# Content based Paddy Leaf Disease Recognition and Remedy Prediction using Support Vector Machine

Farhana Tazmim Pinki
Computer Science and Engineering Discipline
Khulna University
Khulna, Bangladesh
farhana.kucse@gmail.com

Nipa Khatun
Computer Science and Engineering Discipline
Khulna University
Khulna, Bangladesh
nipakhatun130208@gmail.com

S.M. Mohidul Islam
Computer Science and Engineering Discipline
Khulna University
Khulna, Bangladesh
mohid@cse.ku.ac.bd

***Abstract*—Rice is one of the staple foods of the world. But the production of rice is hampered by various kind of paddy diseases. One of the main diseases of paddy is leaf disease. Generally, it is very time-consuming and laborious for farmers of remote areas to identify paddy leaf diseases due to unavailability of experts. Though experts are available in some areas, disease detection is performed by naked eye which causes inappropriate recognition sometimes. An automated system can minimize these problems. In this paper, an automated system is proposed for diagnosis three common paddy leaf diseases (Brown spot, Leaf blast, and Bacterial blight) and pesticides and/or fertilizers are advised according to the severity of the diseases. K-means clustering is used for separating affected part from paddy leaf image. Visual contents (color, texture, and shape) are used as features for classification of these diseases. The type of paddy leaf diseases is recognized by Support Vector Machine (SVM) classifier. After recognition, the predictive remedy is suggested that can help agriculture related people and organizations to take appropriate actions against these diseases.**

***Index Terms*—Paddy Leaf Disease, K-Means Clustering, Color, Texture, Shape, SVM, Remedy**

## I. Introduction

Rice is an important agricultural crop. Most of the world's population consume rice as staple food. More than 90 percent of world's rice is produced and consumed in Asia [1]. In a country like Bangladesh, the population has increased about 1.2 percent annually [2], it is necessary to increase rice production to cope with the country's growing population. But farmers lose an estimated average of 37% of their rice crop to pests and diseases every year [3]. There are many reasons that make slow and less productive of rice production. One of the significant cause is paddy diseases. Generally, paddy leaves are more affected by different kind of diseases such as Brown spot, Leaf blast, and Bacterial blight.

Disease detection and recognition is a demanding task. Generally, diseases are detected manually which is very difficult and time-consuming. The naked eye observation of experts is the main approach adopted in practice which is expensive on large farms [4]. Most of the farmers in rural areas determine disease manually that sometimes causes an error to identify the types of disease. They do not aware of proper management to cure paddy leaf diseases. So, the production of rice is being hampered rapidly in every year. For this reason, a fast and accurate paddy leaf disease recognition system and proper care according to the severity of disease become essential tasks. This research mainly focuses on three most common leaf diseases named Brown spot, Leaf blast, and Bacterial blight. Brown spot is a fungal disease and lesions are initially small, circular and dark brown to purple-brown [4]. Leaf blast is also a fungal disease and lesions are a spindle or elliptical shaped and whitish to gray centers with red to the brownish border [4]. Bacterial leaf blight is caused by bacteria and it has yellow and straw-colored stripes and lesions are near the leaf tip or margin [3].

In a study conducted by Phadikar et al. [5], developed an automated system to classify brown spot and leaf disease of rice plant-based on morphological change. Otsu's segmentation algorithm was used to segment the image. Radial distribution of the hue from the center to the boundary of the spot images were extracted as features to classify disease. In a study conducted by Radhika et al. [4], developed a system for paddy leaf disease identification. They used K-means clustering technique to detect infected parts of a leaf and then extracted features from the segmented image. Gray level co-occurrence matrix and discrete wavelet transform were used for feature extraction. Finally, Back Propagation Neural Network (BPNN) algorithm was applied to classify paddy leaf diseases. Another study which was conducted by R. Islam et al. [6], developed an image processing technique to calculate the percentage of disease affected pixels of paddy leaf. They used K-means clustering method to segment the original image into three cluster images based on color. Among these images, affected and unaffected leaf regions were considered to calculate the percentage of affected pixels. They calculated the total number of pixel of the unaffected and affected region. Finally, the percentage of affected pixels was calculated. They observed disease severity and suggested appropriate measure for treatment for Leaf blast disease.

The main objectives of this research are to provide a robust paddy leaf disease recognition system using image processing techniques and give remedy prediction to take proper cures. The K-means clustering segmentation algorithm is used to segment the image and SVM classifier is applied to classify paddy leaf diseases. This system provides a proper guidance containing instantaneous remedies based on the severity of the disease.

The rest of this paper is organized as follows. Section II

978-1-5386-1150-0/17/$31.00 ©2017 IEEE

represents the proposed methodology in details. Section III presents the experimental results and discussion. Section IV ends the paper with the conclusion.

## II. Proposed Methodology

The proposed system has concentrated on recognizing the paddy leaf diseases which assists the farmers to take a proper measurement and increases the production of paddy. The system has two main phases: Training phase and Testing Phase. The proposed system architecture is shown in Fig. 1.

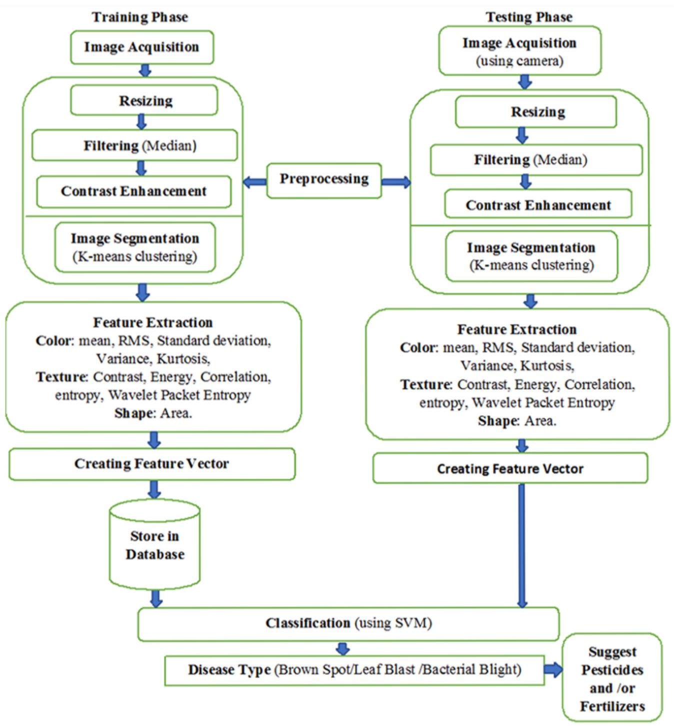



Fig. 1. System architecture of the proposed system

### A. Training Phase

Some disease affected paddy leaf images are used to train the SVM. The steps involved:

*1) Image Acquisition:* Some disease affected paddy leaf images are collected from the Internet and captured by camera from some rice producing area in Bangladesh. Samples of each disease are shown in Fig. 2.

*2) Image Preprocessing:* In this step, the dataset of images is preprocessed for further analysis. This step includes the following processes:

- **Image resizing**: Resizing is needed for image display and storage issues. The images are resized to $900 \times 700$ pixels. Then the images are used for further processing.
- **Image filtering**: When the images are captured, they may contain dust, sediment, slob, dew drops etc. The median filter is used to remove the noise and smooth the images. The median filtering is a technique to smooth the images by replacing the value of a pixel by the median of the intensity value of the neighborhood of that pixel so that the noises are replaced by the image intensity value. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value.
- **Contrast Enhancement**: The contrast of the image is enhanced for better understanding of different part of the image such as normal object and lesion present in the leaf part.
- **Image Segmentation**: Segmentation means representing the image into the different meaningful part that makes easy to analyze the images [7]. In this system, K means clustering (a color-based segmentation algorithm) is used to segment the images so that we can obtain the disease affected part of the paddy leaf images. The images are segmented into 3 clusters depending on the color variation. Among the three clusters, the disease affected part is chosen to extract features. Following are the steps to segment the image [8]:

(a) (b) (c)

Fig. 2. (a) Brown spot (b) Leaf blast (c) Bacterial blight affected paddy leaf images

Step 1: Take the resultant image.
Step 2: Convert image from RGB color space to L*a*b* color space.
Step 3: Classify the colors in a*b* space using K-means clustering.
Step 4: Label every pixel in the image using the result from K-means.
Step 5: Create images that segment the original image by color.
Step 6: Separate the infected and uninfected part.

*3) Feature Extraction:* In image processing, feature plays an important role in dimensionality reduction. When the size of the input data becomes too large but not containing so much information then the input data can be transformed into a set of features [7]. In this system, the visual-based features i.e. color, texture, and shape features are extracted.

*Color features* are one of the representative features of image. There are many approaches to extract color features. Color moment is one of the simplest and effective color features and it's math foundation that any color distribution can indicate with its moment [7]. The three color moments i.e mean, standard deviation and skewness are enough to indicate color distribution [7]. Mean, RMS, Variance, Standard deviation, and Kurtosis values are used as color features in this work. These features are computed using the following [9]:

**Mean**: The mean takes the average level of intensity of the image. In other words, this can be described as the sum of the pixel values divided by the total number of pixels.

$$m = \sum_{i=1}^{L-1} r_i p(r_i) \tag{1}$$

Where $m$ is mean, $r_i$ is the intensity value of the image and $p(r_i)$ is the probability of intensity value.

**RMS**: It is the square root of the mean of the squares of the value.

$$RMS = \sqrt{\frac{1}{MN} \sum_{i=1}^{K} \sum_{j=1}^{K} (I_{ij} - \bar{I})} \tag{2}$$

where intensity $I_{ij}$ is $i^{th}$ and $j^{th}$ element of the two-dimensional image of size $M$ by $N$. $I$ is the average intensity of all pixel values in the image.

**Variance**: The variance is denoted by $\sigma^2$ or $\mu_2$ whereas the mean is a measure of average intensity, the variance is a measure of contrast of an image.

$$\mu_2(r) = \sum_{i=1}^{L-1} (r_i - m)^2 p(r_i) \tag{3}$$

**Standard deviation**: Standard deviation is the square root of variance.

**Kurtosis**: The fourth moment is a measure of relative flatness of the histogram.

$$\mu_4(r) = \sum_{i=1}^{L-1} (r_i - m)^3 p(r_i) - 3 \tag{4}$$

*Texture features* which represent the orientation of the surface are very important for image analysis. Contrast, Energy, Entropy, Correlation from Grey level co-occurrence matrix and Wavelet packet entropy values are used to extract texture feature. These features are computed using the following [9]:

**Contrast**: A measure of intensity contrast between a pixel and its neighbor over the entire image. The range of values is 0 to $(K-1)^2$.

$$Con = \sum_{i=1}^{K} \sum_{j=1}^{K} (i-j)^2 p_{ij} \tag{5}$$

Where $p_{ij}$ is an estimate of the probability that a pair of points satisfying the sum of elements of Grey level co-occurrence matrix and $k$ is the row (or column) dimension of the square matrix.

**Energy**: It is also called uniformity. A measure of uniformity in the range $[0, 1]$. Uniformity is 1 for a constant image.

$$E = \sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij}^2 \tag{6}$$

**Entropy**: It is a statistical measure of randomness to characterize the texture of an image. The entropy is 0 when all the pixel values are 0 and is maximum when all the pixel values are equal. The maximum value is $2log_2 k$.

$$H = -\sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij} log_2 p_{ij} \tag{7}$$

**Correlation**: A measure of how correlated a pixel is to its neighbor over the entire image. The range of values is 1 to $-1$, corresponding to perfect positive and negative correlations.

$$Cor = \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{(i-m_r)(j-m_c)p_{ij}}{\sigma_r \sigma_c} \tag{8}$$

**Wavelet Packet Entropy**: The segmented image is decomposed into four subbands to obtain approximation, horizontal, vertical and diagonal coefficient using discrete wavelet transform. Then texture features are extracted by performing log energy entropy operation over the coefficient of wavelet transform using (7).

*Shape features* represent the contours and physical structure of the image. In the proposed method, the area of disease affected part is used as shape feature.

*4) Creating Feature Vector and Storing in Database:* After extracting features, a feature vector is created and stored it in the database.

### *B. Testing Phase*

Test images are captured by camera from the paddy field. The query image is processed and features of this image are extracted using the same processes of the training phase. Then a feature vector is created for query image. This feature vector is sent to a classifier to recognize the paddy leaf diseases.

*1) Disease Classification using SVM:* The paddy leaf disease is detected and recognized using SVM. SVM performs classification tasks by constructing hyperplane that separates cases of different class labels. Here the training points that are nearest to the separating function are called support vectors. Optimal separating hyperplane maximizes the margin of training data. The separating hyperplane equation [5]:

$$y = wx + b \tag{9}$$

Where $w$ is weight vector and $b$ is offset.

It classifies the query image to one of the types of paddy leaf disease.

*2) Suggest pesticide and/or fertilizer:* To seek out the severity of disease, the area of the whole leaf image and the disease affected area of the image are calculated. Finally, the percentage of disease affected area is measured using the following equation:

$$A_p = \frac{A_d}{A_l} \times 100\% \tag{10}$$

Where $A_p$ is percentage of affected area, $A_d$ is area of disease affected part and $A_l$ is total Area of leaf.

The Predictive measurement that is followed to suggest pesticides or fertilizers is shown in Table I.

TABLE I
PREDICTIVE MEASUREMENT FOR PADDY LEAF DISEASES

| Disease Type | Percentage of affected region | Pesticides or Fertilizer suggestion |
|---|---|---|
| Brown spot [3] [10] | Initial stage (Less than 25%) | (i) monitor soil nutrients regularly (ii) apply required fertilizers (iii) for soils that are low in silicon, apply calcium silicate slag before planting |
| | Severe stage (More than 25)% | Spraying of infected plants with fungicides, such as Benzoyl and Iprodione, and antibiotics, such as Validamycin and Polyoxin, is effective against the disease. |
| Leaf blast [6] | Less than 10% | Close observation and flooding the whole field with huge water |
| | 10% to 15% | (i) Adjust planting time. Sow seeds early, when possible, after the onset of the rainy season. (ii) Split nitrogen fertilizer application in two or more treatments. Excessive use of fertilizer can increase blast intensity. |
| | 15% to 25% | Silicon fertilizers (e.g., calcium silicate) can be applied to soils that are silicon deficient to reduce blast. |
| | More than 25% | Systemic fungicides like triazoles and strobilurins can be used judiciously for control to control blast. A fungicide application at heading can be effective in controlling the disease. |
| Bacterial leaf blight [3] [11] | Initial stage (Less than 25)% | (i) Use balanced amounts of plant nutrients, especially nitrogen. (ii) Apply N in three split doses, 50% basal, 25% in tillering phase and 25%N in panicle initiation stage |
| | Severe stage (More than 25%) | Foliar spray of 0.05 g Streptocycline and 0.05 g Copper Sulfate |

## III. EXPERIMENTAL RESULTS

The experiment is performed in two phases: Training phase and Testing phase. In training phase, the training image is processed using different preprocessing steps. Fig. 3 shows the result after resizing the image to $900 \times 700$, filtering using median filter, and enhancing the contrast.

Fig. 4 shows the result of image segmentation using K-means clustering. From the three segmented images, the disease affected segment is chosen to extract the feature. Since cluster 1 is the disease affected segment, it is chosen to extract features. Fig. 5 shows the result of feature extraction.

After extracting the features, the feature vector is created and stored in the database.

The preprocessing, segmentation and feature extraction of a query image are performed in the same way as training phase.

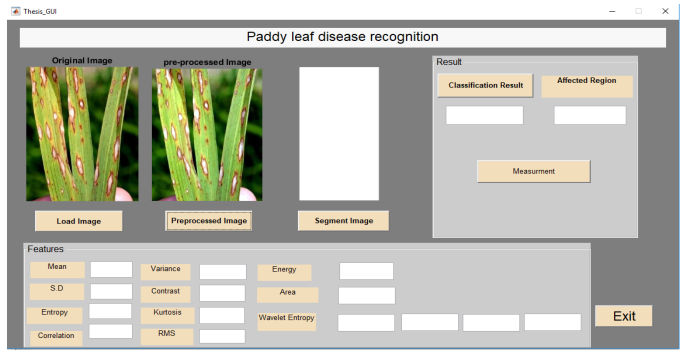



Fig. 3. Result after preprocessing

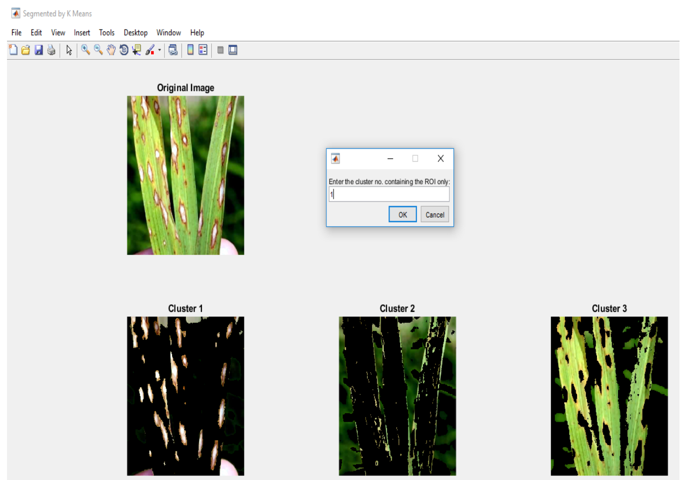



Fig. 4. Result of segmentation

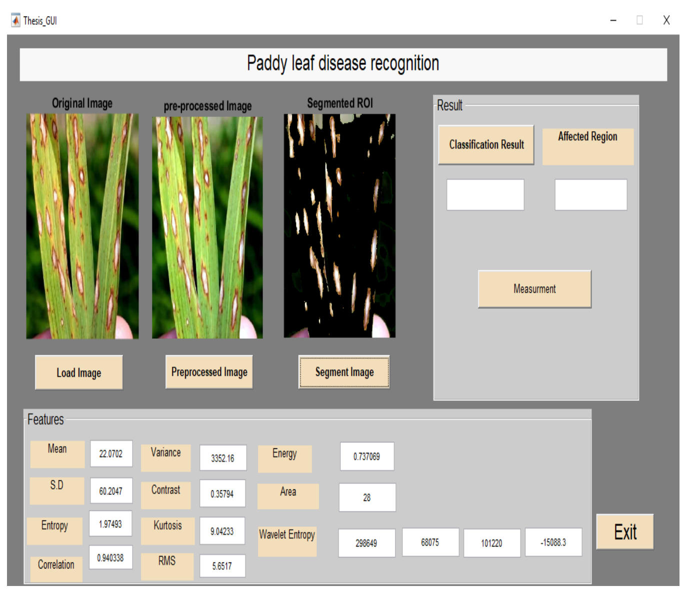
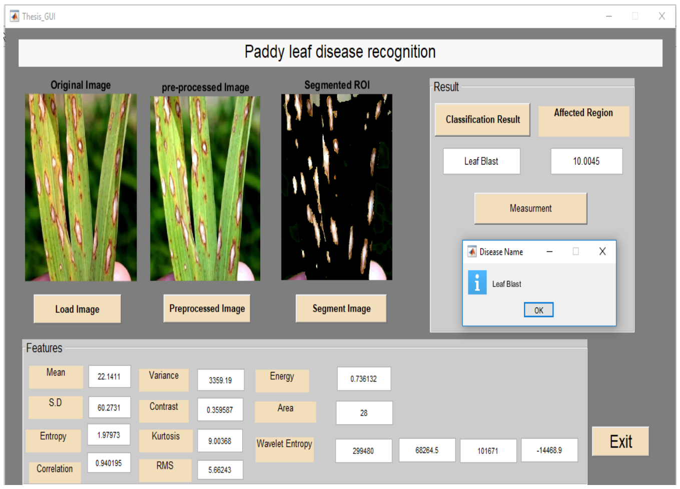


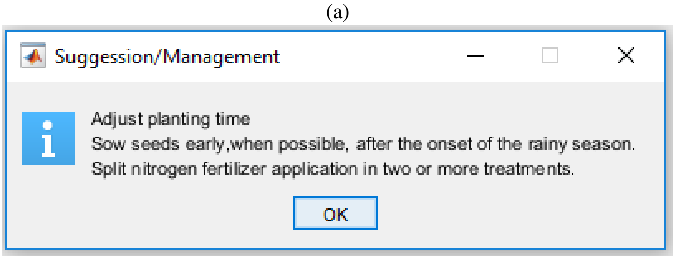

Fig. 5. Result after feature extraction

After creating the feature vector of a query image, the feature vector is sent to the classifier and SVM classifies paddy leaf diseases using training dataset.

Then the percentage of affected area is calculated using (10). Since the affected area is $10.0045\%$ for the query sample (which is recognized as Leaf blast), the predictive measure is suggested according to Table I. This result and remedy are shown in Fig. 6 for the query sample.




(a)




(b)

Fig. 6. (a) Recognized result and percentage of affected area (b) Predicted remedy

The accuracy of the outcome of each individual disease and the overall accuracy of the system is shown in Table II. Table III shows the comparison of this system with some existing works.

TABLE II
ACCURACY OF THE SYSTEM

| Disease Type | Recognition accuracy | Overall Accuracy |
|---|---|---|
| Brown spot | 90.9% | 92.06% |
| Leaf blast | 94.11% | |
| Bacterial leaf blight | 85.71% | |

TABLE III
PERFORMANCE COMPARISON

| Reference | Overall Accuracy |
|---|---|
| Joshi et al. [12] | 87.02% (using Minimum Distance Classifier) |
| | 89.23% (using k-Nearest Neighbor Classifier) |
| T. Suman et al. [13] | 70% |
| This system | 92.06% |

## IV. CONCLUSION

This research aims to develop a robust and user-friendly system for disease classification of paddy leaves. For this purpose, K-means clustering technique is used to choose the disease affected segment. To analyze the paddy leaf disease color, texture, and shape features are extracted from the disease affected segment. Supervised learning-based classifier SVM is used to classify diseases. The percentage of affected region is computed and pesticides and/or fertilizers are suggested based on the severity of diseases. We see that the system shows robust result than some existing methods.

## REFERENCES

[1] *Rice production in the Asia-pacific region: issues and perspectives - M.K. papademetriou**, 2017, last accessed Date: 21-August-2017. [Online]. Available: http://www.fao.org/docrep/003/x6905e/x6905e04.htm

[2] L.Cago, *Bangladesh population (2017) - Worldometers*, 2017, last accessed Date: 17-July-2017. [Online]. Available: http://www.worldometers.info/world-population/bangladesh-population

[3] *Pests and diseases - IRRI Rice Knowledge Bank*, 2017, last accessed Date: 20-July-2017. [Online]. Available: http://www.knowledgebank.irri.org/step-by-step-production/growth/pests-and-disease

[4] R. Deshmukh, "Detection of paddy leaf diseases," *International Conference on Advances in Science and Technology*, vol. 2, pp. 8–10, 2015.

[5] S. Phadikar, J. Sil, and A. K. Das, "Classification of rice leaf diseases based on morphological changes," *International Journal of Information and Electronics Engineering*, vol. 2, no. 3, p. 460, 2012.

[6] R. Islam and M. R. Islam, "An image processing technique to calculate percentage of disease affected pixels of paddy leaf," *International Journal of Computer Applications*, vol. 123, no. 12, pp. 28–34, 2015.

[7] M. P. B. Mr. Girish Athanikar, "Potato leaf diseases detection and classification system," *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 2, pp. 76–88, 2016.

[8] S. R. N.Neelaveni, "Paddy leaf disease detection classification and remedy finder using k means clustering and image mining," *International Journal of Computer Applications*, vol. 4, no. 12, pp. 20 859–20 866, 2016.

[9] R. E. W. Rafael C Gonzalez, *"Digital image processing"*, 3rd ed. Pearson publication.

[10] R. Zainon, "Paddy disease detection system using image processing," Ph.D. dissertation, UMP, 2012.

[11] *Management of Bacterial leaf blight of Rice — agropedia*, 2017, last accessed Date: 08-August-2017. [Online]. Available: http://agropedia.iitk.ac.in/content/management-bacterial-leaf-blight-rice

[12] A. A. Joshi and B. Jadhav, "Monitoring and controlling rice diseases using image processing techniques," in *Computing, Analytics and Security Trends (CAST), International Conference on*. IEEE, 2016, pp. 471–476.

[13] T. Suman and T. Dhruvakumar, "Classification of paddy leaf diseases using shape and color features," *IJEEE*, vol. 7, no. 01, pp. 239–250, 2015.