Detection and Classification of Rice Plant Diseases using Image Processing Techniques

D.Bandara
Department of Computer Science
University of Jaffna.
Jaffna, Sri Lanka
dilshika.bandara@gmail.com

B.Mayurathan
Department of Computer Science
University of Jaffna.
Jaffna, Sri Lanka.
barathym@univ.jfn.ac.lk

***Abstract*—Plant leaf disease detection and classification is one of the interesting research areas in the agriculture sector. An approach based on image processing and machine learning techniques is proposed in this paper to detect and classify paddy leaf diseases. Leaf Blast, Brown Spot and Bacterial Leaf Blight diseases are considered to calculate the performance of this proposed methodology. Colour thresholding is applied to identify the disease area in the paddy leaf. Hence, various feature categories such as colour, texture, and shape features are extracted from the affected area of the diseased image. Support Vector Machine (SVM) and *k*-nearest neighbors (*k*-NN) algorithms are used as classifiers and the performances of the proposed methodology are evaluated using these classifiers. The experimental results are compared with state-of-the-art work approaches. Our proposed approach achieves 89.19%, 82.86%, and 89.19% of accuracy for detecting the rice plant diseases such as leaf blight, brown spot and leaf blast respectively.**

Keywords*— Rice leaf Disease, colour thresholding, disease classification, SVM-classifier, k-NN classifier.

I. Introduction

The main source of the income in Sri Lanka is agriculture and 75% of the Sri Lankan population depends on farming. Also, Rice is the one of the main traditional food in Sri Lanka. Rice is rich in nutrients, vitamins, and minerals. In Sri Lanka there are more than 2000 indigenous rice varieties [1]. There is a significant demand to produce traditional rice varieties in the local and international markets. In Asian Region including Sri Lanka, health threats is a common problem among both rice growers and consumers.

The main reason for the quantity and quality reduction of crops due to crop diseases. Both quality and quantity reduction directly affected the overall production of the crop in the country. Therefore, the identification of plant diseases is most important as early as to increase profit and crop yield.

The types of plant diseases can be categorized as a disease, injury, and disorder [2]. Most of the plant diseases are caused by virus, bacteria, and fungi. Most of the common rice plant diseases are Brown spot, Bacterial leaf blight, Leaf blast , Leaf smut and Sheath blight [3]. Each of the disease has its own characteristics. They differ from colour, shape, and size. But sometimes in two diseases colour can be the same but shape may be different or shape may be the same colour can be different. In such cases, farmers may get confused in identifying the disease.

In this context, an accurate diagnosis is very important in the early stage of rice plant. Naked eye observation [4] is one of the traditional approaches to identify the diseases in rice plant. Here [4], farmers use their own experience to identify the diseases. Hence, this approach is time consuming. Nowadays, image processing techniques are widely used in state-of-the- art-methods [5, 6, 7] to identify the diseased plant leaves [8].

This paper presents an approach based on image processing and machine learning techniques to detect and identify the rice leaf diseases.

The rest of the paper is organized into the following sections: Section 2 presents three different types of rice plant leaf diseases those are considered in this paper to evaluate the performance of the proposed methodology. A brief description of the state-of-the-art researches that have already been done on rice plant leaf disease detection are mentioned in Section 3. Section 4 indicates the detailed methodology we carried out for detecting the rice plant leaf diseases. Section 5 gives the experimental setup and testing outputs of this paper. Conclusion of this paper is given in section 6.

II. Background Theory

The rice plant diseases that we choose in this paper and their characteristics are presented in this section. Figure 1 shows sample image for Brown Spot, Bacterial Leaf Blight and Leaf Blast shows Diseases respectively. Also, the symptoms of the three diseases are shown in Table 1.

Brown Spot *Bacterial Blight* *Leaf Blast*

Fig. 1. Chosen Disease Images

A. Brown Spot

Brown Spot is the most common and dangerous rice disease. It is a fungal plague that infects the coleoptile, leaves, leaf sheath, panicle branches, glumes, and sprinkles. In the starting stage spots are initially small, circular, and dark brown to purple-brown. When the growth stage is reached then the spots are ovular shape with brown to grey center that surrounded by a sepia colour margin.

B. Bacterial Blight

This disease mostly likes to develop in areas that have weeds and stubble of infected plants. Bacterial blight causes the wilting of seedlings and yellowing and drying of leaves. The disease occurs in both tropical and temperature environments, particularly in irrigated and rain fed lowland areas. Leaves colour turn in to yellow to straw-coloured and wilt. Bacterial blight is the most destructive disease of rice. It cankill seedlings or plant up to the tillering stage.

C. Leaf Blast

The blast is affected by fungus. It develops areas which have low soil moisture, frequent and prolonged periods of a rain shower, and cold temperature in the daytime. Symptoms of the rice leaf blast appear as white to gray-green lesions or spots with dark green borders. Older lesions of leaf blast are stem or elliptical shaped with grey to whitish centers that has centerscolour varies between red to brownish.

Table I Symptoms for plant leaf diseases

Disease	Brown Spot	Leaf Blight	Leaf Blast
Symptoms	Small circular spots	Yellowing and dying leaves	Spots with dark green borders
Shape	Circular and oval	Several inches long	Diamond shape
Colour	Change dark brown to purple brown grey, reddish brown margin in the center	Yellow and strawed colour	Grey- green lesions or spots

III. Literature Review

This section presents an overview of research carried out using machine learning and image processing techniques to identify the plant leaf diseases.

Prototype based approach [8] is proposed to recognize plant leaf diseases such as Bacterial Blight, Brown spot and Leaf Smut. The images are captured using NIKON D90 camera with dimension 2848 x 4288. Colour, texture and shape features are considered and totally 88 features are used in this work [8]. First converted rice leaf images into HSV colour space. Then Saturation component of the converted image is extracted. HSV (Hue Saturation Value) colour mask is used to remove background from the images. The mask is then applied to RGB (Red Green Blue) colour space image and again converted into HSV colour space after removing the background. K-means Clustering is used for image segmentation. Three clusters namely diseased portion, background and green portion are set in this experimental setup. Hence, three different image segmentation techniques which are Otsu's segmentation technique, LAB colour space based K-means clustering and HSV colour space based K-Means Clustering are used to image segmentation. During the recognition phase SVM is used for the multi-class classification. Based on the testing results, 73.33% of accuracy is achieved for training images. Also, 5- fold and 10-fold cross-validation are performed and achieve 83.80% and 88.57% accuracy respectively.

Bag of Words (BoW) approach is used in [5] to recognize the rice crop diseases. 400 images are collected from various sources. Images are converted into gray-scale images to make uniform colour of all images. Scale Invariant Feature Transform (SIFT) is used to extract distict points and calculate their descriptors. Clustering is done by K-means clustering algorithm for the similar words making as a group. This approach is applied to classify three rice crop diseases namely Brown Spot, False Smut, Bacterial Blight. SVM is used to classify the diseases and SVM parameters (gamma, nu) are optimized for maximize the efficiency and achieved 94.16% accuracy.

Spot colour and outline of the spot area from diseased leaf are considered to developed an algorithm [9] for detecting rice plant diseases. In this work, images are collected from the China National Rice Research Institute and the resolution of all images were reduced to 800 x 600 pixels. Shape and texture features are extracted after removing the noises from the image. SVM classifier is applied to classify rice blast, bacterial leaf blight and leaf sheath blight disease spots and obtained 97.2% accuracy.

Image recognition system is presented in [10] to recognize paddy leaf diseases. The authors chose three different rice leaf diseases such as rice blast, rice sheath blight, and brown spot to evaluate the performance of their proposed system. Also, mathematics morphology method is used for image segmentation. Then texture, shape and colour features are extracted from disease spot. Classification method with membership function is used to classify the diseases, and based on the testing results, 70% classification accuracy is achieved for 50 sample images.

Software prototype system is developed in [11] for rice plant disease detection. Infected images are collected from various rice plants and then images are enhanced by increasing contrast and brightness. Also, Hue Intensity Saturation (HIS) model is generated for image segmentation. 8-connectivity boundary detection algorithm is used to detect boundary and spots. Self-organizing map (SOM) neural network is used for classification. Three different cases such as RGB of the spots, fourier transform of the spot and arbitrary rotation of the spot are analyzed in this experiment obtained 70% of successful classification rate.

An image processing-based system is developed in [12] to classify the different paddy leaf diseases namely brown spot, leaf blast and bacterial blight. Disease portions are located using Haar-like feature and AdaBoost (Adaptive Boosting) classifier and achieved 83.33% accuracy rate on detection. In recognition phase, SIFT is used to feature extraction and k-NN and SVM approaches are used for classification and achieved accuracy 93.33% and 91.10% respectively.

Fermi energy-based segmentation approach [13] is proposed to differentiate different types of rice diseases. Fermi energy-based region detection concept is based on pauli exclusion principle states that fermi energy is attained when temperature of a material is lowered to absolute zero. Infected region of the image is isolated from its background and colour, shape and position of the infected portion are extracted using developed segmentation algorithm. A rule-based classifier is built to cover all the diseases of the rice plant images

Automatic detection of plant diseases is proposed using an approach based on machine learning techniques [14]. K-means clustering is used for the image segmentation and then mean value, standard deviation and GLCM texture features of disease images are extracted. Healthy images and disease images are differentiated using ANN classification method. The simulation results gives an accuracy of 99% for infected images and 100% for normal images during the training phase. Also, during the testing phase 90% of accuracy and 86% of accuracy are gained for infected and healthy images respectively.

IV. Proposed Methodology

An approach based on image processing and machine learning techniques for identification and classification of rice plant diseases is proposed in this section. Figure 2 depicts the block diagram of our proposed methodology. Following subsections describe each of the steps in our proposed methodology.

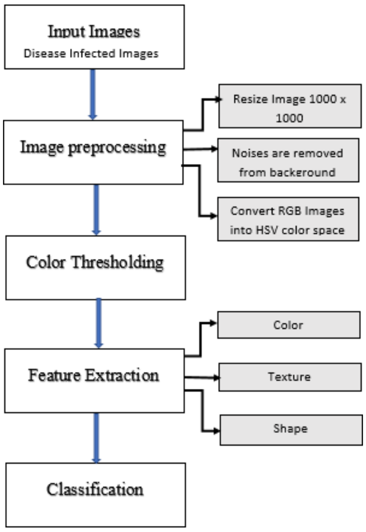

Fig. 2. Proposed Methodology for rice plant disease classification

A. Image Preprocessing

First, images are resized into a dimension of 1000 x 1000 pixels in order to convert all images into the same size. Computational power and memory requirements are reduced from this image resizing.

Then, original images are converted into HSV (Hue Saturation Value) colour space and H, S, V components are extracted from the image.

The disease infected images may contain either black, white, grey or green backgrounds. Binary mask for background removing is obtained the facility of MATLAB colour thresholder. Colour thresholder is used to remove the white and grey backgrounds into black colour. Then, these images are converted into HSV colour space. Common mask to remove the background from the infected rice leaf images is obtained by changing the S (Saturation) component value. Figure 3 shows the background removal using colour thresholding.

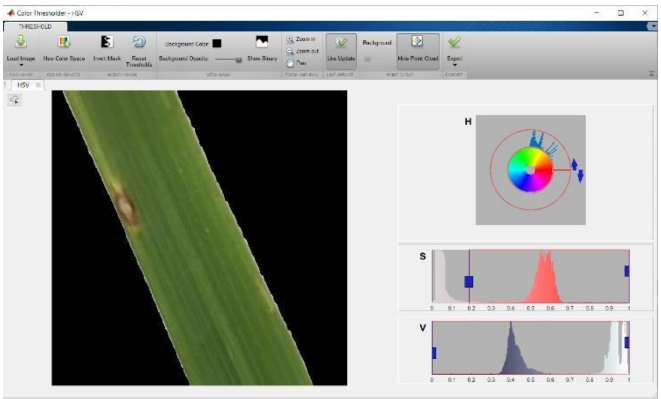

Fig. 3. Background removing using colour thresholding

B. Disease Segmentation

In this proposed methodology, Multiband thresholding is used to detect the diseased area. Then, background removed images are used as an input image and binary mask obtained using colour thresholding are used to separate healthy portion and the diseased portion of the image. Again, HSV colour space is used to remove the remaining noises in the diseased portion of the image. To obtain the disease portion, H value of the background removed image is used. The purpose of colour thresholding is to remove the green colour portion. By removing the green colour portion, the disease portion has been extracted.

C. Feature Extraction

Feature extraction means it represents the image data into a compact feature vector and it's a process of transforming raw data into numerical features. Those numerical values can be used to preserving the information about original image. Here, features are extracted under three categories such as shape, colour, and texture.

1) Shape Feature Extraction: Brown spot has circular or ellipse shaped disease spots, Leaf Blight has long area wilt and Leaf Blast has diamond shaped spots. So, the shape feature also helps to identify the differences between diseases. Disease area ratio is used as shape feature in this methodology. Disease area ratio can be obtained as follows.

$$\textit{Disease area position} = \frac{Ad}{At}$$

Where A_d is the total area of disease spots and A_t is the total area of the leaf image.

2) Colour Feature Extraction: Colour is one of the most widely used image features which can be used for pattern recognition and classification [15]. As mentioned in Table 1, the diseased images have different colours and patterns. Therefore, colour features are extracted from the diseased portion of the image. Mean value is the most basic statistical measure which is used in analysis and geometry [9]. The following colour features are used in this experimental setup

- Mean values of components of R(Red), G(Green), B(Blue)in RGB colour space,
- Mean values of H(Hue), S(Saturation), V(Value) components in HSV colour space,
- Mean values of L(Lightness), A (Red/Green value), B (Blue/Yellow value) components in LAB colour space.

After that extract the Standard Deviation of RGB colour space by extracting each R, G and B components and their respective Standard Deviation values, following values are extracted in order to have all features dimensions are same.

- Kurtosis: - This measure how outlier-prone a distribution is.
- Skewness: - This is a measure of the asymmetry of the data around the sample mean.

3) Texture Feature Extraction: Following texture features are extracted to calculate the performance of our proposed methodology

- Entropy - Entropy is a statistical measure of randomness and it is used to characterize the texture feature from the input image.

- Variance – It is a measure of how far a set of numbers is spread out. It is one of the several descriptors of a probability distribution, describing how far the numbers lie from the mean [16].
- Inverse Difference moment – It measures image homogeneity. This parameter achieves its largest value when most of occurrences in GLCM are concentrated near main diagonal. This is inversely propositional to the GLCM contrast.
- GLCM properties – The GLCM is a tabulation of how often different combinations of pixel grey level values occur in an image. Properties of GLCM such as Contrast, Correlation, Energy and Homogeneity are also considered in this work.

D. Classification

Support Vector Machine (SVM) and k-Nearest Neighbour (k-NN) are used as a classifier in this paper.

In Neural Networks there can be problems that occur due to random weight assignment. But, SVM does not suffer from that problem due to random weight assignment. It groups the training data base on the label classes that are given. In this work, three types of plant leaf diseases are considered. So, Gaussian kernel is used for multi-class classification. In this work, three types of plant leaf diseases are considered. So, Gaussian kernel is used for multi-class classification.

k-NN method is more efficient than SVM as it takes less computational time comparing to SVM.

V. EXPERIMENTAL SETUP AND TESTING RESULTS

Experimental design and its outcomes are presented in this section. Affected Rice plant leaf images were collected from UCI Machine Learning Repository, Kaggle, Shutterstock, and Rice Research Institute. Also, dataset that we developed contains a total of 754 images including 280 images of Brown Spot, 194 images of Bacterial Leaf Blight, and 280 images ofLeaf Blast.

In order to calculate the performance of our proposed methodology, dataset is split into training and testing. For the training phase, 175 Brown Spot images, 120 Bacterial Leaf Blight Images, and 175 Leaf Blast Images are randomly selected from the dataset. Remaining images are considered as testing phase.

First, one shape feature, 15 colour features and 7 texture features in total 23 features are extracted from the images.

A. Comparison of segmentation techniques

Performance of the colour thresholding and K-means techniques are compared in the segmentation phase to obtain the diseased portion of the image from original. Based on the testing results, K-Means Clustering does not always correctly identify the diseased portion and healthy portion. Also, colour thresholding gives more accurate outputs rather than the K- Means clustering.

Figure 4 and Figure 5 show the effect of segmentation by using colour thresholding and K-means respectively. Based on the output of the segmenting disease portion, colour thresholding gives better performance than K-means.

B. Comparison of the Training Classifiers

In order to determine the suitable classifier to train the model, k-NN (k-Nearest Neighbor) and SVM (Support Vector Machine) are compared in this experimental design. Table II

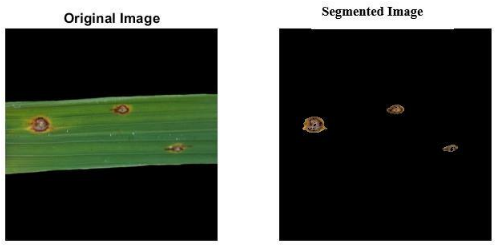

Fig. 4. Segmentation using Colour Thresholding

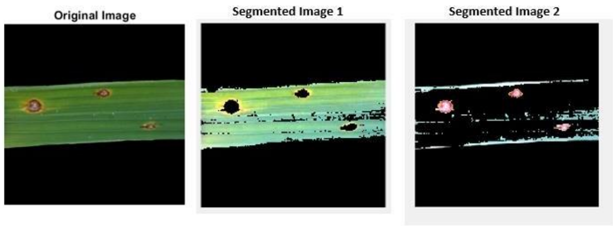

Fig. 5. Segmentation using K-means

shows the performances for training classification. Based on the results shown in Table II, k-NN classifier gives better classification performance than SVM classifier.

Table II classification performance of the trained model

Classifier	k-NN Classifier	SVM Classifier
Brown Spot	95.32%	93.62%
Leaf Blight	94.89%	95.75%
Leaf Blast	94.46%	92.78 %
Average Accuracy	94.89%	94.04%

C. Classification performances of the testing images

Multi-class SVM with Gaussian Kernel and k-NN method with k=7 are used in the testing phase to compare the performance of these classifier in this task. Table III and Table IV shows the performance of the rice plant disease detection using SVM and k-NN classifier accordingly.

According to the classification performance listed in Table III and IV, SVM gives better classification rate for Leaf Blight only. But k-NN shows better recognition performances in the other two classes.

Table III classification performance using SVM

Dataset	Brown Spot	Leaf Blight	Leaf Blast
Number of Training Images	175	120	175
Number of Testing Images	105	74	105
Number of correctly classified	83	66	71
Number of Miss-classified	22	8	36
Performance SVM (%)	79.05%	**89.19%**	67.62%

Also, performance of the proposed methodology is compared with state-of-the-art methods. Table VI shows the classification performances of the state-of-the-art methods to identify plant diseases. According to the results listed in Table VI, our proposed methodology gives better classification performances than others. Also, our proposed methodology uses large size of the dataset to evaluate the classification performances.

Based on the classification rates in Table VI, [12] shows high accuracy than our methodology. But, [12] used very small size of the dataset to compute the performances.

Table IV classification performance using k-NN

Dataset	Brown Spot	Leaf Blight	Leaf Blast
Number of Training Images	175	120	175
Number of Testing Images	105	74	105
Number of correctly classified	87	64	78
Number of Miss-classified	18	10	27
Performance k-NN	**82.86%**	83.78%	**74.28%**

Table VI Performance comparison with the State-of-the-art methods

Ref	Training images	Testing Dataset	classifier	Disease	Accuracy
[8]	40	-	SVM	Brown Spot	80%
	40	-	SVM	Leaf Blight	100%
	40	-	SVM	Leaf Smut	40%
[12]	40	10	KNN	Brown Spot	96.66%
	40	10	KNN	Leaf Blast	90%
	40	10	KNN	Leaf Blight	93.33%
[10]	50	11	NN	Brown Spot	85%
	50	15	NN	Sheath Blight	60%
	50	20	NN	Leaf Blast	80%
Ours	**280**	**105**	**k-NN**	**Brown Spot**	**82.86%**
	194	**74**	**k-NN**	**Leaf Blight**	**83.78%**
	280	**105**	**k-NN**	**Leaf Blast**	**74.28%**
Ours	**280**	**105**	**SVM**	**Brown Spot**	**79.05%**
	194	**74**	**SVM**	**Leaf Blight**	**89.19%**
	280	**105**	**SVM**	**Leaf Blast**	**67.62%**

VI. Discussion and Conclusion

Diseases in the rice plant leaf causes a huge loss in the agriculture domain. In this paper, an approach based on image processing and machine learning techniques is proposed to detect the rice leaf diseases. There are three types of rice leaf diseases such as Brown spot, Bacterial Leaf Blight, and Leaf Blast are considered in our experiments. Also, images are collected from different resources.

During the disease portion separation, performance of HSV colour thresholding and K-means are compared. Hence, this paper showed that HSV colour thresholding segments disease portion more accurately than K-means. Based on the classification performance of the trained model, k-NN gives better performance to recognizing Brown spot and Leaf Blast and SVM performs well in the other. So, performance of the k-NN and SVM are compared in the testing phase also.

According to the testing outputs, k-NN gives better classification performance than SVM. Therefore, classification performance of our approach is compared with the state-of-the-methods. Based on these experimental designs, we can confidently conclude that, our proposed approach gives high classification accuracies than others.

For future improvements, we would like to consider a large number of plant leaf disease classes and optimizing the parameter values to obtain higher accuracy than this work. Moreover, we intend to extract more features that would help to obtain higher accuracy on training and testing.

References

[1] Rajkumar, G., Weerasena, J., Fernando, K. and Liyanage, A,. "Assessment of genetic diversity among Sri Lankan rice varieties by AFLP markers", *Plant Genetic Resources*, *9*(2), pp.224-228, 2011.

[2] Gangadharan, K., Rosline, G.,Dhanasekaran, D. and Malathi, K., "Automatic detection of plant disease and insect attack using effta algorithm". volume 11,01 2020.

[3] Rice Production (Peace Corps): Chapter 14 - Diseases of rice. Last accessed on 23 November 2015. [Online]. Available: http://www.nzdl.org

[4] Barbedo JG, "Digital image processing techniques for detecting, quantifying and classifying plant diseases", *Springerplus*, Vol.2, No.1, (2013), pp.660–671

[5] Komal Bashir, Maram Rehman, and Mehwish Bari, "Detection and classification of rice diseases: An auto-mated approach using textural features",Mehran University Research Journal of Engineering and Technology, pp.239-250.

[6] Khirade, S.D. and Patil, A.B., "Plant disease detection using image processing". In *2015 International conference on computing communication control and automation* (pp. 768-771). IEEE.

[7] Kadir, A., Nugroho, L.E., Susanto, A. and Santosa, P.I., 2013, "Leaf classification using shape, colour, and texture features" arXiv preprint arXiv:1401.4447.

[8] Harshadkumar B Prajapati, Jitesh P Shah, and Vipul KDabhi, "Detection and classification of rice plant diseases".Intelligent Decision Technologies, pp.357-373.

[9] Q. Yao, Z. Guan, and Y. Zhou, "Application of support vector machine for detecting rice diseases using shape and colour texture features," in International Conference on Engineering Computation, pp. 79-83 2009.

[10] G Anthonys and N Wickramarachchi, "An image recog-nition system for crop disease identification of paddyfields in sri lanka", In2009 International Conference onIndustrial and Information Systems (ICIIS), pages 403–407. IEEE, 2009.

[11] Santanu Phadikar and Jaya Sil, "Rice disease identification using pattern recognition techniques", In *2008 11th International Conference on Computer and Information Technology*, pages 420–423,2008.

[12] K Jagan Mohan, M Balasubramanian, and S Palanivel, "Detection and recognition of diseases from paddy plantleaf images", International Journal of Computer Applica-tions, 144(12), 2016.

[13] Phadikar, S., Sil, J. and Das, A.K., 2013, "Rice diseases classification using feature selection and rule generation techniques", Computers and electronics in agriculture, 90, pp.76-85.

[14] S Ramesh, "Rice blast disease detection and classification using machine learning algorithm" ,In20182nd International Conference on Micro-Electronics andTelecommunication Engineering (ICMETE), pages 255–259. IEEE, 2018.

[15] Phadikar, S., Sil, J. and Das, A.K., 2013, "Rice diseases classification using feature selection and rule generation techniques", *Computers and electronics in agriculture*, *90*, pp.76-85.

[16] Vijay Kumar and Priyanka Gupta, "Importance of sta-tistical measures in digital image processing", I*nternational Journal of Emerging Technology and Advanced Engineering*, *2*(8), pp.56-62.