VOL. 20, NO. 2-3, 2021, 13-17
www.elektrika.utm.my
ISSN 0128-4428

ELEKTRIKA
Journal of Electrical Engineering

# Paddy Plant Disease Detection Using Image Processing

**Goh Chi Yen** and **Leow Pei Ling***

School of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.

*Corresponding author: leowpl@utm.my, Tel: 607-5557170, Fax: 607-5566262

**Abstract:** Rice is the most significant human food crop and almost every Malaysian consume it daily. However, paddy plant diseases have significantly caused damage to the crop, reduced the rice production and threatening the food security in Malaysia. The traditional way of identifying the disease is inspecting it manually using naked eyes, which produce inaccurate results and it is time-consuming. Hence, development of image processing algorithm is vital to assist the farmers in detecting the diseases on paddy plant. In this paper, the Random Forest classifier is proposed to classify the three common paddy plant diseases which are rice blast, brown spot and narrow brown spot disease. The proposed method effectively detects and classify paddy plant disease with an overall accuracy of 93%.

**Keywords:** Classify, Image processing, Paddy plant disease, Random Forest



*Article History: received 25 May 2021; accepted 12 June 2021; published 15 October 2021*

## 1. INTRODUCTION

Rice or Oryza Sativa (scientific name) plays an important role in many Asian cuisine and it is the staple food for nearly one third of the world population [1]. Rice is not just solely a source of carbohydrate, but it is rich in nutrients, vitamins and minerals. From planting to harvesting, the yield of rice can be affected by factors such as soil fertility, weather condition and pests. However, in well-irrigated environments, the cause of severe rice production reduction is paddy plant diseases [2].

Proper plant disease diagnosis is essential to improve the rice productivity [2]. Other than relying only on the expert, image processing method has been proposed where farm operator can collect images and let the machine to do the diagnosis. Since image processing can process image faster, cost-effective and save human effort, hence image processing plays an important role in detecting plant disease [3], [4]. Without image processing techniques, farmers have to observe the paddy plants manually using naked eyes or pass the plant samples to experts for verification, which create tremendous work because paddy field has large area. Some of the farmers use excessive pesticide because they have no idea what type of disease the infected plant contains [5]. By using image processing, the paddy plant diseases can be detected and diagnosed in the early stage to facilitate the plant disease control through proper disease management. Hence, an effective paddy plant disease detection system with image processing is very beneficial for paddy farmers.

## 2. DETECTION OF PADDY PLANT DISEASES

Several studies have been conducted to detect paddy plant diseases using image processing techniques together with Artificial Intelligent. In [6], a paddy disease diagnosing system has been developed. The characteristics of paddy plant disease such as spot colour, boundary colour, broken paddy leaf colour and lesion type were extracted for classification. The lesion type is determined by measure the ratio of width and height of the lesion spot. Besides, two thresholding techniques, local entropy-based and Otsu method are applied to compare the result. The classification is done by utilizing the production rule developed with agricultural expert.

Besides, Pinki et al. reported a method of recognizing paddy leaf diseases through prediction technique [7]. There are two set of data in the system which are the training set and testing set. Training set is used to undergo training phase and train the model, whereas testing set is used in testing phase to test the trained model. In both phases, the input image is first pre-processed, then the disease part is segmented using K-means clustering. A total of eleven features based on shape, texture and colour are extracted and supervised learning-based classifier, Support Vector Machine (SVM) is implemented for classification.

Meanwhile Islam et al. in [5] is introducing a faster image processing technique in detecting rice diseases. The steps of rice diseases detection include image acquisition, image pre-processing and masking of green pixels with blue pixels. Next, the blue pixels containing unaffected

area are discarded. RGB percentage extraction and calculation is then done for remaining diseases affected area and finally classification using Gaussian Naïve Bayes together with neural network. By using this method, the rice diseases can be effectively detected using only a small portion of the diseased leaf. However, this method is limited if the RGB percentage for two plant diseases are similar, such as brown spot and narrow brown spot diseases.

Meanwhile, Pothen and Pai proposed a method to classify three paddy plant diseases, namely bacterial leaf blight, brown spot and leaf smut disease [8]. Otsu's method is applied in image segmentation, then the features from segmented images are extracted using Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). Lastly, SVM is applied to classify the paddy plant diseases.

In [9], another method for automatic paddy leaf diseases identification is proposed. The leaf image is first preprocessed, then the Fuzzy C-means algorithm is applied to detect the edge for image segmentation. The feature of segmented image is extracted based on the texture using Gray Level Concurrence Matrix (GLCM) and Speeded Up Robust Feature (SURF) techniques. The disease is then classified using Artificial Neural Network (ANN).

The research done by [10] presents different classification techniques for plant diseases detection. After images are preprocessed, K-means clustering technique and Hue Saturation Value (HSV) Alteration are used for image segmentation, and GLCM techniques for feature extraction. The result showed that Random Forest classifier have the higher accuracy rate compared to Kernel SVM, K-nearest neighbors and decision tree for plant disease classification.

In this paper, image processing with Random Forest classifier method is proposed for paddy plant diseases detection and classification. The proposed methodology includes image pre-processing, image segmentation using K-means clustering, and feature extraction based on colour, shape and texture of the diseased part.

## 3. PADDY PLANT DISEASES

Plant disease is defined as an occurrence of physiological disorder [5] that disrupts the growth of plant [6]. This section describes the symptoms of the three common paddy plant diseases, which are rice blast, brown spot and narrow brown spot disease.

Rice blast is caused by the fungus *Magnaporthe Oryzae*. Infection of rice blast can occur on leaves, collars, necks and seeds of paddy plant [11], and potentially results in crop death. Figure 1 shows the rice blast disease.

Figure 1. Rice blast diseae

From Figure 1, the lesions have elliptical or spindle shape. The colour of the lesions is red to brownish for borders while whitish to grey for centres [7].

As for the brown spot disease. The spot on the paddy leave is caused by the fungus *Bipolaris Oryzae*, which result in unfilled grains. Figure 2 shows the brown spot disease.

Figure 2. Brown spot disease

Initially, the shape of the lesions is small and circular with the colour of dark brown to purple brown [7]. It is then surrounded by reddish brown margin after the lesions fully developed [12]. Meanwhile, narrow brown spot is caused by the fungus *Sphaerunlina Oryzina*. Figure 3 shows the narrow brown spot disease.

Figure 3. Narrow brown spot disease

The shape of the lesions has a width of approximate 1 to 1.5mm wide with 2 to 10mm long. It tends to be parallel with the veins of paddy leaf with dark brown colour [12].

## 4. PROPOSED METHOD

The proposed methodology consists of two phases for training and testing purposes. Firstly, image acquisition is done to collect all the paddy plant images needed. The dataset is then split into training set and testing set. An algorithm is developed in training phase to observe the patterns in training set, and a Random Forest model is constructed to learn these patterns and their corresponding output. The testing set that undergo testing phase is then fed into the Random Forest model for paddy plant diseases prediction and performance evaluation. MATLAB R2017b is implemented to develop and execute the system. Figure 4 shows the flowchart of the system design.

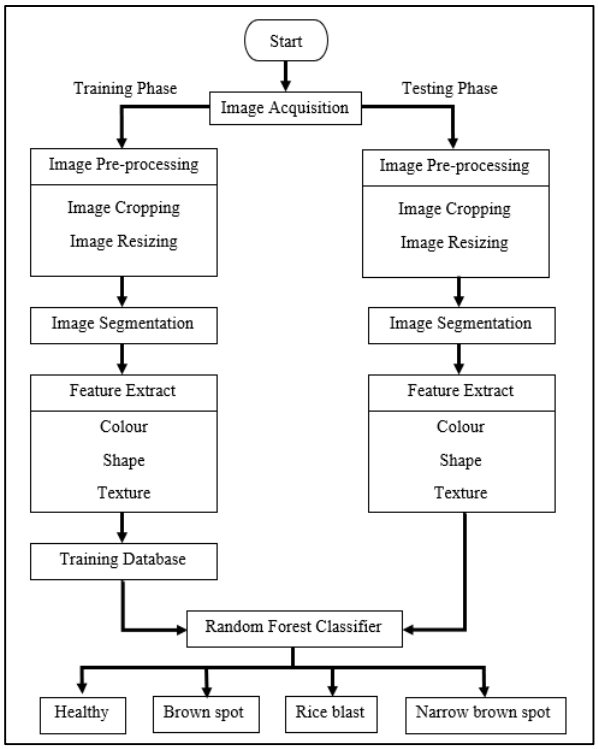


Figure 4. Flowchart of the system design

Features such as colour, shape and texture are extracted from the images and feed into the Random Forest Classifier where the extracted features of the data are to be classified to be healthy plant or with the three diseases.

### 4.1 Image Acquisition

First, image acquisition is done to acquire all the images needed for the dataset of this system. The images are collected from Internet which include images of rice blast, brown spot, and narrow brown spot disease, as well as healthy paddy plant images.

### 4.2 Image Pre-Processing

To improve the image data for further processing, image pre-processing techniques such as image cropping and image resizing are performed. Every image is cropped manually to remove the unwanted part of the images. Next, the image is resized to 256 × 256 pixels to standardize the image resolution and reduce the computational complexity.

### 4.3 Image Segmentation

Colour-based segmentation is performed to get the diseased part of the paddy plant from the images. The original colour space for image is RGB, which contains three colour components (red, green and blue) for each individual pixel of image. To distinguish these colours easily, colour transformation is performed to convert the RGB image into L*a*b* colour space. Layer "a*" is defined as the colour fall along the red-green axis while layer "b*" for colour fall along the blue-yellow axis [13]. Since the a*b* layers contain all the information, hence these layers are used for classification using K-means clustering technique.

Three clusters are created for the K-means cluster, and one of the clusters contains the disease. The distance between two colours (a* and b* values) are calculated using Euclidean Distance Metric, and K-means clustering is used to assign the value to the nearest cluster. The result from K-means is then used for labelling of every pixel in the image.

### 4.4 Feature Extraction

Every paddy plant disease has different symptoms, therefore the features are important parameters for classification. In this system, features based on the texture, colour and shape of the segmented image are extracted. The texture extraction is done using Gray-Level Co-Occurrence Matrix (GLCM) technique, which is a method that computes the spatial relationship between pixels [14]. Gray scale conversion of selected cluster image is done before analysed using GLCM. By using the statistics of GLCM, image information such as correlation, homogeneity, contrast, and energy are derived. Entropy which known as the randomness of an image is also extracted.

Colour features are the most suitable features to describe the lesions of paddy plant disease. Colour moments which are measures that indicate the colour distribution of an image by using the foundation of mathematics [7] are the simple ways to extract the colour features. Hence, colour moments such as mean and standard deviation are used in this system. Mean is defined as the average colour value of an image while standard deviation is the square root of variance [7]. Besides, variance and root mean square are used to extract the colour features.

Shape features such as area, eccentricity and perimeter are extracted as they represented the physical structure of the diseased part of paddy plant image. These shape features are measured using the binary image of diseased part.

After total of 12 features are extracted, a feature vector for each image is created and store into database. In training phase, the features vectors are stored into training set for training purpose in classification. In testing phase, the features vectors are stored into testing set to test the model.

### 4.5 Classification

Classification is applied to classify the dataset into classes which they belong to [9]. Figure 5 shows the structure of Random Forest classifier.

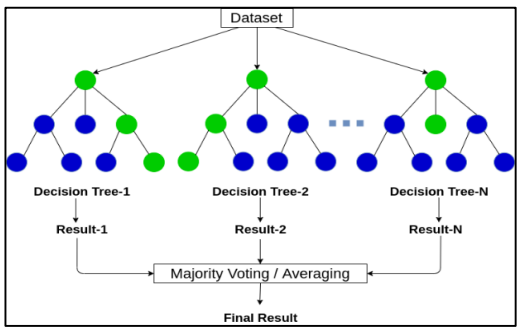


Figure 5. Structure of Random Forest model

Random Forest classifier which is a supervised machine learning model is implemented to classify the paddy plant diseases. Random Forest performs the classification task by constructing multiple independent decision trees during training time. Then, training samples are selected by bootstrap sampling randomly from the training set. Every decision tree contains of one sample, and features are selected randomly to split the nodes of the decision tree [15]. Out-of-bag error is estimated to measure the prediction error of Random Forest. Testing set is then used to test the trained Random Forest model. The dataset is fed into the model, and the model selects and predicts the result with the most votes.

## 5. RESULT AND ANALYSIS

A total of 180 images of different format, size and resolution are collected from Internet based on the three paddy plant diseases, as well as healthy paddy plant images. The dataset is then split into 120 images for training and 60 images for testing the classifier. Figure 6 shows the result of image segmentation for rice blast disease.

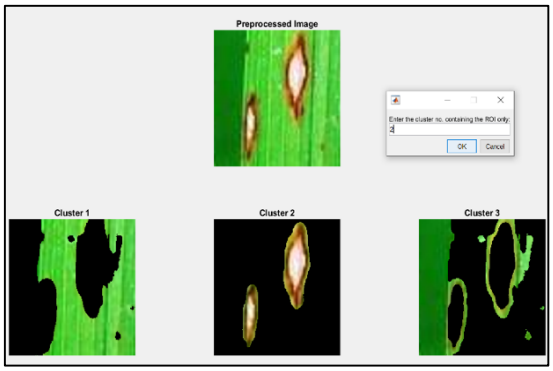


Figure 6. Image segmentation result for rice blast disease

All the images are first cropped and resized. Then, K-means clustering is applied to segment diseased part from the images. From Figure 6, Cluster 2 is selected to process further as it contains the diseased part. Figure 7 shows the out-of-bag error plot for the trained Random Forest model.

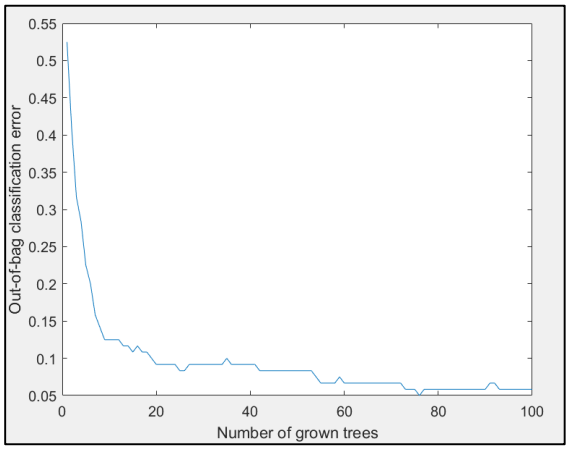


Figure 7. Out-of-bag error plot

After extracting the features of selected cluster image, Random Forest is implemented to classify the paddy plant disease. During the training time, the out-of-bag classification error is estimated as 0.0583 with 100 decision trees grown. The predicted result for the selected cluster image of Figure 6 is shown in Figure 8 as a pop out message.

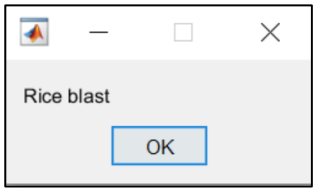


Figure 8. Predicted result for selected cluster image of the rice blast disease.

The accuracy of detection and classification of paddy plant disease using Random Forest classifier is shown in Table 1.

Table 1. Random Forest classifier performance

| Paddy plant sample | Classification result | | | | Accuracy |
|---|---|---|---|---|---|
| | Brown spot | Rice blast | Narrow brown spot | Healthy | |
| Brown spot | 14 | 0 | 1 | 0 | 93.33% |
| Rice blast | 0 | 14 | 1 | 0 | 93.33% |
| Narrow brown spot | 2 | 0 | 13 | 0 | 86.67% |
| Healthy | 0 | 0 | 0 | 15 | 100% |
| Overall | | | | | 93.33% |

The system effectively detects and classifies the paddy plant disease with an overall accuracy of 93.33%. The performance is greatly affected due to the similar texture between brown spot and narrow brown spot disease. Approximately 13.33% of narrow brown spot disease are wrongly classified into brown spot disease. The system can be improved by expanding the dataset images as only 15 images for each disease are used to test this system.

## 6. CONCLUSION

Three type of paddy plant diseases have been successfully classified by using image processing and Random Forest Classifier technique. The developed algorithm detects and classify the three common paddy plant diseases namely rice blast, brown spot and narrow brown spot disease through the image segmentation process using K-means clustering technique. Then a total 12 features based on colour, shape and texture are extracted and classification is implemented using Random Forest classifier. The method

is tested and can effectively detect and classify the paddy plant diseases with an overall accuracy of 93%. With this method, farmers can easily capture the paddy plant images on the spot with their mobile phone and identify the plant disease accurately at the early stage.

## ACKNOWLEDGMENT

The authors would like to acknowledge and thank the Universiti Teknologi Malaysia for all financial support through Research University Grants vote Q.J130000.2451.04G93, Q.J130000.2551.20H93 and Q.J130000.2451.09G14.

## REFERENCES

[1] Payal and P. Kukana, "Hybrid machine learning algorithm-based paddy leave disease detection system," in *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020*, 2020, no. Icosec, pp. 512–519.

[2] M. Suresha, K. N. Shreekanth, and B. V Thirumalesh, "Recognition of diseases in paddy leaves using knn classifier," in *2017 2nd International Conference for Convergence in Technology (I2CT)*, Apr. 2017, pp. 663–666.

[3] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Comput. Electron. Agric.*, vol. 153, no. April, pp. 69–81, 2018.

[4] S. Phadikar and J. Sil, "Rice disease identification using pattern recognition techniques," in *Proceedings of 11th International Conference on Computer and Information Technology, ICCIT 2008*, 2008, no. Iccit, pp. 420–423.

[5] T. Islam, M. Sah, S. Baral, and R. Roychoudhury, "A faster technique on rice disease detection using image processing of affected area in agro-field," in *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, 2018, no. Icicct, pp. 62–66.

[6] F. T. Pinki, N. Khatun, and S. M. M. Islam, "Content based paddy leaf disease recognition and remedy prediction using support vector machine," in *20th International Conference of Computer and Information Technology, ICCIT 2017*, 2018, vol. 2018-Janua, pp. 1–5.

[7] M. E. Pothen and D. M. L. Pai, "Detection of rice leaf diseases using image processing," *Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020*, no. Iccmc, pp. 424–430, 2020.

[8] G. Jayanthi, K. S. Archana, and A. Saritha, "Analysis of automatic rice disease classification using image processing techniques," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 3, pp. 15–20, 2019.

[9] P. Panchal, V. C. Raman, and S. Mantri, "Plant diseases detection and classification using machine learning models," *CSITSS 2019 - 2019 4th Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. Proc.*, pp. 1–6, 2019.

[10] M. S. G. Premi, R. Narmadha, and T. Bernatin, "A brief survey on diseases of paddy plant," *J. Pharm. Sci. Res.*, vol. 11, no. 7, pp. 2739–2743, 2019.

[11] R. P. Narmadha and G. Arulvadivu, "Detection and measurement of paddy leaf disease symptoms using image processing," in *2017 International Conference on Computer Communication and Informatics, ICCCI 2017*, 2017, pp. 7–10.

[12] G. Saradhambal, R. Dhivya, S. Latha, R. Rajesh, "Plant disease detection and its solution using image classification," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 14, pp. 879–884, 2018.

[13] R. Meena Prakash, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari, and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," in *Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 44 2017*, 2018, vol. 2018-Janua, pp. 1–4.

[14] X. Mai and M. Q. H. Meng, "Automatic lesion on random forest," in *2016 IEEE International Conference on Real-Time Computing and Robotics, RCAR 2016*, 2016, pp. 255–259.