Classification of Thrips-infected Chili Plant Using Image Processing on Embedded System

M.H. Almaswari[1], N.Nordin[1,2*] and Mariam Md Ghazaly[1,2]

[1] Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
[2]Center for Robotics and Industrial Automation (CeRIA), Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.

*Corresponding author's email: majd714040908@gmail.com

ABSTRACT: In chili plantations, infestation by insects bacterial, and fungal diseases are the major constraints in chili production. This paper embarks on classifying the leaves infested by thrips from healthy leaves. Moreover, images were captured using the camera and processed onboard the Raspberry Pi. The K-means algorithm is utilized for image segmentation. Texture features and shape features were computed from the processed images before it is fed to Linear Support Vector Machine algorithm (SVM) for supervised learning. The results were analyzed in the form of a confusion matrix. Based in the results, the Linear SVM classifier shows an accuracy of 80%, specificity of 75%, precision of 78% and Sensitivity of 84%.

Keywords: *chili; precision farming; leaves classification; pest control.*

1. INTRODUCTION

Precision agriculture is at the helm of the interest in modern farming technology. According to Abdullahi et al. in his proposed cycle of precision agriculture [1], stress detection algorithm that caters the needs of water, nutrient, detection and removal of weeds and insects are the required step before the prescription plan of how much water, fertilizer, herbicide and pesticide is executed.

In Malaysia, the average domestic demand for chillies was at 55,420 tonnes a year in 2018, while the local production was only at 24,428 tonnes [2]. The needs to increase the production has been at all-time high. Recognizing the needs to impart precision agriculture in this industry, this paper explains the attempt to classify the thrips-infected leaves of the chilli plant as part of the stress detection algorithm to improve chilli production.

When the chilli plants are infected by thrips, the leaves become curled which can be seen in Figure 1. The infected leaves are also difficult to be recognized from afar because the color of the infected ones are like healthy ones. Sometimes the leaves of the infected plants are overlapped with the healthy ones which further poses the challenge to identify them.

Hence, the objectives of this study are to determine the features that differentiate between thrips-infected leaves and healthy leaves, to classify the thrips-infected leaves from the healthy leaves, and to evaluate the classification performance with four parameters, specificity, accuracy, sensitivity, and precision.

Figure 1 Thrips-infected chili leaves (right) and healthy ones (left)

2. METHODOLOGY

The process of classifying the leaves infested by thrips from healthy leaves goes through several stages to achieve the intended results. Figure 2 shows the flowchart of this research methodology.

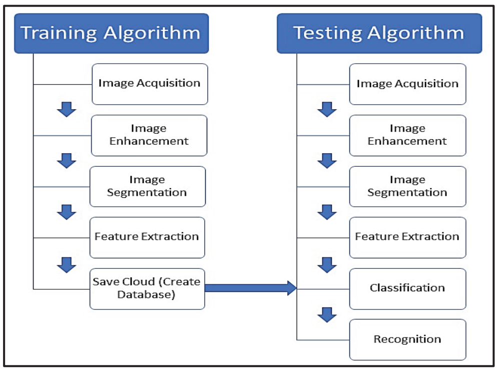

Figure 2 Flowchart of suggested work

The image is first acquired using a mobile phone camera (Samsung Note 10, 12 MP, f/1.5-2.4, 27mm lens) at 1 meter from the subject under daylight condition. Image enhancement is then performed to improve image contrast. Then, the leaves are extracted from the background using K-means segmentation. The algorithm identified 3 centroids of color from the image, then allocates every pixel's three-color-channel intensity value to the nearest cluster, while keeping the centroids as small as possible. Since the pictures were taken with the intention that the leaves are approximately 70-80% from the field of view, the green cluster was always chosen as one of the color clusters, thus segmented the leaves from the background.

After that, the segmented images are converted to grayscale for the computation of Gray-Level Co-Occurrence Matrix (GLCM), C_M. For any grayscale image, I(k,k) with the central pixel (n_c,m_c), the co-occurance matrix, C_M is defined as in Equation 1.

$$C_M = \sum_{n=1}^{k} \cdot \sum_{m=1}^{k} \begin{cases} 1, \text{ if } I(n,m)=k \text{ and } I(n+D_x, m+D_y)=k \\ 0, \text{ otherwise} \end{cases}$$

where (D_x, D_y) are defined as

$$D_x = D\cos\theta, D_y = D\sin\theta \tag{1}$$

From this matrix, the correlation, energy, and homogeneity are computed [3] and are taken as features for classification using Support Vector Machine (SVM) with linear, quadratic and Gausssian kernels.

3. RESULTS AND DISCUSSION

Image enhancement was conducted to improve the contract of the leaves from the background as can be seen in Figure 3.

Figure 3 Image Filtering for Chili Plant

Figure 4 depicts the one of the results of the k-means segmentation that was conducted based on the improved image.

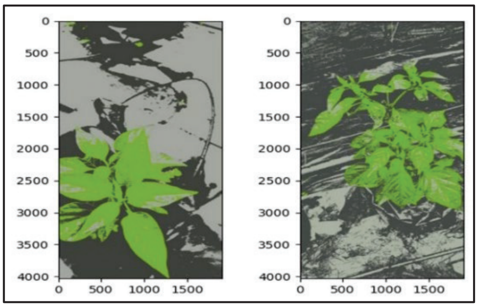

Figure 4 Infected (right) and healthy (left) after segmentation

The segmented images are then converted to grayscale to construct the GLCM and the features for classification. The are 200 images that were used in this study with a balanced count of infected and healthy leaves. 140 images were used for the supervised training which is 70% of the dataset. Three different kernels; linear, quadratic and Gaussian are compared when training the SVM.

Table 1 depicts the performance of each kernel. The model was then used on 60 images that were randomly taken out from the original 200 images for the testing. It can be concluded that linear kernel has outperformed other kernels in classifying the infected leaves.

Confusion matrix along with the measure of accuracy, precision, specificity and sensitivity were used to evaluate the performance of classified images. Figure 7 depicts the confusion matrix for the classification of the thrips-infected leaves on the unseen test data.

Table 1 Performance Evaluation of Results

Kernel	Accuracy	Precision	Recall	F1
Linear	72%	71%	70%	72%
Quadratic	60%	60%	61%	69%
Gaussian	63%	62%	64%	53%

		Predicted label	
		Healthy	Infected
True label	Healthy	20	8
	Infected	9	23

Figure 7 Confusion Matrix Results

4. ANALYSIS

Based on the training result, it is evident that the accuracy of the classification is merely 72%. This is expected considering the difficulty of outlining the curly shaped leaves especially when it is not dominant in the picture. However, when tested with the unseen data, the accuracy rose to 80%, which may suggest that the model used was not overfitted considering the test data was almost balanced.

5. CONCLUSION

The use of GLCM and its corresponding measures are beneficial as features to outline the infected leaves from healthy ones. However, the current state of achievement is not yet optimal. Further works should include increasing the database and optimizing the features for the classification,

ACKNOWLEDGEMENT

Authors are grateful to Universiti Teknikal Malaysia Melaka and Ministry of Education Malaysia for the financial support through FRGS grant no. FRGS/2018/FKE-CERIA/F00353.

REFERENCES

[1] H. Abdullahi and R. Sheriff, "Case Study to Investigate the Adoption of Precision Agriculture in Nigeria Using Simple Analysis to Determine Variability on a Maize Plantation," 2017.

[2] S. A. Shah, "Mardi launches new technologies for better banana, chilli production," The Malaysian Reserve, Feb. 19, 2021.

[3] L. S. Athanasiou, D. I. Fotiadis, and L. K. Michalis, "4-Plaque Characterization Methods Using Intravascular Ultrasound Imaging," in Atherosclerotic Plaque Characterization Methods Based on Coronary Imaging, Oxford: Academic Press, 2017, pp. 71–94.