International Journal of Computer (IJC)
ISSN 2307-4523 (Print & Online)

http://ijcjournal.org/

Automatic Paddy Leaf Disease Detection Based on GLCM Using Multiclass Support Vector Machine

Venuja Satgunalingam[a]*, Rajeetha Thaneeshan[b]

[a,b]Vavuniya Campus of the University of Jaffna, P.O.Box 43000, Srilanka
[a]Email: venujakumar@gmail.com
[b]Email: rajeetha19@gmail.com

Abstract

The paddy leaf diseases have increased rapidly in the recent years because of globalization, environmental pollution and climate changes which reduce the production of rice and economy of the country. For healthy growth of rice plants there is a need of automatic system which can detect the paddy diseases automatically on time to give the proper treatment for the affected plants. In this paper, we proposed a methodology to develop an automatic system for detect the paddy disease which are Paddy Blast Disease, Brown Spot Disease, Narrow Brown Spot Disease using MATLAB. This paper concentrate on the image processing techniques used to enhance the quality of the image and Multiclass Support Vector Machine to classify the paddy diseases. The methodology involves image acquisition, pre-processing, segmentation, feature extraction and classification of the paddy diseases. Image segmentation technique is used to detect infected parts of leaf by using canny edge detection, multilevel thresholding and region growing techniques. We extract texture features using GLCM (grey level co- occurrence matrix) techniques, additionally we extract color and shape features to improve the accuracy of the framework and use Multiclass Support Vector Machine for classification. We achieved 87.5% accuracy for the test dataset.

Keywords: Paddy Blast; Brown Spot; Thresholding; Support Vector Machine.

1. Introduction

Srilanka is agriculture based country that has many people working in the agriculture industry. The agricultural sector plays an important role in economic development by providing rural employment. Paddy is one of the nation's most important products as it is considered to be one of Srilanka's staple food and cereal crops and because of that, many efforts have taken to ensure its safety, one of them is crop management of paddy plants. Paddy plants are affected by various fungal and bacterial diseases.

----

* Corresponding author.

An abnormal condition that injures the plant or leads it to function improperly is called as a disease. Diseases are readily recognized by their symptoms. There are a lot of paddy disease types which are Bakanae, red disease virus, brown spot disease and many more. Image processing and computer vision technology are very beneficial to the agricultural industry. They are more potential and more important to many areas in agricultural technology. This research work focuses on detecting three paddy plant diseases namely paddy Brown Spot Disease, Paddy Blast Disease, Narrow Brown Spot Disease using image processing techniques and Multiclass Support Vector Machine. Brown spot is a fungal disease that infects the coleoptile, leaves, leaf sheath, panicle branches, glumes, and spikelets. Its most observable damage is the numerous big spots on the leaves which can kill the whole leaf. When infection occurs in the seed, unfilled grains or spotted or discolored seeds are formed. Paddy blast, caused by a fungus, causes lesions to form on leaves, stems, peduncles, panicles, seeds, and even roots. So great is the potential threat for crop failure from this disease that it has been ranked among the most important plant diseases of them all. Other grasses, including crabgrass, are infected with closely related fungi (Magnaporthe grisea, Magnaporthe poae, Magnaporthe rhizophila and Magnaporthe salvinii), which cause nearly identical symptoms on their respective hosts. Narrow brown spot (also called narrow brown leaf spot, or rice Cercospora leaf spot) is caused by the fungus Sphaerulina oryzina (syn. Cercospora janseana, Cercospora oryzae) and can infect leaves, sheaths, and panicles. It leads to premature death of leaves and leaf sheaths, premature ripening of grains, and in severe cases, lodging of plants. So, the proper detection and recognition of disease is very important in applying required fertilizer.

2. Literature Review

Reference [1] This paper presents experimental results on an algorithm that was developed using Artificial Bee Colony (ABC). Results showed that ABC produced better percentage of correctness and detection time than Canny, Sobel, Roberts and Prewitt [2]. The proposed framework of this paper is image-processing-based and is composed of the following main steps; in the first step the images at hand are segmented using the K-Means technique, in the second step the segmented images are passed through a pre-trained neural network [3]. This paper presents a survey of different image processing and machine-learning techniques used in the identification of rice plant diseases based on images of disease infected rice plants. This paper presents not only survey of various techniques but also concisely discusses important concepts of image processing and machine learning applied to plant disease detection and classification [4]. This paper has the stages pre-processing step, background removal technique is applied on the image in order to remove background from the image. Then, the background removed images are further processed for image segmentation using otsu thresholding technique. Different segmented images will be used for extracting the features such as color, shape and texture from the images. At last, these extracted features will be used as inputs of classifier. Expand [5]. This paper presents an application of image processing techniques and Support Vector Machine (SVM) for detecting rice diseases. Rice disease spots were segmented and their shape and texture features were extracted. The SVM method was employed to classify rice bacterial leaf blight, rice sheath blight and rice blast. The results showed that SVM could effectively detect and classify these disease spots to an accuracy of 97.2% [6]. The classification and recognition of paddy diseases are of the major technical and economical importance in the agricultural industry. To automate these activities, like texture, color and shape, disease recognition system is feasible. The goal of this research is to develop an image recognition system that can recognize paddy diseases [7]. This article presents a prototype system for detection and classification of rice diseases based on the images of infected rice plants. This prototype system is developed after detailed experimental analysis of various techniques used in image processing operations. They consider three rice plant diseases namely Bacterial leaf blight, Brown spot, and Leaf smut [8]. This paper empirically evaluates four techniques of background removal and three techniques of segmentation. To enable accurate extraction of features, they propose centroid feeding based K-means clustering for segmentation of disease portion from a leaf image. They enhance the output of K-means clustering by removing green pixels in the disease portion. They extract various features under three categories: color, shape, and texture. They use Support Vector Machine (SVM) for multi-class classification [9]. This paper proposes a new variation level set formulation in which the regularity of the level set function is intrinsically maintained during the level set evolution. The level set evolution is derived as the gradient flow that minimizes an energy functional with a distance regularization term and an external energy that drives the motion of the zero level set toward desired locations. The distance regularization term is defined with a potential function such that the derived level set evolution has a unique forward-and backward (FAB) diffusion effect, which is able to maintain a desired shape of the level set function, particularly a signed distance profile near the zero level set [10]. This paper presents the different image processing techniques for paddy disease identification and further classification. The challenges involved in each step of diseases detection and classification are analyzed and discussed.

3. Materials and Methodology

3.1. Dataset and Software Used

The RGB color images of paddy leaf has captured using a Canon PowerShot G2 digital camera, with pixel resolution 768x1024. Those images were cropped into a smaller image. We have collected about 200 data samples. It consists of three types of paddy diseases. We used MATLAB R2015a as developing tool.

3.2. Methodology

Proposed methodology works in three major parts, image preprocessing, image segmentation, feature extraction and classification using multiclass Support Vector Machine as shown in Figure 1.

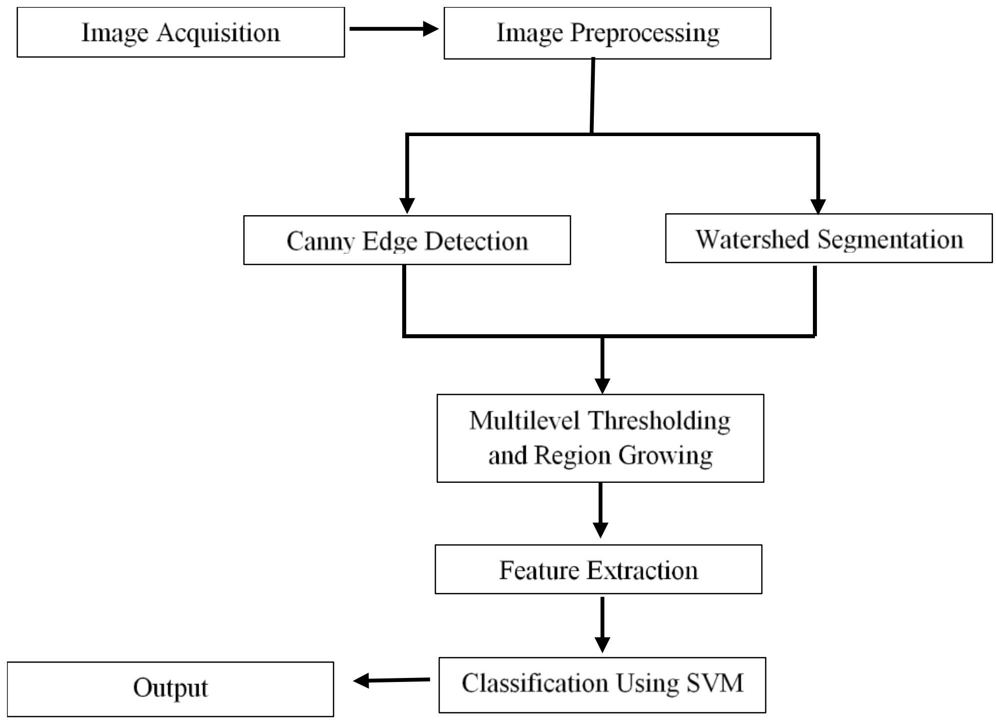

Figure 1: Overview of the proposed methodology

3.2.1. Image Acquisition

The RGB colour images of paddy leaf captured using a Canon PowerShot G2 digital camera, with pixel resolution 768x1024. Those images are cropped into a smaller image. We collect about 200 data samples. It consists of three types of paddy diseases.

3.2.2. Image Preprocessing

The images can be acquired under various lighting conditions and may be affected by the noises. Therefore the pre-processing module is designed to enhance the image quality by removing the uneven intensity and noises in the image

3.2.3. Image Segmentation

Segmentation is grouping the pixels which has the similar features can be done in many ways and a single method would not be robust to segment the affected region from the paddy leaf image. Because of the properties associated with the affected area shows variation in shape, texture and color. Hence combination of multiple features will provide robust grouping of the ROI. To address this challenging characteristics of image conditions, multiple image processing algorithms are employed such as:

- Canny edge detection

Canny edge detector is the optimal and most widely used algorithm for edge detection. Compared to other edge detection methods like Sobel, canny edge detector provides robust edge detection, localization and linking. It is a multi-stage algorithm and in the first stage it will filter out any noise in the original image before trying to locate and detect any edges using Gaussian filter. In the second stage it will find the edge strength by taking gradient of the image using sobel operator. In the third stage it will find the gradient angle which is used to find the edge direction. Once the edge direction is known, in the next stage non-maximum suppression is applied which is used to trace along the gradient in the edge direction and compare the value perpendicular to the gradient. Two perpendicular pixel values are compared with the value in the edge direction. If their value is lower than the pixel on the edge then they are suppressed i.e. their pixel value is changed to 0, else the higher pixel value is set as the edge and the other two suppressed with a pixel value of 0. In the last stage it decides which are all edges are really edges and which are not. For this, we need two threshold values, minVal and maxVal. Any edges with intensity gradient more than maxVal are sure to be edges and those below minVal are sure to be non-edges, so discarded. Those who lie between these two thresholds are classified edges or non-edges based on their connectivity. If they are connected to "sure-edge" pixels, they are considered to be part of edges. Otherwise, they are also discarded. This stage also removes small pixels' noises on the assumption that edges are long lines. So finally we get strong edges in the image.

- Watershed segmentation

Any grayscale image can be viewed as a topographic surface where high intensity denotes peaks and hills while low intensity denotes valleys. Start filling every isolated valleys (local minima) with different colored water (labels). As the water rises, depending on the peaks (gradients) nearby, water from different valleys, obviously with different colors will start to merge. To avoid that, build barriers in the locations where water merges. Continue the work of filling water and building barriers until all the peaks are under water. Then the barriers give the segmentation result. This is the "philosophy" behind the watershed. But this approach gives over segmented result due to noise or any other irregularities in the image. So we used marker-based watershed algorithm where we specified which are all valley points are to be merged and which are not. It is an interactive image segmentation. In this algorithm here we give different labels for our object we know. Label the region which we are sure of being the foreground or object with one color (or intensity), label the region which we are sure of being background or non-object with another color and finally the region which we are not sure of anything, label it with 0. That is our marker. Then apply watershed algorithm. Then our marker will be updated with the labels we gave, and the boundaries of objects will have a value of -1. Finally, the region where they touch are segmented properly Segmentation using the watershed transform works well if you can identify, or "mark," foreground objects and background locations. Marker-controlled watershed segmentation is also a multi stage algorithm. Initially, it uses the gradient magnitude as the segmentation function to find the borders inside the affected region which has high gradient value. After that, it will find the foreground markers, which must be connected blobs of pixels inside each of the foreground objects. In this framework we use morphological techniques called "opening-by-reconstruction" and "closing-by reconstruction" to "clean" up the image. In the next stage, it will find the background

- Multilevel thresholding

This Multilevel thresholding segmentation approach is based on region growing which combines region growing and merging to implement the process. To extract the region seed, the image is first transformed from RGB space to HSV space, the hue value difference between adjacent pixels is calculated, a multi-threshold approach is applied to obtain the seed candidate region, finally region Centre pixel is regarded as the region seed point. At region growing phase, a region homogeneity criterion is applied to implement the region growing, the color intensity value in R, G, B channel is regarded as a three-dimension vector, and three vector space angles are calculated corresponding to R, G, B channel respectively. Only the differences between adjacent space vector angles on each channel are smaller than a preset threshold, then the homogeneity condition is met and the adjacent pixel is added. Edge which is already found in the edge detection is regarded as the stop conditions while region growing is implemented.

3.2.4. Feature Extraction

Feature extraction plays a main role in disease detection to distinguish the affected lesion from the paddy leaf. In this research work mainly we focus on texture features extraction using GLCM (grey level co- occurrence matrix) techniques and we extract color and shape features also.

- Extraction of Texture features

We applied GLCM techniques for the cluster images which are normally grey scale image to analyze the texture features. The extracted features are such as contrast, correlation, energy and homogeneity done by grey co matrix. Contrast is used to differentiate an element by intensity variation. In SGDM, Energy is termed sum of square elements. Whereas Homogeneity is measured based on the distribution of elements in SGDM. Correlation returns a measure of however correlative an element is to its neighbor over the full image.

- Extraction of Color features

We extract the mean values of R, G, and B components in RGB image, H, S and V components in HSV image and L, A and B components in LAB image.

- Extraction of Shape features

The total area of the lesion of a paddy leaf is calculated. Then the number of affected blobs are calculated using blob detection techniques. After that we found proportion area.

3.2.5. Classification and Detection

In improving our models accuracy ability to correctly predict, training is required to be done using a model. In the classification of images, multi class support vector machine (SVM) is a model used to train and predict if an image is affected based on the extracted features. SVM is a supervised learning model used in classification and regression based on characteristics of defined decision boundaries. For a wide range of classification tasks, SVMs are presently among the better performers among various classification models. The standard format of the SVM solves two class problems. From the binary problems can be extended to multiclass SVM with K classes. Where K>2. It has a two approaches, it is one-against-one and one-against-all. After the training phase, the features were extracted to classify the database in multiclass SVM classification. From the 200 images, 160 were used for training and the remaining 40 for testing.

4. Results and Discussion

4.1. Segmentation Results

The input images of paddy leaf captured using a digital camera, with pixel resolution 768x1024. Those images are cropped into a smaller image. We collect about 200 data samples. It consists of three types of paddy diseases. The sample input image is given in Figure 2 and Input image after preprocessing is shown in Figure 3.

Figure 2: Input Image

Figure 3: Input image after preprocessing

After the preprocessing the Segmentation algorithms are applied such as Canny edge detection, Watershed algorithm and multilevel thresholding. The segmented image is shown in Figure 4.

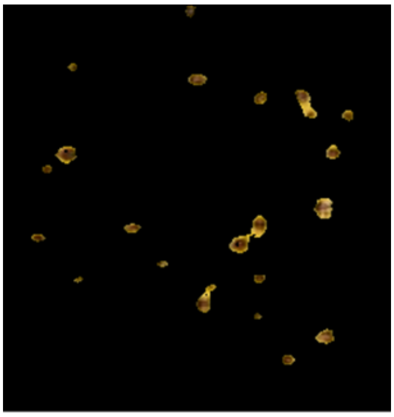

Figure 3: Segmented image

4.2. Classification Results

The accuracy of the classification is calculated by the matrices TP, FP, TN and FN. The classification results of our testing data are: TP = 18, FN = 2, FP = 3 and TN = 17.

The following matrices are calculated from this comparison process.

True positive (TP): Correctly classified the paddy leaf disease

False positive (FP): Falsely classified the paddy leaf disease

True negative (TN): Correctly classified the paddy leaf without any disease

False negative (FN): Falsely classified the paddy leaf without any disease

According to these matrices the sensitivity, specificity, and accuracy are calculated by the following equations:

Sensitivity = TP/ (TP+FN).

Specificity = TN/ (TN+FP).

Accuracy: (TP+TN)/ (TP+FP+TN+FN)

Table 1: Defines sensitivity and specificity parameters for evaluating accuracy for this test data set

	Specificity	Sensitivity	Accuracy	Accuracy rate
And another entry	8.5	0.9	8.75	87.5%

5. Conclusion

Paddy plants are affected by various fungal and bacterial diseases which make a big amount of loss in agriculture and economy. Main objective of this proposed methodology is to develop an automatic system for detect the paddy disease which are Paddy Blast Disease, Brown Spot Disease, Narrow Brown Spot Disease. We prepared our own dataset of paddy leaf images as input of our system. The diseased regions were identified and segmented using canny edge detection, multilevel thresholding and region growing techniques. We extracted texture features using GLCM (grey level co- occurrence matrix) techniques, additionally we extracted color and shape features to improve the accuracy of the framework and used Multiclass Support Vector Machine for classification. We achieved 87.5% accuracy for the test dataset.

6. Limitations of the study

We proposed a methodology to develop an automatic system for detect the paddy disease which are Paddy Blast Disease, Brown Spot Disease, Narrow Brown Spot Diseases. Even though there are more paddy diseases available but we find only three major types of paddy diseases. There are some specific cases in segmentation which are hard to segment the lesion accurately such as overlapped lesions and irregular boundary lesions.

7. Recommendations

We have applied this methodology only for the paddy leaf. In the Future, we are planning to extend this project to detect more paddy diseases and classify disease symptoms affected on fruits, vegetables, commercial crops etc. We will improve the accuracy of this methodology through minimize the drawbacks in segmentation such as overlapping, irregular boundary.

References

[1]. S. Weizheng, W. Yachun, C. Zhanliang, and W. Hongda, "Grading methodof leaf spot disease based on image processing," in Computer Science andSoftware Engineering, 2008 International Conference on, vol. 6. IEEE,2008, pp. 491–494.

[2]. D. Al Bashish, M. Braik, and S. Bani-Ahmad, "A framework for detection and classification of plant leaf and stem diseases," in Signal and Image Processing (ICSIP), 2010 International Conference on. IEEE, 2010, pp. 113–118.

[3]. J. P. Shah, H. B. Prajapati, and V. K. Dabhi, "A survey on detection and classification of rice plant diseases," in 2016 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC), March 2016, pp. 1– 8.

[4]. B. S. Prajapati, V. K. Dabhi, and H. B. Prajapati, "A survey on detection and classification of cotton leaf diseases," in International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016, March 2016, pp. 1–8.

[5]. Q. Yao, Z. Guan, Y. Zhou, J. Tang, Y. Hu, and B. Yang, "Application of support vector machine for detecting rice diseases using shape and color texture features," in Engineering Computation, 2009. ICEC'09. International Conference on. IEEE, 2009, pp. 79–83.

[6]. G. Anthonys and N. Wickramarachchi, "An image recognition system for crop disease identification of paddy fields in sri lanka," in 2009 International Conference on Industrial and Information Systems (ICIIS). IEEE, 2009, pp. 403–407.

[7]. R. A. D. Pugoy and V. Y. Mariano, "Automated rice leaf disease detection using color image analysis," in 3rd international conference on digital image processing. International Society for Optics and Photonics, 2011, pp. 80 090F–80 090F.

[8]. C. Charliepaul, "Classification of rice plant leaf using feature matching," International Journal On Engineering Technology and Sciences, vol. 1, no. 2, pp. 290–295, nov 2014.

[9]. C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," IEEE transactions on image processing, vol. 19, no. 12, pp. 3243–3254, 2010.

[10]. S. T and D. T, "Classification of paddy leaf diseases using shape and color features," International Journal of Electrical and Electronics Engineers, vol. 7, no. 1, pp. 239–250, 2015.

[11]. A. K. Singh, R. A, and B. S. Raja, "Classification of rice disease using digital image processing and svm," International Journal of Electrical and Electronics Engineers, vol. 7, no. 1, pp. 294–299, 2015.

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

[13]. M. A. A. KAHAR, S. MUTALIB, and S. ABDUL-RAHMAN, "Recent advances in mathematical and

computational methods," in Recent Advances in Mathematical and Computational Methods, pp. 248–257.

[14]. J. W. Orillo, J. D. Cruz, L. Agapito, P. J. Satimbre, and I. Valenzuela, "Identification of diseases in rice plant (oryza sativa) using back propagation artificial neural network," in Humanoid, Nanotechnology,

[15]. Technology, Communication and Control, Environment and Management (HNICEM), 2014 International Conference on. IEEE, 2014, pp. 1–6.

[16]. L. Liu and G. Zhou, "Extraction of the rice leaf disease image based on bp neural network," in Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on. IEEE, 2009, pp. 1–3.

[17]. N. N. Kurniawati, S. N. H. S. Abdullah, S. Abdullah, and S. Abdullah, "Investigation on image processing techniques for diagnosing paddy diseases," in Soft Computing and Pattern Recognition, 2009. SOCPAR'09. International Conference of. IEEE, 2009, pp. 272–277

[18]. K. Majid, Y. Herdiyeni, and A. Rauf, "I-pedia: Mobile application for paddy disease identification using fuzzy entropy and probabilistic neural network," in Advanced Computer Science and Information Systems (ICACSIS), 2013 International Conference on. IEEE, 2013, pp. 403–406.

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

[20]. S. Phadikar and J. Sil, "Rice disease identification using pattern recognition techniques," in Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on. IEEE, 2008, pp. 420–423.

[21]. G. Maharjan, T. Takahashi, and S. Zhang, "Classification methods based on pattern discrimination models for web-based diagnosis of rice diseases," Journal of Agricultural Science and Technology, vol. 1, no. 1, pp. 48–56, 2011.