 8 images are used for testing and the accuracy rate is 98.39%. For the experimentation purpose 70% images were used for training, and 30% images were used for testing. By using the proposed system different plant diseases such as a brown spot, a bacterial blight, and a blast were detected. During experimentation, features are extracted from the query image. The paddyleaf disease diagnosis is performed using a classification approach. The image processing technique is used to identify thediseases of the paddy leaf. The proposed system gives the accuracy rate of 98.39%.

Acknowledgment

This work has been supported by the Exploration Focus of the Division of Medical Electronis, Sri Siddhartha Academy of Higher Education, Tumakuru, Karnataka, India.

REFERENCES

[1] S.Pavithra, A.Priyadharshini,V.Praveena and T.Monika, ”Paddy Leaf Disease Detection Using SVM Classifier”, International Journal of communication and computer Technologies, PP.16-20, Vol.03, karur, 26 June 2015.

[2] G.Arulvadivu and R.P.Narmadha,”Detection And Measurement of Paddy Leaf Disease Symptoms using Image Processing”, In Preceeding of InternationalConference on Computer Communication and Informatics, PP. 1-4, India, 5-7 January, 2017.

[3] M.P.Vaishnnav, K.Suganya Devi, and P.Srinivasa, ”Detection and Classification of Groundnut Leaf Diseases using KNN classifie”, In Preceeding of InternationalConference on Systems ComputationAutomation Networking, PP. 1-5, India, 29-30 March 2019.

[4] Sachin D Khirade, ”Plant Disease Detection Using Image Processing”, In Preceeding of International Conference on Computing Communication Control and Automation, PP. 768- 771, India, 26-27, February, 2015.

[5] Farhana Tazmim Pinki, Nipa Khatun, and S.M. Mohidul Islam, “Content-based Paddy Leaf Disease Recognition and Remedy Prediction using Support Vector Machine”, In Preceeding of International Conference of Computer and Information Technology,Bangladesh, 22-24 December2017.

[6] Rashedul Islam and Rafiqul Islam, “An Image Processing Technique to Calculate Percentage of Disease Affected Pixels of Paddy Leaf”, International Journal of Computer Applications, PP. 28-34, Volume 123, No.12, August 2015.

[7] N.Kavya, S.Naveen Kumar, and L.Yasaswi Venkat, N.Pranay Kumar,“Detection and Classification of Rice Leaf Diseases Using Deep Learning”, International Journal of Advanced Science and Technology, PP. 5868-5874, volume 29, Mylavaram, March 2020.

[8] Melike Sardogan, and Yunus Ozen, “Plant Leaf Disease Detection and Classification based on CNN with LVQAlgorithm”, In Preceeding of International Conference of Computer Science and Engineering, PP. 382-385, Bosnia, 20-23 September 2018.

[9] Amandeep Singh, and Maninder LalSingh, “Automated Blast DiseaseDetection from Paddy Plant Leaf - A ColorSlicing Approach”, In Preceeding of International Conference on Industrial Technology and Management, PP. 339- 344, UK, 7-9, March 2018.

BIOGRAPHIES OF AUTHOR

Raju A S is Professor, Department of BioMedical Engineering at Sri Siddhartha Institute of Technology, Tumakuru, India. He Holds a Ph.D degree with specialization in Biomedical image & signal processing from Visvesvaraya Technological University, Belagavi, India. His research areas are Biomedical Instrumentation, Biomedical Signal & Image Processing, Computer Vision, Pattern Recognition. Acadamic Council member at Sri Siddhartha Academy of Higher Education, Tumakuru. He is a recipient of Best Faculty award by Wipro Mission 10X, High Impact Teachiching Skills. Member of Biomedical Engineering Society of India, IETE, MIE, IBSC, ISTE and FIETE. He has chaired many international and national conferences and published research journals. He can be contacted at email: asraju74@gmail.com.