

## Automatic classification of paddy leaf disease

Shafaf Ibrahim<sup>1</sup>, Nurnazihah Wahab<sup>2</sup>, Ahmad Firdaus Ahmad Fadzil<sup>3</sup>,  
Nur Nabilah Abu Mangshor<sup>4</sup>, Zaaba Ahmad<sup>5</sup>

<sup>1,2,3,4</sup>Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka  
(Kampus Jasin), Malaysia

<sup>5</sup>Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Perak  
(Kampus Tapah), Malaysia

---

### Article Info

#### Article history:

Received Jan 28, 2018

Revised Apr 26, 2019

Accepted Mei 15, 2019

---

#### Keywords:

Automatic classification paddy  
leaf disease

Feature extraction

SVM

---

### ABSTRACT

Rice is a staple food in most of the Asian countries. It is an important crop, and over half of the world population relies on it for food. However, paddy leaf disease can affect both the quality and quantity of paddy in agriculture production. The classification of paddy leaf disease is an important and urgent task as it destroys about 10% to 15% of production in Asia. Thus, a study on automatic classification of paddy leaf disease using image processing is presented. Feature extraction techniques of color, texture, and shape were implemented to analyze the characteristics of the paddy leaf disease. In another note, a Support Vector Machine (SVM) is used to classify the four types of paddy leaf disease which are the brown spot, bacterial leaf blight, tungro virus, and leaf scald. The performance of the proposed study is evaluated to 160 testing images which returned 86.25% of classification accuracy. The outcome of this study is expected to assist the agrotechnology industry in early detection of paddy leaf disease in which an appropriate action could be taken accordingly.

Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Shafaf Ibrahim,  
Faculty of Computer and Mathematical Sciences,  
Universiti Teknologi MARA Cawangan Melaka (Kampus Jasin),  
77300 Merlimau, Melaka, Malaysia.  
Email: shafaf2429@uitm.edu.my

---

## 1. INTRODUCTION

Rice is a staple food in most of the Asian countries. Paddy covers around 69% of the cultivated area, and the main field covers around 63% of the total region under the food grains [1]. However, there are many factors that make paddy rice production become slow and less productive. One of the main factors is paddy leaf disease [2].

The paddy leaf disease might be caused by the bacteria, viruses, and fungi [3]. The disease on the paddy leaf may have some similar symptoms which lead to confusion in classifying the disease [4]. Thus, an early stage diagnosis of paddy leaf disease may need more money and a lot of time [5].

Classification of paddy leaf disease is an important and urgent task. It can affect both quality and quantity of paddy in agriculture production [6], [7]. The common disorder found in the paddy usually appears at the panicle initiation stage which shows on the paddy leaves [8]. It is due to mineral deficiency and infections caused by the pest, and it is visualized by discoloration and dead spots on the paddy leaves. Thus, it is beneficial to classify the paddy leaf disease by the symptoms found on the surface of the paddy leaf.

Lesion area and leaf area of the paddy leaf diseases are frequently measured by the ratio [9]. In making sure that the leaf diseases do not affect the production, the management should keep a close supervision of the crops [10]. These diseases occur naturally, and their symptoms differ

extremely. Plant scientists should keep a track on the estimation of the damaged plant by keeping an eye on the percentage of the affected area [9].

Traditionally, the paddy leaf diseases are identified using a naked eyes observation method [11]. There are visually classified by the experts by identifying the changes in the paddy leaf color. However, different experts may classify the same part as a different disease. Thus, to increase accuracy, a paper grid method is used. Yet, the method is found to be laborious, time-consuming and impractical for the large field [12].

Based on the problems discussed, the visual recognition of diseases on leaves is observed to be less accurate and it requires more experienced workers. Thus, a fast and accurate approach to classify paddy leaf disease is highly needed. Therefore, a study on automatic classification of paddy leaf disease using image processing technique is proposed. The image processing technique is very effective and dependable day by day. Feature extraction techniques of color, texture, and shape were implemented to analyze the characteristics of the paddy leaf disease. Whereas, a Support Vector Machine (SVM) technique is used to classify the four types of paddy leaf disease which are the brown spot, bacterial leaf blight, tungro virus, and leaf scald. The outcome of this study is expected to assist the agrotechnology industry in early detection of paddy leaf disease where an appropriate action could be taken accordingly. In another note, the automatic classification is an extra advantage as it may reduce a large work of monitoring in the large crop of paddy.

## 2. RESEARCH METHOD

The aim of this study is to automatically classify the paddy leaf disease using image processing technique, and to evaluate the performance of the disease classification. Figure 1 depicts the proposed process flow of the automatic classification of paddy leaf disease.

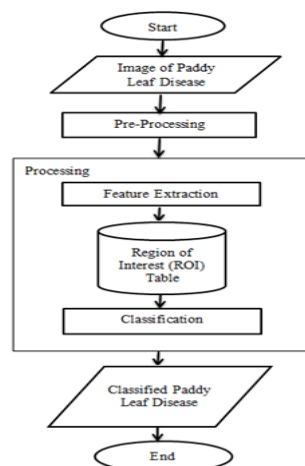






Figure 1. The proposed process flow of automatic classification of paddy leaf disease

The proposed process of automatic classification of paddy leaf disease begins with the insertion of paddy leaf disease image. The inserted image will go through the two main stages which are pre-processing and processing. The pre-processing includes image enhancement. On the other hand, the processing stage comprises of two sub-processes which are feature extraction and classification. During the feature extraction process, the features of each paddy leaf disease will be extracted. The process is used to study the characteristics of each Region of Interest (ROI) which consequently produced the ROI table. Next, the classification process will automatically classify the paddy leaf disease which produces the final outcome of the classified paddy leaf disease subsequently. The detail explanation of each process involved is elaborated further in the next subsections.

### 2.1. Testing Images

Hundred and sixty testing images of four types of leaf diseases which are bacterial leaf blight, brown spot, tungro virus and leaf scald were collected. Table 1 tabulates the sample images for each type of paddy leaf disease as mentioned.

Table 1. Types of Paddy Leaf Disease


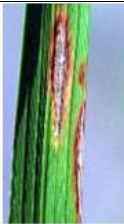
Type	Sample Image
Bacterial Leaf Blight	
Brown Spot	
Tungro Virus	
Leaf Scald	

**2.2. Pre-Processing**

The different image may have different edges, has locally varying statistics and smoothness in it [13]. Thus, image enhancement plays an important role in the field of image processing. It is used to improve the visibility of low-contrast features, and the digital quality of the image [14]. The quality of the image is utilized to judge whether the image is capable enough for utilization.

A method of contract stretching is proposed for the image enhancement. It is a simple image enhancement technique that attempts to improve the contrast in an image. It is done by stretching the range of intensity values it contains to span the desired range of values [15]. Table 2 shows the implementation of contrast stretching on paddy leaf disease sample image.

Table 2. Contrast Stretching

Before enhancement	After enhancement
	

**2.3. Processing**

The processing stage involved two sub-processes which are feature extraction, and classification.

**2.3.1 Feature Extraction**

Feature extraction assumes as an essential part for recognition of an object [16]. Feature extraction techniques of color, texture and shape were implemented to analyze the characteristics of the paddy leaf disease. Different techniques of Color Moments, Grey Level Co-Occurrence Matrices (GLCM) and Regionprops were proposed for color, texture, and shape extraction respectively. The flowchart of feature extraction processes is illustrated in Figure 2.

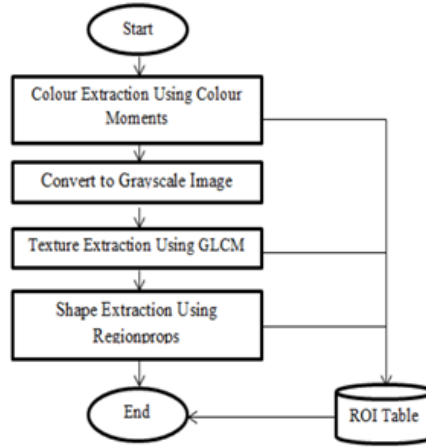




Figure 2. Feature extraction processes

The feature extraction process begins with the color feature extraction using a technique of Color Moments. The Color Moments is one of the simplest yet very effective features. It is utilized to separate the pictures based on the feature of the colour [17]. These strategies offer a measurement for colour similarity between the pictures [13]. Out of a few parameters in color moments, the standard deviation values for red, green and blue (RGB) as in (1) were selected due to its simplicity.

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - u_i)^2\right)^{1/2} \tag{1}$$

The next process is grayscale conversion. Grayscale images are distinct from one-bit bitonal black-and-white images, which in the context of computer imaging are images with only the two colors, black and white [18]. Grayscale images have many shades of gray in between. In this part of study, the grayscale conversion is necessary for extracting the texture features. Table 3 depicts a sample of grayscale conversion on paddy leaf disease sample image.

Table 3. Grayscale Conversion

Enhanced image	Grayscale conversion
	

The second feature extraction is texture. The texture of the paddy leaf disease is distinguished by a powerful texture extraction technique which is GLCM. A GLCM indicates the probability of a gray-level  $i$  occurring in the neighbourhood of gray-level  $j$  given distance  $d$ , angle  $\theta$  and the total number of gray levels  $N$  [19]. There are three parameters of GLCM selected which are contrast, homogeneity, and correlation. Table 4 illustrates the details and equations of the three selected features.

Table 4. GLCM Equations

Features	Details	Equations
Contrast	Represents the amount of local gray level variation in an image.	$\sum_{n=0}^{Ng-1} (i - j)^2 \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} p(i, j)$ (2)
Homogeneity	Also known as Inverse Difference Moment. It is high when the local gray level is uniform and inverse GLCM is high.	$\frac{\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p_{ij}}{1+(i-j)^2}$ (3)
Correlation	Measures the linear dependency of grey levels of neighbouring pixels.	$\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i, j)p(i, j) - \mu_x \mu_y$ (4)

The third feature is shape. A shape is an important case to identify and recognize an object, which it is the purpose to encode simple geometrical forms [20]. It is used to calculate the area, perimeter, circularity and to calculate using connected regions in the image [21]. Regionprops technique is employed to measure the properties of a selected region of an image in pixel count [22]. There are three features extracted which are centroid, major axis length, and minor axis length. A function called *bwboundaries* is used to trace the boundary of a selected region in an image. The function *imfill* is used to fill any holes so that Regionprops can be used to estimate percentage area enclosed by each of the boundaries. The details of the shape features extracted are tabulated in Table 5.

Subsequently, the extracted feature values of color, texture, and shape were collected and summarized in the ROI table which includes the minimum and maximum range values of all the paddy leaf disease features as in Table 6. The ROI table is acted as a feeder in the SVM classification afterward.

Table 5. Parameters of Regionprops

Parameter	Details
Area	Represents a number of white pixels in a binary image.
Major Axis Length	The length (in pixels) of the major axis of the ellipse that has the same second-moments as the region.
Minor Axis Length	The length (in pixels) of the minor axis of the ellipse that has the same second-moments as the region.

Table 6. ROI Table

Type	Features	Range Value	
Brown Spot	Colour	Standard Deviation Red	35.941 - 87.517
		Standard Deviation Green	40.391 - 97.2
		Standard Deviation Blue	36.631 - 90.788
	Texture	Contrast	0.0856974 - 1.58631
		Homogeneity	0.760194 - 0.957238
		Correlation	0.845 - 0.984995
	Shape	Area	60300 - 6822140
		Major Axis Length	346.41 - 3510.29
		Minor Axis Length	232.095 - 2992.98
Bacterial Leaf Blight	Colour	Standard Deviation Red	43.456 - 94.622
		Standard Deviation Green	36.633 - 87.867
		Standard Deviation Blue	36.235 - 102.827
	Texture	Contrast	0.0643622 - 0.850459
		Homogeneity	0.785532 - 0.967881
		Correlation	0.824413 - 0.988628
	Shape	Area	49278 - 1127680
		Major Axis Length	271.355 - 1385.64
		Minor Axis Length	193.99 - 923.76
Leaf Scald	Colour	Standard Deviation Red	38.704 - 92.487
		Standard Deviation Green	36.597 - 90.365
		Standard Deviation Blue	36.192 - 103.462
	Texture	Contrast	0.164881 - 2.51397
		Homogeneity	0.694004 - 0.971598
		Correlation	0.790685 - 0.980476
	Shape	Area	23616 - 665052
		Major Axis Length	123.553 - 1087.73
		Minor Axis Length	81.9837 - 815.219
Tungro Virus	Colour	Standard Deviation Red	41.258 - 74.019
		Standard Deviation Green	42.695 - 74.358
		Standard Deviation Blue	43.509 - 82.857
	Texture	Contrast	0.146147 - 5.26955
		Homogeneity	0.558186 - 0.929856
		Correlation	0.571602 - 0.982052
	Shape	Area	22500 - 2662000
		Major Axis Length	173.205 - 1385.64
		Minor Axis Length	136.255 - 1536.91

### 2.3.2 Classification

Classification refers to a process of sorting information obtained from the images into classes [23]. In this part of study, the classification is a process where it determines the type of paddy leaf disease of the inserted paddy leaf image based on the extracted features beforehand. The SVM is implemented as it is a useful machine learning tool for classification [24]. It classifies the given data samples (in the form of vectors) by mapping them to high dimensional spaces and constructs hyper-planes that divide the

data into partitions. The data belonging to the same class will be put into the same partition with high probability and those with different class's likely end up in the different partitions [25].

The *trainImageCategoryClassifier ()* function is used to create the image category classifier. Each element of the image sets defines an image category. Two classes of data which are training data and testing data are built separately for the purpose of maximizing the distance. The SVM is trained and the support vectors (SVs) are obtained for each class. Next, the testing phase calculates the average distance between the test samples and SVs for each class. Finally, the type of paddy leaf class is decided based on the minimal average distance of the test samples. This procedure is repeated until all images are classified.

#### 2.4. Performance Evaluation

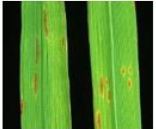



The performance of the paddy leaf disease classification is evaluated using a truth table. It is performed by comparing the disease classification result with the actual disease. Based on the truth table obtained, the classification accuracy for each type of paddy leaf disease is calculated using (5):

$$\% \text{ of Accuracy} = \frac{\text{No of TRUE Accuracy Result}}{\text{Total No of Testing Images}} \quad (5)$$

### 3. RESULTS AND ANALYSIS

Forty testing images are tested for each type of paddy leaf disease. Table 7 shows some of the results plotted by the truth table from each type of disease.

Table 7. The Proposed Truth Table for Performance Evaluation

No.	Image	Disease Classification	Actual Disease	Accuracy
1		Brown Spot	Brown Spot	TRUE
2		Tungro Virus	Tungro Virus	TRUE
3		Bacterial Leaf Blight	Bacterial Leaf Blight	TRUE
4		Leaf Scald	Leaf Scald	TRUE

The performance of the paddy leaf disease classification is demonstrated in Table 8. From the calculation of accuracy, it is observed that the study produced a good performance with 90% of accuracy for the brown spot, 95% of accuracy for grade tungo virus, and 80% of accuracy for both bacterial leaf blight and leaf scald. The tungo virus returned the highest percentage of accuracy, whereas, the bacterial leaf blight and leaf scald appear the moderate percentage of accuracy. The overall mean percentage of accuracy is observed to produce a good percentage of accuracy which is 86.25%.

Table 8. Accuracy Result

Paddy Leaf Disease	No. of TRUE Classification	% of Accuracy
Brown Spot	36	90
Tungro Virus	38	95
Bacterial Leaf Blight	32	80
Leaf Scald	32	80
MEAN		86.25

Despite a good performance of paddy leaf disease classification presented, several improvements are suggested. The implementation and incorporation of other feature extraction and classification techniques such as Gabor filter and Convolutional Neural Network (CNN) are recommended to improve the classification results in the future.

#### 4. CONCLUSION

This paper proposed a study of automatic classification of paddy leaf disease using image processing. Feature extraction techniques of color, texture and shape were implemented to analyze the characteristics of the paddy leaf disease. In another note, a Support Vector Machine (SVM) is used to classify the four types of paddy leaf disease which are the brown spot, bacterial leaf blight, tungro virus, and leaf scald. The application to a variety of testing images has been successful. The performance of the paddy leaf disease is evaluated using a truth table. The performance obtained exhibit a little variation in classifying the type of paddy leaf diseases. The overall mean percentage of accuracy demonstrated a good percentage of accuracy which is 86.25%. Therefore, it can be concluded that the proposed implementation of image processing techniques for the automatic classification of paddy leaf disease is found to be successful. Yet, an implementation and incorporation of the current feature extraction and classification techniques are recommended.

#### ACKNOWLEDGEMENTS

The research was supported by Ministry of Education Malaysia (MoE), and Universiti Teknologi MARA through the Fundamental Research Grant Scheme (FRGS) (600-IRMI/FRGS 5/3 (215/2019)).

#### REFERENCES

- [1] P. Kumar, *et al.*, "Detection of Healthy and Defected Diseased Leaf of Rice Crop using K-Means Clustering Technique," *International Journal of Computer Applications*, vol. 157(1), pp. 24-27, 2017.
- [2] B. Mukherjee and N.U. Maheswari, "Biological Control of Narrow Brown Leaf Spot and Leaf Smut Disease in Paddy Crops by Some Antagonistic Fungi," *Global Journal of Medical Research*, 2018.
- [3] J. P. Shah, *et al.*, "A Survey on Detection and Classification of Rice Plant Diseases," *IEEE International Conference on Current Trends in Advanced Computing (ICCTAC)*, 2016.
- [4] A. A. Joshi, and B. Jadhav, "Monitoring and Controlling Rice Diseases Using Image Processing Techniques," *International Conference on Computing, Analytics and Security Trends (CAST)*, pp. 471-476, 2016.
- [5] P. B. Padol and A. A. Yadav, "SVM classifier based grape leaf disease detection," *2016 Conference on Advances in Signal Processing (CASP)*, Pune, pp. 175-179, 2016.
- [6] P. Chaudhary *et al.*, "Color Transform Based Approach for Disease Spot Detection on Plant Leaf," *International Journal of Computer Science and Telecommunications*, vol. 3(6), pp. 65-71, 2012.
- [7] A. Singh and M. L. Singh, "Automated blast disease detection from paddy plant leaf - A color slicing approach," *2018 7th International Conference on Industrial Technology and Management (ICITM)*, Oxford, 2018, pp. 339-344.
- [8] P. Kumar and N. Bhoi, "Detection of Healthy and Defected Diseased Leaf of Rice Crop using K-Means Clustering Technique," *International Journal of Computer Applications*, vol. 157(1), pp. 24-27, 2017.
- [9] N. Rishi and J. S. Gill, "An Overview on Detection and Classification of Plant Diseases in Image Processing," *International Journal of Scientific Engineering and Research (IJSER)*, vol. 3(5), pp. 114-117, 2015.
- [10] Jadhav R. *et al.*, "Risk Factors for Agricultural Injury: A Systematic Review and Meta-analysis," *Journal of Agromedicine*, vol. 20(4), pp. 434-449, 2015.
- [11] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4(1), pp. 41-49, 2017.
- [12] P. Chaudhary *et al.*, "Color Transform Based Approach for Disease Spot Detection on Plant Leaf," *International Journal of Computer Science and Telecommunications*, vol. 3(6), pp. 65-71, 2012.
- [13] B. Chitradevi and P. Srimathi, "An Overview on Image Processing Techniques," *International Journal of Innovative Research in Computer and Communication Engineering*, pp. 6466-6472, 2014.
- [14] A. S. A. Salam, M. N. M. Isa, M. I. Ahmad, "Spatial domain image enhancement techniques for acute myeloid leukemia (M1,M4,M5,M7)," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14(1), pp. 250-257, April 2019.
- [15] A. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989, Chap. 7, pp. 235.
- [16] H. A. Nugroho *et al.*, "Shape analysis for classification of breast nodules on digital ultrasound images," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13(2), pp. 837-844, Feb 2019.
- [17] A. Singh and K. Gupta, "A Contrast Enhancement Technique for Low Light Images," *2nd International Conference on Communication Systems (ICCS 2015)*, 2015.

- [18] M. Bouillon *et al.*, "Grayification: A meaningful grayscale conversion to improve handwritten historical documents analysis," *Pattern Recognition Letters*, vol. 121, pp. 46-51, 2019.
- [19] P. Mohanaiah *et al.*, "Image Texture Feature Extraction Using GLCM Approach," *International Journal of Scientific & Research Publication*, vol. 3(5), pp. 1-5, 2013.
- [20] D. P. Tian, "A Review on Image Feature Extraction and Representation Techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8(4), Jul 2013.
- [21] P. S. Kumar and Vs. Dharun, "Extraction of Texture Features using GLCM and Shape Features using Connected Regions," *International Journal of Engineering and Technology*, vol. 8. pp. 2926-2930, 2016.
- [22] J. Na`am *et al.*, "An Automatic ROI of The Fundus Photography," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8(6), pp. 4545-4553, Dec 2018.
- [23] N. L. I. Rusli, A. Amir, N. A. H. Zahri, R. B. Ahmad, "Snake species identification by using natural language processing," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 13(3), pp. 999-1006, Mar 2019.
- [24] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in image classification," *International Journal of Information Technology*, 2018.
- [25] H. T. Sencar and N. Memon, "Digital image forensics: There is more to a picture than meets the eye," *Springer*, (H. T. Sencar & N. Memon, Eds.), New York, NY: Springer New York, 2013.

## BIOGRAPHIES OF AUTHORS



Shafaf Ibrahim is a senior lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka, Kampus Jasin. She holds a Diploma, Bachelor's Degree, Masters and PhD in Computer Science. Her research interests are Artificial Intelligence, Evolutionary Algorithms, Machine Learning, and Image Processing.



Nurnazihah Wahab is a final year student of Bachelor's Degree of Computer Science in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Malaka, Kampus Jasin. Her research interest covers Image Processing and Artificial Intelligence.



Ahmad Firdaus Ahmad Fadzil is a lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka, Kampus Jasin. He holds a Diploma, Bachelor's Degree, and Masters in Computer Science. His research interests are Artificial Intelligence, Evolutionary Computing, Parallel Processing and Image Processing.



Nur Nabilah Abu Mangshor is a lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka, Kampus Jasin. She holds a Diploma, Bachelor's Degree, and Masters in Computer Science. Her research interests are Artificial Intelligence and Image Processing.



Zaaba Ahmad is a lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Perak, Kampus Tapah. He holds a Diploma, Bachelor's Degree, and Masters in Computer Science. His research interests are Artificial Intelligence, Social Media Analytics, E-commerce and Social Commerce.