Rice Leaf Disease Detection Using MATLAB

[1]Priya Seema Miranda, [2]Jayalakshmi K P, [3]K Aarya Shri

[1,2,3]Assistant Professor
[1,2,3]Department of Electronics and Communication Engineering , St Joseph Engineering College,Mangalore,India

Abstract: Rice and wheat are two of the main crops produced in India, where agriculture is very important to the country's economy. 60% of people rely on crop farming for their livelihood. Of the Indian population and serves as a source of food for the rest of the globe, robust crop development and the early diagnosis of crop diseases are therefore crucial to ensuring that production is unaffected. Although the illness is not unique to leaves, it can extend to other plant components and have an impact on a plant's life cycle. For a wide range of crops, numerous disease detection, diagnosis, and classification techniques have been created and put to use. The ability to automatically identify plant leaf diseases is a crucial study area since it could help with monitoring vast fields of crops. The suggested system is a computer programme for identifying and identifying diseases in rice plant leaves. The technique consists of four basic steps: contrast adjustment for image processing, k-means clustering for image segmentation, GLCM for feature extraction, and KNN for image classification.

Keywords: K-means clustering, ROI, GLCM, KNN, GUI

1 Introduction

The impact of agriculture on the economy and the nation's GDP is significant. The majority of plants contract bacterial and fungal infections. The climatic circumstances have an impact on plant growth because of the exponential population increase tendency. The main obstacles are lowering the use of pesticides and raising their quality and quantity. Early diagnosis could boost output while reducing the need for pesticides. Farmers make up the vast bulk of the population, and their economic progress is based on the yield they produce. For a country to grow, it needs to produce goods of high quality. The growth of crops has recently been hampered by harsh weather or a shortage of water, which has caused crop damage.

In terms of both production and consumption, rice is one of the most important crops in India. Oryza Sativa is the scientific name for this kharif crop. It grows primarily in India's southern and eastern states. Many diseases, including Blast, one of the damaging diseases, can affect various areas of the plant and can impact rice at various phases of the harvesting process.

The traditional technique of detection involves visually inspecting the leaf to look for illness, although this method is not always reliable. The use of a computer-aided approach to detect the sickness is therefore advantageous. It is suggested to use a method for precise and automatic disease detection. This technique classifies the disease using a picture as input.

Damaged rice can greatly reduce yield. It is mainly caused by viruses, bacteria or fungus. The different diseases classified for the work are

A. Brown Spot

A plant fungal disease called brown spot of rice typically affects the host plant's leaves and glumes, as well as its seedlings, sheaths, stems, and grains. The symptoms of Cochliobolus miyabeanus can be very varied. Together with the presence of brown spot, many different components of the plant, including leaves, seeds, stems, and inflorescences, can be seen to exhibit general symptoms seen on the hosts. Another indication of rice disease caused by brown spot is stem discoloration. On the leaves of the host, oval-shaped brown patches with a grey centre are the telltale signs of fungal growth.

B. Blight caused by bacteria

Xanthomonas oryzae pv. oryzae is the reason behind Bacterial Blight. Seedlings wilt as a result, and leaves turn yellow and dry out. The disease is most prone to spread to areas with weeds and contaminated plant residue. It can appear in both tropical and temperate settings, especially in lowlands that get irrigation and rainfall. It is frequently seen when the disease can spread due to strong winds and persistent, heavy rainfall.

C. Leaf Smut

The fungus Mangnaporthe oryzae is responsible for the blast disease in rice. It affects the rice panicle's leaf, collar, node, and neck. They occur frequently in situations with low soil moisture, frequent rain, extended rainstorms, and chilly temperatures. In highland rice, the variation in temperature between day and night allows dew to collect on the leaves, and the generally cooler climate encourages the growth of this disease.

D. Tungro

A virus type belonging to the Caulimoviridae family's Ortervirales order is the tungro virus. In this genus, the rice tungro bacilliform virus is the sole species. Stunting, yellow to orange leaf discolouration, and fewer tillers are symptoms. Tungro, which is Filipino for "degenerated growth," was first noticed there.

2 Related Work

Using an SVM classifier, Shweta.S. Kothawale et al. [1] devised a method for detecting grape leaves. Using a histogram and thresholding to determine the leaf area, feature extraction is carried out using a GLCM and Random Transform Approach. SVM classifier compares the features obtained from training and testing to determine if a leaf is healthy or unhealthy. A strategy to help plants recover from infections was proposed by Saradhambal.G et al. [2] and used image processing to highlight the parts of the leaf that were afflicted.The process is guided by a voice navigation system as well.Histogram

matching was suggested as a method of accuracy by Shital Bankar et al. [3]. It consists of two phases: training and testing, during which a comparison approach is used to build a histogram for each phase. If a leaf is sick, it is compared to a healthy sample first to show the sickness before the test sample and healthy sample are compared again. A technique to identify the afflicted area of an infected arecanut picture was proposed by S Siddesha et al. [4]. For separating foreground from background, K-means is employed. The Otsu technique is used to find arecanut regions that are contaminated. This technique is strongly supported by theoretical approach.

Random forests classifier was used to develop the technique suggested by Shima Ramesh et al. [5]. Both classification and regression can be done using this classifier. This method provides greater accuracy with fewer data sets. Using HoG feature extraction, the feature vector is retrieved for the training dataset, and the random forest classifier is used to train it. The test dataset is used with the same methodology. For the purpose of estimating the outcomes, the features retrieved from the test image are contrasted with characteristics extracted from the trained classifier.

The method proposed by Minu Eliz Pothen et al. [6] employs various machine learning and image processing approaches. Using SVM, features from LBP and HOG are categorised. Three kernel functions—linear, polynomial, and radial basis function—are used during training.

To achieve precision, each kernel function is combined with LBP and HOG.Taohidul Islam et al. [7] proposed system detects and classifies disease using percentage of RGB value of diseased portion. In this system after pre-processing of image, masking of green pixel with blue pixel is done to differentiate affected and unaffected region of plant. Now the affected region is in its RGB form and sent for extraction of RGB percentage. These RGB values are classified one of the following classes; Color A, Color B, Color C and Color D. This percentage of different color is fed to Gaussian Naïve Bayes and using neural network disease are detected and classified. Vinaya Mahajan et al. [8] in this system the image obtained is resized, and then filtered using median filter to remove noise followed by segmentation using Otsu's model. Features extracted from images are contrast, dissimilarity, energy, homogeneity, entropy. Classification is done using fuzzy logic this method gives good accuracy. Pooja V et al. [9] have proposed a system using combination of various features. Segmentation is done using k-means and Otsu's detection, three clusters are formed and appropriate region is chosen by user. Various features are extracted which describes diseased region based on colour, shape and textural features. Classification is done using SVM. Amrita S. Tu;shan et al. [10] have proposed a comparative study using KNN and SVM classifier. Here five types of diseases are considered and converted to grey scale. The images are segmented using k-means because they are more noise immune. KNN classification shows exact disease name for any number of diseases. While linear is better for detecting two diseases not more than that. S. Pavithra et al. [11] have compared performance of both SVM and KNN classifier. Image is pre-processed to remove noise. Image segmentation is done using Otsu's method and k-means clustering. Feature extraction plays a major role in identifying the object. Kawcher Ahmed et al. [12] in this system a comparative study between four machine learning algorithm is made which includes KNN, Decision tree, Logistic regression and Naïve Bayes is made, where decision tree gave best accuracy. Suresh Metal. [13] proposed system detects only fungal diseases, RGB image is converted to HSV for segmentation. Classification is done using knn classifier with geometrical features like area, Major and minor axis and perimeter. M.P Vaishnanavel et al. [12] have proposed a system for detecting the disease and for disease management. Here images are acquired using mobile phone, the image is pre-processed to increase contrast, and input images are converted to binary images. In segmentation step binary masked image is converted to HSV image. Using fast feature method features are extracted from groundnut and classified using KNN.

3 Proposed Methodology

The system's architecture is as follows:

1. Image Acquisition
2. Image Pre-processing
3. Image Segmentation using K-means clustering
4. Feature Extraction using GLCM
5. Classification using KNN Classifier

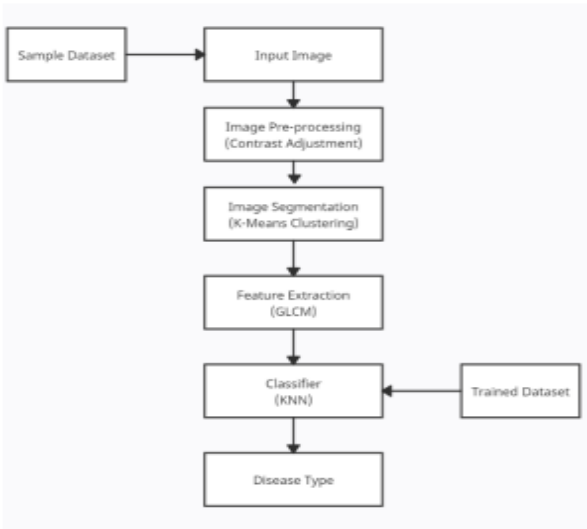

Fig 1: Flowchart

A. Image Acquisition

Image acquisition is the first stage of the image processing process. The system accepts input in the form of 2-D images. as input Images come in a variety of sizes, but they are all adjusted to the same size. Many pre-processing techniques are used to improve the quality and produce accurate results after the image has been acquired. At this step, the picture files of illnesses affecting rice leaves are acquired. This includes pictures of three different illnesses, including leaf blast, brown spot, and bacterial blight. There were 755 total photos utilised.

B. Image Pre-processing

Differentiating objects from their background is done using contrast adjustment. Using a single mapping function that is applied to every pixel, it remaps the original pixel intensities or colours into a new range of intensities or colours. Making the image's details visible is the goal of accomplishing this. The subject's border and outline, as well as any other components, must be visible. Prior to utilising the contrast modification approach to improve the photographs' quality, the images were first downsized to 250x250 as shown in Fig 2 and Fig 3.

Fig 2. Resized Image

Fig 3. Contrast Adjusted image

C. Image Segmentation

The sick part of the leaf and the healthy leaf region are separated using the K-means clustering technique. The number of clusters is indicated by the "K" in K-means clustering. The sick portion, which is the area of interest, is located in one of the clusters (ROI). This cluster ought to be chosen and used in the following steps. Reallocating each pixel to the closest clusters minimises the sum of distances and recalculates the centroids of the clusters, dividing the image into three areas. Different leaf picture parts make up each cluster. Using the findings from K means, each pixel in the image is labelled using index values from three clusters. Because cluster 2 has a greater ROI than the other two, it is chosen for further consideration.

Fig 4. Cluster Image

D. Feature Extraction

The dimensionality reduction method, which divides and condenses a starting set of raw data into smaller, easier-to-manage groupings, includes feature extraction. To make the procedure simpler, this is done. The fact that a huge number of variables are present in large data sets is one of their defining traits. Processing these variables takes a lot of computing power. So, by integrating variables into features, feature extraction significantly reduces the amount of data while still obtaining the optimal

feature from those enormous data sets. Although these qualities are simple to use, they nevertheless fall short of accurately and uniquely describing the real data set.

GLCM

The GLCM algorithm, which Haralick first introduced in 1973, is one of the most well-known methods for surface analysis. The basic concept is to create highlights based on the co-occurrence matrix of grey levels (GLCM). There can only be 13 features taken at once because the GLCM method requires 13 features. [9]

Feature extraction can be carried out using the Grey Level Co-occurrence Matrix, or GLCM. The Area of Interest is chosen from the three clusters, and then using the GLCM approach, various features are retrieved from it. The sick area is described by its colour, shape, and textural characteristics. The model is trained using a variety of features, including contrast, correlation, energy, entropy, homogeneity, IDM, kurtosis, mean, RMS, skewness, smoothness, standard deviation, and variance.

E. Classification using KNN Classifier

k-Nearest Neighbor is a Machine Learning algorithm based on Supervised Learning technique. It compares the similarity between the new data and the available data and puts the new data into the group/class that is most similar to the available ones. KNN algorithm stores all the available data. It is easy to classify new data points into the nearest similar class by using the algorithm. Each data point in the 'k' closest samples casts a vote and the class with the maximum votes wins.

The working of KNN algorithm can be explained as follows:

Step-1: Select the value of 'k' neighbors

Step-2: Estimate the Euclidean distance of the 'k' neighbors

Step-3: Take the 'k' nearest neighbors as per the calculated Euclidean distance.

Step-4: Among these 'k' neighbors, count the number of the data points in each class.

Step-5: Assign the new data points to that class for which the number of the neighbors is maximum.

There is no particular method to determine the best value for "k", so various values are trailed to find the best one. Noise will have a higher influence on the result if the value of 'k' is small i.e., the probability of overfitting is very high. A large value of K makes it computationally expensive and defeats the basic idea behind KNN (that points that are near might have similar classes). To optimize the results, we can use Cross Validation.

Using the cross-validation technique, we can test the KNN algorithm with different values of 'k'. The model which gives good accuracy can be considered to be an optimal choice. It depends on individual cases, at times best process is to run through each possible value of k and test our result.

The main advantages of using this algorithm are:

1. KNN is simple to implement.
2. Small training datasets are executed quickly.
3. Any prior knowledge about the structure of data in the training set is not required.
4. Retraining is not required if the new training data is added to the existing trained dataset.

Here, the MATLAB software's Classification Learner application is used for training purposes. An excel sheet is prepared and saved with the necessary datasheet of the characteristics used to categorise diseases. For the training dataset photographs and the testing dataset images, two excel sheets are made. The software is subsequently loaded with these files to perform the classification. The trained model is then exported to the workspace when classification is complete, and an image is then selected to be checked for the presence of one of the three diseases. It shows the precise location of the disease as well as the affected area.

4 Results and Discussion

The noise from the original RGB image was removed as the first phase in this research. K-means clustering was used for segmentation, and this technique produced three clusters as a result. These three clusters' Regions of Interest (ROI) were carefully scrutinised, and the infected cluster was found. For further phases, this cluster is utilised.

This cluster should be provided as an input to the Grey Level Co-occurrence Matrix (GLCM) method's feature extraction step. A matrix containing the various intensities present in an image is the outcome of this procedure. The output of this approach is sent as an input to the classification step, which is the last step. This procedure makes use of the KNN algorithm. The image is processed by the k-NN algorithm, which then labels it with the appropriate disease. Training was done using classifier learner app, which performs automatic training and seeks for the best classification model.

	Trained Images	Test Images
Bacterial Blight	200	25
Brown Spot	200	25
Leaf Blast	200	25
Tungro	65	15

Fig 5 Distribution of Images

In all, 755 pictures were used for this work. 665 of them were used for training purposes, while 95 were utilised for testing purposes. The model's testing yielded a 91% accuracy.

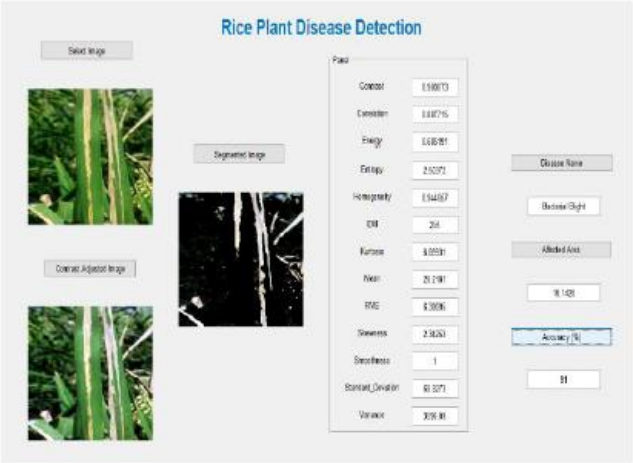

Fig. 6. GUI Output for Bacterial Blight

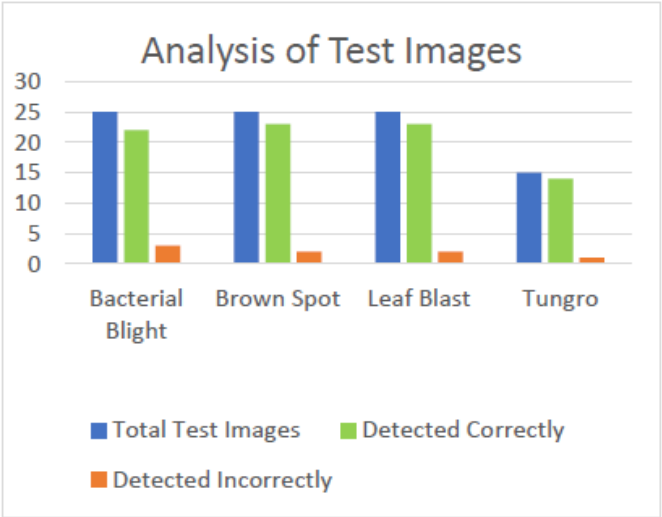

Fig. 7. Graph of Analysis of Test Images

The distribution of the photos among the four disorders is shown in the Fig. 7. It can be seen that 15 photos are checked for Tungro, and 25 images are examined for Bacterial Blight, Brown Spot, and Leaf Blast, respectively. Of 90 photos, 82 are accurately identified, of which 22, 23, 23, and 14 respectively, from bacterial blight, brown spot, leaf blast, and tungro.

5 Conclusion

The research makes use of a computer-aided system for classifying rice leaf diseases into four categories: bacterial blight, brown spot, leaf blast, and tungro. A total of 755 photos were used, of which 665 were used for testing and 90 for training. The pre-processing stage involved resizing and adjusting the contrast of the acquired photos. The best cluster among the three clusters was determined after these photos were further separated using the k-means clustering technique. With the GLCM approach, features were taken from the chosen cluster and utilised to train the pictures. In order to test the model using the KNN algorithm, test photos were used, and diseases were displayed.

The proposed system complies with the requirements, although there is room for improvement going forward. By training the model on more high-resolution photos that also show the medication for the specific ailment, the accuracy of the model can be increased. Software or an app can be created to make it simple for users and farmers to utilise.

References

1. Shweta S. Kothawale, S.R. Barbade and Pradnya .P.Mirajkar, “Grape Leaf Disease Detect- ion Using SVM Classifier”, International Journal of Innivative Research in Computer and Communication Engineering, Vol. 6, Issue 4, April 2018, pp. 3288-3294.
2. Saradhambal.G Dhivya.R, Latha.S and R.Rajesh, “Plant Disease Detection and its Solution Using Image Classification”, in International Journal of Pure and Applied Mathematics, Vol. 119, No. 14, 2018, pp. 879-884.
3. Shital Bankar, Ajita Dube, Pranali Kadam and Sunil Deokule, “Plant Disease Detection Techniques Using Canny Edge Detection & Color Histogram Image Processing”, in Shital Bankar et al, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol 5 (2), 2014, pp. 1165-1168.
4. S Siddesha and S K Niranjan, “Detection Of Affected Regions Of Disease Arecanut Using K-Means And Otsu Method”, in International Journal of Scientific and Technology Research Volume 9, Issue 2, February 2020, pp.3404-3408.

5. Shima Ramesh, Ramachandra Hebbar, Niveditha M, Pooja R, Prasad Bhat N, Shashank N and P V Vinod, "Plant Disease Detection Using Machine Learning", in International Conference on Design Innovations for 3Cs Compute Communicate Control, 2018, pp. 41-45.
6. Minu Eliz Pothen and Dr.Maya L Pai, "Detection of Rice Leaf Diseases Using Image Processing", in Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 424-430.
7. Taohidul Islam, Manish Sah, Sudipto Baral and Rudra Roy Choudhury, "A Faster Technique on Rice Disease Detection Using Image Processing of Affected Area in Agro Field", in Proceedings of the 2nd International Conference Mangaluru Inventive Communication and Computational Technologies (ICICCT), 2018, pp. 62-66.
8. Vinaya Mahajan and N.R. Dhumale, "Leaf Disease Detection using Fuzzy Logic", International Journal of Innovative research in Science, Engineering and Technology, Vol. 7, Issue 6, June 2018, pp. 6801-6807.
9. Pooja V, Rahul Das and Kanchana V, "Identification of Plant Leaf Diseases using Image Processing Techniques", in IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR), 2017, pp 130-133.
10. Amrita S. Tulshan and Nataasha Raul, "Plant Leaf Disease Detection using Machine Learning", in 10th ICCCNT, July 6-8 2019, IIT Kanpur
11. S.Pavithra, A.Priyadharshini, V.Praveena and T.Monika, "Paddy Leaf Disease Detection Using SVM Classifier", in International Journal of Communication and Computer Technologies, Vol. 3, Issue 1, 2015, pp. 16-20.
12. Kawacher Ahmed, Tasmia Rahman Shahidi Md. Irfanul Alam and Sifat Momen "Rice Leaf Disease Detection Using Machine Learning Techiques", 2019 International Conference on Sustainable Technologies for Industry 4.0
13. Suresha M, Shreekanth K N and Thirumalesh B V "Recognition of Diseases in Paddy Leaves Using kNN Classifier" in 2017 2nd International Conference for Convergence in Technology (I2CT) [14] M.P.Vaishnnave, K.Suganya Devi, P.Srinivasan and G.ArutPerumJothi "Detection and Classification of Groundnut Leaf Diseases using KNN classifier" in Proceeding of International Conference on Systems Computation Automation and Networking 2019.