Evolution in Electrical and Electronic Engineering Vol. 2 No. 2 (2021) 834-841

EEEE

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/eeee
e-ISSN : 2756-8458

The Disease Detection for Maize-Plant using K-Means Clustering

Muhamad Amirul Mohd Yusof[1], Ain Nazari[1]*

[1]Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/eeee.2021.02.02.100
Received 04 July 2021; Accepted 07 October 2021; Available online 30 October 2021

Abstract: The quality care for maize plant should be emphasized by the farmers to ensure the maize that produced is healthy and non-diseased. People usually use guide books or website to recognize the maize plant disease and this consuming a lot of time and hard to bring the accurate result. An automated detection system can help to identify the types of maize plant disease with more efficient. The purpose of this research is to design a system that can analyze plant image using image processing method. The proposed method consists of five stage which is pre-processing, image segmentation, feature extraction and classification. The histogram equalization and median filtering algorithm is used in pre-processing process. Then for the segmentation, there are two part which are leaf area and disease area. The thresholding, masking and k-means algorithm is used to perform this segmentation. In feature extraction process, there are 13 features need to extract from the image to perform classification in Support Vector Machine (SVM) classifier. Hence, from the findings of the study, it indicates that the system is able to identify the type of maize plant disease correctly with the average accuracy, sensitivity and specificity for healthy leaf is 97.53%, 90.08% and 85.4%. While for the overall disease leaf is 97.27%, 90.82% and 84.93%. In conclusion, this system has developed an automated system to identify the types of diseases for maize plant.

Keywords: SVM Classifier, *k*-means Clustering, Maize-Plant Disease

1. Introduction

Maize plant is the one of agricultural product which is in a lot demand in domestic or international market. Beside has been one of food source in human food chain, maize plant also has declared as agricultural product that can contributes in country economy [1]. So to maintain the productivity of maize plant, the farmer needs to ensure the maize produced are in good quality and free from disease. The early identification of maize plant disease is the one of technique that can prevent the maize plant produce infected by disease. The manual identification requires expertise, time and accuracy [2]. Hence, an automated detection system is needed to help farmer identifying the types of maize plant disease with more efficient.

*Corresponding author: ain@uthm.edu.my

The uses of image processing technology in many industries now days has been rising from day to day. This is because this technology can perform the system that can identifying and recognition the object from image. Image processing method works by applying the operation on the image and then will analyze to perform desire output. For example, such as in medical researches, the image processing has been use to recognize and detect the tumor in the human body. Related to the current technology developments, this method also has been most useful in face recognition and fingerprint recognition [2]. In this research, the image processing method will be focus on detection disease of the maize-plant such grey leaf spot, rust and leaf blight.

This paper about to propose an automated system that can identify the types of diseases for maize plant by implement of *k*-means clustering and SVM classifier. *K*-means clustering is the algorithm that used to cluster the object based on the feature of the leaf in *k* number of groups and it's done by using the Euclidean distance metric [3]. By applying this clustering the disease part can be segmented from the image. While SVM classifier is use to make a classification for the image that has been segmented. SVM classifier is set of related supervised learning method used for the process of classification and regression [4]. Based on the research from Budiarianto Suryo Kusumo and et al. support vector machine (SVM) is a classifier that aims to find a hyperplane that separate the support vector the farthest [1].

From Saradhambal. G, Dhiyvya. R, Latha. S, R. Rajesh, they proposed project that used to detect the plant disease and provide solution to recover from disease. The steps that involve is such image acquisition, image pre-processing, image segmentation, feature extraction and classification. For the pre-processing, the image has been enhancing by eliminate the noise from image. The Otsu classifier and k-means clustering is use to create segmentation. The feature extraction is use to predict the infected region [5]. While for proposed from S.Jayamoorthy and Dr. N.Palanivel, the detection system that been design is consists of image acquisition, pre-processing image, thresholding, masking and genetic algorithm to segment components. The clipping process has been applied to get the interested image region. Then the image smoothing is done by using smoothing filter. The proposed has use Spatial Fuzzy C-Means clustering and genetic algorithm to create segmentation on leaf [6].

The detection system for the flower recognition by Huthaifa Almogdady, Dr. Saher Manaseer, Dr.Hazem Hiary has proposed a flower recognition approach based on image processing technique and Artificial Neural Networks (ANN) algorithm. The image enhancement is performing to making flower images more data extractable, clearer and more useful. The Gray Level Co-occurrence Matrices (GLCM) is use to extract the flower image. By using Back Propagation ANN classification, the type of flower can be recognizing [2]. Besides that, the plant disease detection that proposed by Sachin D. Khirade and A. B. Patil is also perform the classification by using neural network. The back propagation algorithm, is apply to modified self-organizing map (SOM). As the result Multiclass Support vector machines can be used [7].

The structure for this study is presented as follows; Section 2 is a part explanation for method use in the study. While for section 3, it discusses about the result that achieve from study. Finally, the overall summary for this study is conclude in Section 4.

2. Research Methods

The proposed system generally is divided to five main process which is image acquisition, preprocessing, segmentation, feature extraction and classification. The MATLAB software is used to design algorithm for this system such histogram equalization technique, threshold and masking method, GLCM algorithm, k-means clustering and SVM classification. Figure 1 illustrates the block diagram for the proposed system.

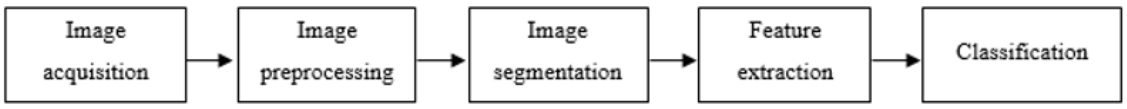

Figure 1: Block diagram for proposed system

2.1 Software Development

To develop system of this project, MATLAB software is used to implement the algorithm for each process in the system. Each method and algorithm that been design in this system is related and play an important role to achieve the expected result. The tools in MATLAB software such image processing tools, make the process to analyze the image of maize leaf more efficient. The variety of MATLAB tools also make this software widely used as a computational tool in science and engineering field [8]. By using MATLAB software, it allows for this system to analyze the image data, performing algorithm and design applications.

2.2 Image Acquisition

Image acquisition is the process of retrieving image from some source. For this system, the image of the maize plant that has been select must be in RGB color space and the image need to be collected in one file of dataset. The image of maize plant for train data and test data is collect from Kaggle website [10]. There are 4188 images of the maize plant consist of 1162 healthy leaf, 1146 blight leaf, 574 grey leaf spot and 1306 common rust leaf image from this website. From this total, 30 images for each condition of maize leaf has been choose to create a new dataset for the system.

2.3 Image Pre-processing

Image pre-processing is the process to make improvement for image data to be use in further analysis. In this project, the image clipping is applying to create the region of the leaf that infected and called as clipping image. This make the analyze process for leaf image easier. From the output of clipping image, the image will apply contrast enhancement by using Histogram Equalization algorithm. This method is use to adjust image intensity to enhance contrast. By applying this method, the image color can be improving and make the segmentation process more accurate. The image that enhanced will perform median filtering to reduce the noise from the image. The step is by apply the noise into the image before the median filter is perform into the image. This process can improve the image processing process such in edge detection. Figure 2 shows the output by applying clipping, histogram equalization and median filter.

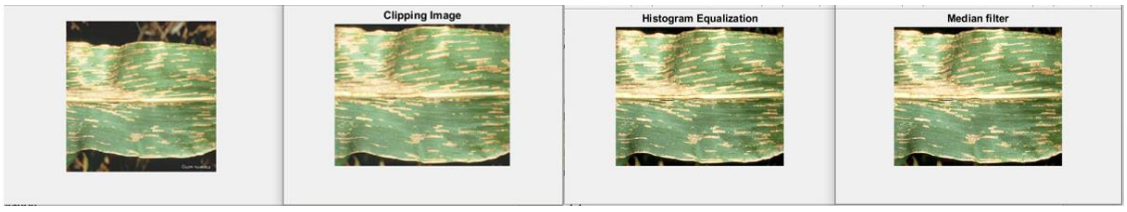

Figure 2: Output for image pre- processing

2.4 Image segmentation

The image segmentation in this project is divided into 2 phase. The first phase is focus on separation of leaf image with their background. Color Thresholder apps from MATLAB can be perform to create this segmentation part. This process is necessary to simplify the process of detection area disease on the maize leaf image. In this Color Thresholder apps, the threshold and masking process has been perform to obtain the leaf image. Figure 3 shows the result from the threshold and masking process.

Figure 3: Output for threshold and masking

The second phase, the segmentation is creating to segment the disease part from the maize leaf. For this phase, k-means clustering algorithm is applying to identify disease part based on set of feature that has been clustered. The value of the clustering image is depending on k-number of class that set. For this project the k number is set into two cluster. So, the result from this clustering will displays two region of interest (ROI) that identified from the leaf image. The *k*-mean clustering for healthy and disease part is display in Figure 4.

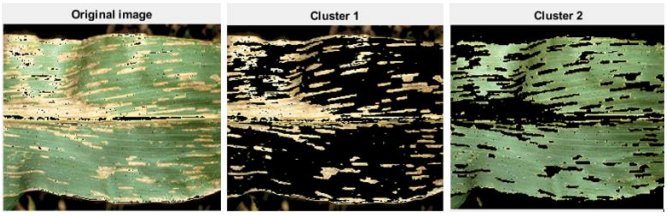

Figure 4: Output for *k*-mean clustering

2.5 Feature Extraction

To perform feature extraction, the image that produce from the segmentation need to apply Gray level co-occurrence matrix [7]. Feature extraction technique is needed in this system to perform classification by using SVM classifier. In this system, texture feature is calculated based on parameter of contrast, energy, homogeneity, mean, standard deviation, entropy, RMS, variance, smoothness, IDM, correlation, kurtosis and skewness for the image analyze. For the two dimensional equation, f(a ,b) image is refer as image, N is refer as total number gray levels, h(i) is refer as intensity level and p(i) is refer as probability density.

- Contrast

Contrast feature can be determining by using intensity of threshold and the nearest pixel of the image. The contrast is expressed in mathematical form like Eq. 1.

$$\text{Contrast} = \sum_{n=0}^{N-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i,j)^2 \quad \textit{Eq. 1}$$

- Energy

Homogeneity is calculated by the energy. It is called as angular second moment or uniformity. The energy is mathematically expressed in Eq. 2.

$$\text{Energy} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i,j)^2 \quad \textit{Eq. 2}$$

- Mean

Eq. 3 represents the mathematical form of mean. It clearly shows that the mean is a function probability density. Mean is the average level of intensity of image.

$$\text{Mean} \quad \mu = \sum_{i=0}^{N-1} i, p(i) \quad \textit{Eq. 3}$$

- Standard deviation

The mean value of the pixels and their probability densities are used to measure the standard deviation like expressed in Eq. 4.

$$SD \text{ or } \sigma = \sqrt{\sum_{i=0}^{N-1} (i-\mu)^2 . p(i)} \quad Eq.\ 4$$

- Entropy

Entropy is measure by uncertainty of the random variable. It depends on the probability density p (i) and the mathematical equation for entropy is express as Eq. 5.

$$Entropy \text{ or } En = -\sum_{i=0}^{N-1} p(i) \log_2[p(i)] \quad Eq.\ 5$$

- Variance

Eq. 6 shows the mathematical equation for variance feature and its calculate by squaring the standard deviation. The variation intensity is measured with the help of variance.

$$Variance \text{ or } \sigma^2 = \sum_{i=0}^{N-1} (i-\mu)^2 . p(i) \quad Eq.\ 6$$

- Inverse Difference Moment (IDM)

The local homogeneity of an image is calculated by IDM and the Eq. 7 performs the mathematical calculation for IDM.

$$IDM = \sum_{i=0}^{N-1} \sum_{i=0}^{N-1} \frac{1}{1+(i-j)^2} p(i,j) \quad Eq.\ 7$$

- Correlation

Eq. 8 expresses the mathematical form of correlation. It is used to measure the relationship between the threshold and nearest pixel in images.

$$Correlation = \frac{1}{\sigma_a \sigma_b} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i,j).p(i,j)^2 - \mu_a \mu_b \quad Eq.\ 8$$

- Kurtosis

The histogram flatness is measured by kurtosis. Eq. 9 expresses the mathematical form kurtosis which depends on the standard deviation, mean and probability density.

$$Kurtosis \text{ or } \mu^4 = \sigma^4 \sum_{i=0}^{N-1} ((i-\mu)^4 . p(i)) - 3 \quad Eq.\ 9$$

- Skewness

Symmetry of an image is defined by the skewness. It is denoted by μ3 and is shown in Eq.10.

$$Skewness \text{ or } \mu^3 = \sigma^{-3} \sum_{i=0}^{N-1} ((i-\mu)^3 . p(i) \quad Eq.\ 10$$

2.6 Classification

In the classification process, the algorithm use is Support Vector Machine (SVM) classifier. The classification is make to identify either the maize plant is healthy or unhealthy. There are 3 type of maize plant disease is choosing for this study which are leaf blight, corn grey leaf spot and corn common rust. In SVM algorithm, each data that has point in number of feature for dimensional space is plotted with the value of particular coordinate. The classification is performing by hyper-plane finding that differentiates the data into two classes. The implementation of Kernel Trick also is needed in this type classifier to solve the not separable problem. This because Kernel Trick can transform the low dimensional input space into the higher dimensional [1].

2.7 Performance Measurement

To analyze the result and performance of this system, each algorithm has been measure the effectiveness by using a specific parameter [9]. From pre-processing until segmentation process, the result is measure based on mean-square error (MSE) and peak signal-to-noise ratio (PSNR) of the image. The equation is express in Eq. 11 and Eq. 12. The lower value of MSE, the higher performance of algorithm. While the higher PSNR value, the better image quality will produce.

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N} \quad \textit{Eq. 11}$$

$$PSNR = 10 \log_{10}\left(\frac{R^2}{MSE}\right) \quad \textit{Eq. 12}$$

For feature extraction and classification process, the result is analyzed based on their accuracy, sensitivity and specificity. The terms true positive (TP), false positive (FP), true negative (TN) and false negative (FN) is use as reference for the accuracy, sensitivity and specificity calculation. Accuracy measurement in image processing is define as percentage of correctly classified in classification system. For sensitivity, it has defined as measurement value that measure the quantity of positives that are correctly identified for the image while specificity is define as measurement value that measure the quantity of negatives that are correctly identified for the image. Table 1 shows the formula list to calculate the system performance.

Table 1: System performance equation

Measure	Performance formula (%)
Accuracy	$\frac{(TP + TN)}{(TP + FP + TN + FN)} \times 100$
Sensitivity	$\frac{TP}{(TP + FN)} \times 100$
Specificity	$\frac{TN}{(TN + FP)} \times 100$

3. Results and Discussion

The result and performance for each process has been evaluated based on their parameter. For pre-processing, thresholding and masking segmentation and *k*-mean clustering, the performance has been measure by using MSE and PSNR. While for the classification process, the performance has been measure by using accuracy, specificity and sensitivity. Resulting from this measurement, the effectiveness of method and algorithm used can be analyzed by comparing the result between healthy and unhealthy leave. The training process for this system has been conduct by using 30 image for healthy leave and 30 image for each type of disease leave.

Table 2 compared the result of the MSE and PSNR average for healthy and unhealthy leave in each process. From the result, it shows that the MSE value for each process is close to zero and this indicates the good result for each output of algorithm. The close MSE value to zero, the better result will produce. For PSNR, the result shows the quality of image has been improving from pre-processing to *k*-mean clustering. The high value of PSNR, the better quality image will produce [9].

Table 2: MSE and PSNR results for each process

Process	Healthy Leaf		Blight disease		Grey leaf spot		Rust	
	MSE	PSNR	MSE	PSNR	MSE	PSNR	MSE	PSNR
Pre-processing	0.0009	78.81	0.0009	79.04	0.0008	79.97	0.0008	79.61
Threshold/Mask	0.0266	64.92	0.0277	64.95	0.0173	66.80	0.0172	66.64
k-mean	0.0799	60.15	0.1064	58.03	0.1926	55.54	0.1030	58.31

The performance for classification part has been measure by accuracy, sensitivity and specificity. To identify accuracy of the system, 500 iterations has been creating to identifying accuracy of linear Kernel. From the 500 iteration, the maximum value for the accuracy is choose as the final accuracy for the classification. From the average accuracy result in Table 3, it shows the high percentage for the classification is on healthy leaf. However, for the disease leaf condition, the accuracy also has been produce in high percentage. This result reveals that this system able to classify the type of maize plant disease correctly.

For the sensitivity and specificity result, it has been analyzed that the performance to measure the proportion of true positive and true negative correctly in each image has result above 80% such in result in Table 3. Although, the value of specificity and specificity has produce not consistent value, but it still produces a good result because the average result for sensitivity is high than specificity result. This result indicates the system can measure the true positive rate in training image very well. However, there are some image that has high specificity than sensitivity. This is effect from a few image that cannot be identifying their feature properly based on dimensional space of the system.

Table 3: Average performance result for classification

Disease/Symptoms	Average performance measure		
	Accuracy	Sensitivity	Specificity
Healthy leaf	97.53%	90.08%	85.4%
Blight disease	97.3%	91.64%	83.47%
Grey leaf spot	97.2%	91.75%	84.73%
Rust	97. 3%	89.07%	86.6%

Figure 5 shows the comparison of accuracy, sensitivity and specificity for healthy leaf and disease leaf. The comparison has been made by compared the average of healthy and total average of disease leave performance. The graph illustrates that the accuracy performance for healthy leave is higher than disease leave. For the sensitivity, the disease leave has higher sensitivity than healthy leave while for specificity performance, the result is contrast which is the healthy leave has higher specificity than healthy leave. Based on this three performance result, it can see that there is not much difference in comparison of performance for healthy leave and disease leave. But, the classification for healthy leave still has high accuracy performance than the disease leaves.

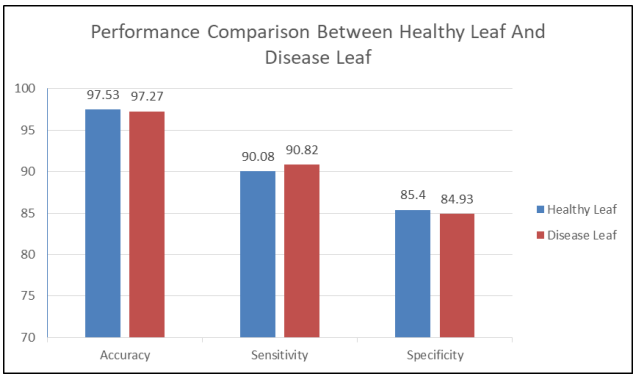

Figure 5: Performance comparison between healthy leaf and disease leaf

4. Conclusion

This study is about the automated detection system for maize-plant disease by implementation of *k*-means clustering. By using image processing tools in MATLAB R2016a software, pre-processing, segmentation, feature extraction and classification process is successful to be perform in designing the disease detection system for maize plant. The SVM classifier algorithm is take action in classifying type of disease based on the image that segmented. To design a system that can classify healthy and unhealthy maize plant, segmentation process has played an important role to separate disease part and healthy part from the image. By creating segmentation from the image, the process to identifying type of disease has been easier. Based on the result of accuracy, specificity and sensitivity, it shows that the process to identifying and classifying the disease part has been done successfully. However, there are some lack in segmentation part which is the PSNR and MSE value need to improve to achieve the better performance. The using of suitable algorithm for pre-processing and segmentation process is the suggested to produce the high accuracy, sensitivity and specificity in classification part.

Acknowledgement

The authors would like to thank the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia for its support in development of this project.

References

[1] B. S. Kusumo, A. Heryana, O. Mahendra, and H. F. Pardede, 'Machine Learning-based for Automatic Detection of Corn-Plant Diseases Using Image Processing', 2018 International Conference on Computer, Control, Informatics and its Applications, p. 93-97, 2018

[2] H. Almogdady, D. S. Manaseer, and D. Hiary, 'A flower recognition system based on image processing and neural networks', International Journal of Scientific & Technology Research, vol. 7, p. 166-173, 2018

[3] S. R, Y. S. Kumar, and G. U. Akhil, 'Leaf Disease Detection using Image Processing', Journal of Chemical and Pharmaceutical Sciences, vol. 10, p. 670-672, 2017

[4] S.Pavithra., A.Priyadharshini., V.Praveena., and T.Monika., 'Paddy Leaf Disease Detection Using SVM Classifier', International Journal of communication and computer Technologies, vol. 03, p. 16-20, 2015

[5] Saradhambal.G., Dhivya.R., Latha.S., and R.Rajesh., 'Plant Disease Detection and its Solution using Image Classification', International Journal of Pure and Applied Mathematics, vol. 119, no. 14th ed., p. 879-884, 2018

[6] S.Jayamoorthy., and D. N.Palanivel, 'Identification of Leaf Disease Using Fuzzy C-MEAN and Kernal Fuzzy C-MEAN and Suggesting the Pesticides', International Journal of Advanced Research in Science, Engineering and Technology, vol. 4, p. 3852-3855, 2017

[7] S. D. Khirade, and A. B. Patil, 'Plant disease detection using image processing', International Conference on Computing Communication Control and Automation, vol. 153, p. 768-771, 2015

[8] B. Larkins and W. Harvey, "Introductory computational science using MATLAB and image processing," Procedia Computer Science, vol. 1, no. 1, pp. 1–9, 2010

[9] Nur Syazlin, Ain Nazari and et al., "Retina Blood Vessel Extraction based on Kirsch's Template Method," Indonesian Journal of Electrical Engineering and Computer Science, vol 14, No. 3, pp. 1-8, 2019

[10] S. Ghose, "Corn or Maize Leaf Disease Dataset", Kaggle. 11-Nov.-2020. [Online]. Available: https://www.kaggle.com/smaranjitghose/corn-or-maize-leaf-disease-dataset. [Accessed: 11-Oct.-2020]