IDENTIFICATION OF DISEASE ON LEAVES SOYBEAN USING MODIFIED OTSU AND LEARNING VECTOR QUANTIZATION NEURAL NETWORKS

[a]Candra Dewi, [b]Muhammad Sa'idul Umam, [c]Imam Cholissodin

[a,b,c] Informatics Department, University of Brawijaya, Malang, 65145, Indonesia
[a] Essential Oil Institute, University of Brawijaya, Malang, 65145, Indonesia
E-mail: dewi_candra@ub.ac.id

Abstract

Disease of the soybean crop is one of the obstacles to increase soybean production in Indonesia. Some of these diseases usually are found in the leaves and resulted to the crop become unhealthy. This study aims to identify disease on soybean leaf through leaves image by applying the Learning Vector Quantization (LVQ) algorithm. The identification begins with preprocessing using modified Otsu method to get part of the diseases on the leaves with a certain threshold value. The next process is to identify the type of disease using LVQ. This process uses the minimum value, the maximum value and the average value of the red, green and blue color of the image. The testing conducted in this study is to identify two diseases called Peronospora manshurica (Downy Mildew) and phakopsora pachyrhizi (Karat). The result of testing by using 50 training data and the value of all recommendations parameters obtained the highest accuracy of identification is 95% %, but the more stable accuracy is 90%. This result shows that the method perform quite well identification of two mentioned disease.

Keywords: digital image, Learning Vector Quantization, leaf soybean disease, modified Otsu.

INTRODUCTION

According to data from the Central Bureau of Statistics, Indonesia experienced a decline in soybean production from 2011 until 2013. In 2011, soybean production was 851 thousand tons with the harvest area of 622 thousand hectares. But, in 2012 the production decreased to 843 thousand tons with 568 thousand hectares of harvest area. The production dropped to 779 thousand tons with 554 thousand hectares of harvested area in 2013 [1]. Meanwhile, according to the Center for Socio-Economic and Agricultural Policy, the consumption or the needs of soybeans for Indonesia in 2011 reached 2 million tons. The research conducted by the Centre for Socio-Economic and Agricultural Policy, one of the factor causes the decreasing in domestic soybean production is the high risk of pests and diseases attack [2].

The diseases in soybean are caused by pathogenic fungi, bacteria, viruses, and mycoplasma. The efforts to distribute the information to farmers about pathogens that cause plant diseases and technological control is very important [3], [4]. Moreover, early detection is also needed so that the disease is not rapidly spread to the other leaves or the other plant. The identification can be done by looking directly at the leaves that affected by the disease, but sometimes quite difficult to distinguish between one disease to another. To ensure the types of diseases that attack, commonly the farmers send affected leaf samples to the laboratories. This of course requires considerable time and costly. To that end, automatic identification of the disease in soybean leaves by using current technology can be used to simplify this process. This method of course will help the farmer quickly identify the infected leaves and reduce the cost. By quickly identification, it will help farmer to remove the infected leaves as soon as possible and avoid the spreading of the disease to other leaves and thus it may help to increase the production of soybean.

In line with current advanced technology, the use of intelligent systems is needed to automate the process of identification of the disease on soybean leaves. Implementing a learning algorithm could contribute to recognize more precise physical characteristics of leaves compared with human vision. Learning Vector Quantitation is a classification method on artificial neural network that work based on learning process to identify the target. This method had been applied for some fields such as classifying acrosome status of boar spermatozoa [5], identifying of nonwoven uniformity [6], generating prototype-based rules [7], predicting protein interactions from protein sequences [8]. There also some research that implement LVQ and give optimum result [9]–[11].

Beside, the comparation of LVQ and the other method also been discussed. The research conducted by Desylvia perform comparation of SOM and LVQ for face recognition using wavelet as feature extraction. The results show that LVQ give better accuracy than SOM with the accuracy of 97.894% for SOM and 100% for LVQ [12]. The other research compares LVQ and backpropagation for classification of diabetes mellitus. This research conclude that LVQ provide higher accuracy rate than backpropagation with the result of 82.56% for LVQ and 73.25% for backpropagation [13].

LVQ recognition results will be optimal if the network received a great input. In the case of leave image that contain disease, it will have hundreds to thousand pixels with high value of variation. Here, we need preprocessing method to reduce the number of input and to obtain the best values for features. This paper uses modified Otsu method to select only the part of disease on leave image as an input of recognition. Otsu is wide range used for segmentation and provide the optimal result [14] [15]–[17]. This method search the optimal threshold values that can distinguish the pixels value of the leaves and the diseases. This paper implement modified Otsu proposed by [18] to create two threshold values for two disease, i.e downy meldew and karat. After that, to obtain the optimum feature is needed the feature extraction. Based on physical observations, both diseases can be distinguished by the color. Therefore, this paper perform color feature extraction to get the optimum features.

MATERIAL AND METHODS

In this research, we develop the application to identify the disease on the leaves. The application works by performing several stages as shown in Figure 1.

The input of the application is the image of leaf. The first step is preprocessing to take part of disease that attacks the leaves with the Modified Otsu thresholding techniques. The preprocessing is necessary to determine whether the leaves contain the disease or not. If the leaves are not diseased, then the identification process is stopped and otherwise the process is continued. The next step is to take the value of each pixel of the disease and perform calculation to obtain the specified value of color feature (color feature extraction). Then is done LVQ Neural Network algorithm to recognize the type of disease. To determine the performance of the used algorithms and methods, testing accuracy were done by comparing results of the application with actual data of disease. The explanations of each part of the process is described in the following sections.
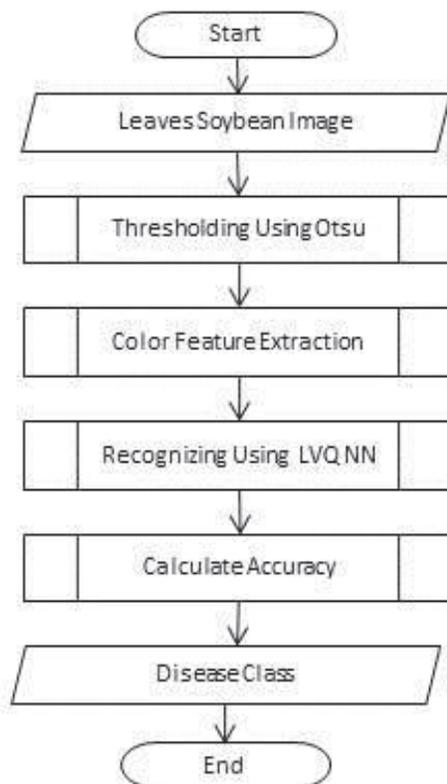

Figure 1. General flow of process

Data Input

The leaves data having the disease were collected from Institute for Agricultural Technology (BPTP) *Karangploso* and the Research Institute for Legumes and Tuber (*Balitkabi*) *Kendalpayak* in Malang, East Java. There are two kinds of diseases that attack the leaves to be identified in this study called *Peronospora manshurica* (Downy mildew) and *phakopsora pachyrhizi* (Karat).

The image of the leaves were acquired using mobile phone camera 5 Mega Pixel with resolution 2592 x 1944 pixel. The leaves are still attached on the plant at the time the photograph was taken. The distance of shooting is approximately 10-15 cm. After that, the background of the image changed to white using image processing software.

The total image having disease captured in the field are 30 images, which each leaf normally consists of several spotting disease. Then the image of the leaves are cropped to the size of 200 x 200 pixels to get a variation example of the form of each disease that will be used in the recognition process. The total number of cropped image is 96 that consist of 48 images have Karat and 48 images have Downy mildew disease. The example of leave images with disease and cropped images are shown in Figure 2.
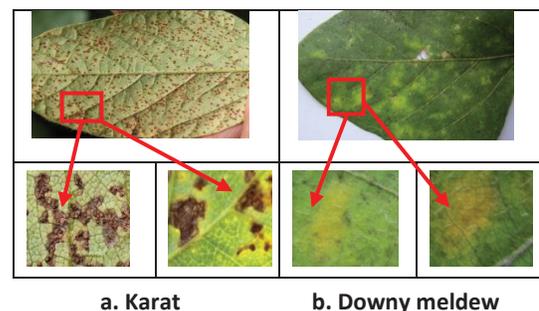

a. Karat   **b. Downy meldew**

Figure 2. The example images of karat and downy mildew deasease

Thresholding Using Modified

Otsu This study perform preprocessing in the form of image thresholding by using modified Otsu method. This process is aimed to determine the threshold value (T) automatically based on the pixel value of image. The threshold obtained is used to distinguish disease stain and leaf

background. Modifications Otsu method is described as follows [18].

First, calculate the i-th gray level value called y1 and y2 based on Red, Green and Blue (RGB) in each pixel of the image. The formula of y1 and y2 are defined as in equation (1) and (2).

$$y1 = 2g - r - b \quad (1)$$
$$y2 = 2r - g - b \quad (2)$$

Where r, g, and b is the value of Red, Green, and Blue of each pixel.

Second, enter the value of y1 and y2 alternately in Otsu as the value of gray level i (ni). This step result two threshold values T1 and T2.

Third, modify the function of the segmentation process as shown at equation (3).

$$g(x,y) = \begin{cases} pr, y1 < T1 + C \ dan \ y2 > T2 - C \\ (0,0,0), \ the \ other \end{cases} \quad (3)$$

Where pr is the value of Red, Green, and Blue (RGB) at each pixel of original image and C is threshold values variable of T1 and T2.

Color Feature Extraction

Karat and Downy mildew disease can be distinguished by it's color. Therefore, the color feature in RGB value is used as input of recognition process. In this study, we use nine color features as a characteristic of the image which are the average value of red, the average of green, the average blue, the minimum of red, the minimum of green, the minimum of blue, the maximum of red, the maximum of green, and the maximum of blue.

Learning Vector Quantization (LVQ)

Learning Vector Quantization is a method of learning for some supervised competitive layer. Automatically, the competitive layer learns to classify a given input vector. Some of the input vectors will be grouped in the same class, if both vectors have a very close distance. The LVQ architecture used has two layers namely input layer and output layer as shown in Figure 3.

The number of nodes in the input layer is nine in accordance with the number of features resulted from color feature extraction. The number of nodes in the output layer is one in accordance with a specified number of diseases

that are downy meldew and karat. Steps of LVQ learning algorithm uses the algorithm proposed by [19].

1. Initialize the variable of learning such as reference vector and learning rate (α).
2. Repeat the step 3 to step 6 since learning rate is not less than the minimum learning rate.
3. For each training vector x, do Step 3 to Step 4.
4. Select class of j (called Cj), where euclidean distance between the input vector and the weight vector for output j (‖x-wj‖) is the minimum.
5. Update weights of the j-th class.(wj) as follows:

   If T = Cj, then
   $$w_{j(new)} = w_{j(old)} + (\propto * w_{j(old)}) \quad (4)$$
   If T ≠ Cj, then
   $$\boldsymbol{w_{j(new)}} = \boldsymbol{w_{j(old)}} - (\propto * \boldsymbol{w_{j(old)}}) \quad (5)$$

   with T is categories or classes for training vectors

6. Reduce learning rate using certain deduction value as folows:

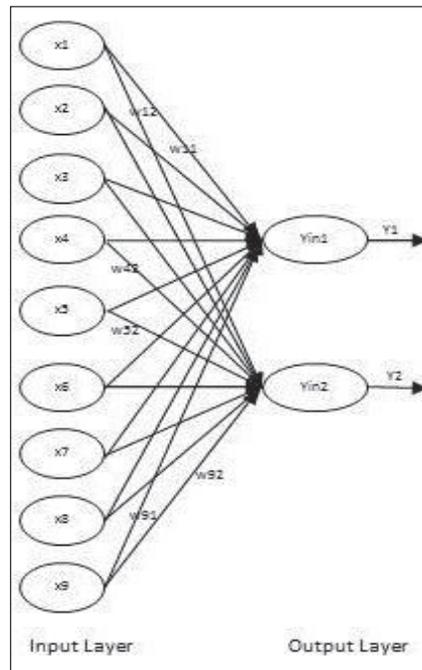$$\propto = \propto * deduction \quad (6)$$

Figure. 3 The LVQ architecture of study

RESULT AND DISCUSSION

As mentioned above, this study developed application for the identification of disease on soybean leaves. Designed application consists of several parts that can be used to perform some tests such as test of diseases, test of Otsu threshold value and test of LVQ parameters. The application also displays the results of image processing at each stage. The example of user interface of the application is shown in Figure 4.

To obtain the best performace of the proposed algorithm, some testing in the form of the selection the threshold value variable (C) on Otsu and the LVQ parameters value is important. As shown in equation (3) in which the use of different threshold values will produce different segmentations area. In addition to a threshold value, there is a variable on LVQ algorithms such as learning rate and decrease of learning rate also needs to be adjusted to obtain optimum accuracy. Thus, the choice of the best value of these three variables is necessary.
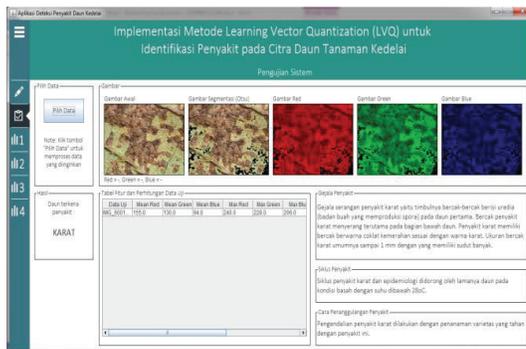

Figure 4. The user interface of application

In this study, testing of the these three variables is done sequentially. The best value resulted from the first variable test will be used for the second and so on. The number of dataset used for testing as much as 60 and the test results are described as follows.

The Result of Threshold Value Testing

The testing was done on some threshold value variable (C) from 10 to 90. This value is chosen after conducting several experiments with several other ranges of value. Whereas, the value of the LVQ parameters on this test used are learning rate value of 0.1 and a learning rate deduction of 0.1.

The best value of C was choosen based on the highest value of accuracy that is calculated during testing process. Table 1 show that the accuracy of test results on all threshold value is quite high. The highest accuracy is 93.33% and obtained when the threshold value is 60, 80 and 90. Therefore, these threshold values then are recommended as optimum value of C for further testing.

Table 1. The result of testing of threshold value variable

Threshold Value Variable (C)	Accuracy
10	90%
20	90%
30	90%
40	90%
50	90%
60	93.33%
70	90%
80	93.33%
90	93.33%

The Result of Learning Rate Testing

As seen in equation 4 and equation 5, the learning rate value is a very decisive variable to get the optimal weight value. Therefore, the selection of appropriate learning values need to be done so that the weight gained is optimal.

Learning rate testing was done by using optimum threshold value generated from previous testing, which is 60, 80 and 90. This test is performed at some learning rate value from 0.1 to 0.9. Table 2 shows the result of this test.

It shown that the average accuracy resulted is 93.33% at learning rate 0.1 to 0.6. But, the accuracy fell sharply on the value of learning rate above 70, except at learning rate 0.8 and threshold value 60. Although the highest accuracy obtained is 96.67%, but this value appears only once in 30 trials. The result shows the stable value of accuracy is 93.33%, so these value is used as a reference in the selection of the optimal learning rate. It can be decided that the value of learning rate being recommended for the next test is between 0.1 and 0.6.

Table 2. The result of testing of learning rate

Learning Rate	Accuracy		
	C=60	C=80	C=90
0.1	93.33%	93.33%	93.33%
0.2	93.33%	93.33%	93.33%
0.3	93.33%	93.33%	93.33%
0.4	93.33%	93.33%	93.33%
0.5	93.33%	93.33%	93.33%
0.6	93.33%	93.33%	93.33%
0.7	50%	50%	50%
0.8	96.67%	50%	50%
0.9	50%	50%	50%

The Result of Learning Rate Deduction Testing

The third variable must be found to get the best performance of system is deduction rate. This value affects the change of the learning rate value in each of iteration.

This test is performed at some learning rate deduction value from 0.1 to 0.9. The other parameter such as threshold value and learning rate use the best value resulted from previous testing that are 0.6, 0.8, 0.9 for threshold value and 0.1 for learning rate. The result at Table 3 shows the best learning rate deduction is 0.1, which has highest accuracy about 93.33%.

Table 3. The result of testing of learning rate deduction

Learning Rate Deduction	Accuracy		
	C=60	C=80	C=90
0.1	93.33%	93.33%	93.33%
0.2	90%	90%	93.33%
0.3	90%	90%	90%
0.4	90%	90%	90%
0.5	90%	90%	90%
0.6	86.67%	86.67%	86.67%
0.7	86.67%	86.67%	86.67%
0.8	86.67%	86.67%	86.67%
0.9	86.67%	86.67%	86.67%
Average	**90.00%**	**90.00%**	**90.00%**

Table 4. The result of testing of the number of data

Number of Data	Accuracy		
	C=60	C=80	C=90
10	83.33%	75.67%	85%
20	89.67%	88.67%	85.67%
30	90%	89.33%	89.67%
40	90.33%	89%	93.33%
50	90%	93.33%	95%
60	92.33%	93.33%	93.33%
Averege	**90%**	**89%**	**90%**

Overall, the application can identify the two diaseses with high accuracy which is between 86.67% and 93.33%. This indicates that the methods offered to help in the identification of disease on soybean plants. The result on Table 3 shows the optimum accuracy is 90%, while it has more consistent value on some testing variable of Otsu and LVQ. This means that the network or the system will give more accurate result while it further tested by using this parameter value.

To determine the effectiveness of these three values of the variableis, then tested by using a different combination of data. The result of this test are shown in Table 4. The table shows the average accuracy is between 89% - 90% and the best accuracy is 95%. Futhermore, the number of data that give better result is 50 and the value of C is 90.

CONCLUSION

This study build the application to identify the disease on leaves soybean through leaves image. To detect the availability of disease on leave image is used modified Otsu method, which is consists of two threshold to make a segment between a disease and leaves that has nearly same color. The color features are then used as input in the process of introduction of the disease with LVQ algorithm. Based on the testing has been done, we found the highest accuracy is 95%, but the more stable accuracy is 90%. According to this accuracy, we conclude this method is quite appropriate to use as alternative on identifying a disease on leave soybean image.

This study only uses the minimum, maximum and average values of the input RGB image values as a feature. The further research it is necessary to consider the use other color features such as color moments to obtain more optimal features as input of the identification process. Besides that it is necessary to perform the identification with other algorithms in the Neural Network such as Backpropagation and Extreme Learning Method to find out the most effective method.

REFERENCES

[1] "Badan Pusat Statistik." [Online]. Available:https://www.bps.go.id/tnmn_pgn.php.

[2] K. Pertanian, "Kinerja produksi dan harga kedelai serta implikasinya untuk perumusan kebijakan percepatan pencapaian target sukses Kementerian Pertanian [Internet]," *Jkt. ID Kementeri. Pertan.*, 2012.

[3] N. Saleh and S. Hardaningsih, "Pengendalian penyakit terpadu pada tanaman kedelai," *Kedelai Teknol. Dan Pengemb. Puslitbangtan*, pp. 319–344, 2007.

[4] S. H. Marwoto and A. Taufiq, "Hama, Penyakit, dan Masalah Hara pada Tanaman Kedelai: Identifikasi dan Pengendaliannya," *Dep. Pertan. Badan Penelit. Dan Pengemb. Pertan. Pus. Penelit. Dan Pengemb. Tanam. Pangan*, vol. 68, 2006.

[5] E. Alegre, M. Biehl, N. Petkov, and L. Sánchez, "Automatic classification of the acrosome status of boar spermatozoa using digital image processing and LVQ," *Comput. Biol. Med.*, vol. 38, no. 4, pp. 461–468, 2008.

[6] J. Liu, B. Zuo, X. Zeng, P. Vroman, and B. Rabenasolo, "Nonwoven uniformity identification using wavelet texture analysis and LVQ neural network," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2241–2246, 2010.

[7] M. Blachnik and W. Duch, "LVQ algorithm with instance weighting for generation of prototype-based rules," *Neural Netw.*, vol. 24, no. 8, pp. 824–830, 2011.

[8] A. Yousef and N. M. Charkari, "A novel method based on new adaptive LVQ neural network for predicting protein–protein interactions from protein sequences," *J. Theor. Biol.*, vol. 336, pp. 231–239, 2013.

[9] B. Biswal, M. Biswal, S. Hasan, and P. K. Dash, "Nonstationary power signal time series data classification using LVQ classifier," *Appl. Soft Comput.*, vol. 18, pp. 158–166, 2014.

[10] P. Melin, J. Amezcua, F. Valdez, and O. Castillo, "A new neural network model based on the LVQ algorithm for multi-class classification of arrhythmias," *Inf. Sci.*, vol. 279, pp. 483–497, 2014.

[11] B. Mokbel, B. Paassen, F.-M. Schleif, and B. Hammer, "Metric learning for sequences in relational LVQ," *Neurocomputing*, vol. 169, pp. 306–322, 2015.

[12] S. N. Desylvia, "Perbandingan SOM dan LVQ pada identifikasi citra wajah dengan wavelet sebagai ekstraksi ciri [skripsi]," *Bogor ID Inst. Pertan. Bogor*, 2013.

[13] A. Nurkhozin, M. I. Irawan, and I. Mukhlas, "Komparasi hasil klasifikasi penyakit diabetes mellitus menggunakan jaringan syaraf tiruan backpropagation dan learning vector quantization," in *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA*, 2011, pp. M33–M40.

[14] S. M. E. Harb, N. A. M. Isa, and S. A. Salamah, "Improved image magnification algorithm based on Otsu thresholding,"

Comput. Electr. Eng., vol. 46, pp. 338–355, 2015.

[15] M. A. Guerroudji and Z. Ameur, "A new approach for the detection of mammary calcifications by using the white Top-Hat transform and thresholding of Otsu," Opt.-Int. J. Light Electron Opt., vol. 127, no. 3, pp. 1251–1259, 2016.

[16] A. M. A. Talab, Z. Huang, F. Xi, and L. HaiMing, "Detection crack in image using Otsu method and multiple filtering in image processing techniques," Opt.-Int. J. Light Electron Opt., vol. 127, no. 3, pp. 1030–1033, 2016.

[17] W. M. Nabella and J. Sampurno, "Analisis Citra Sinar-X Tulang Tangan Menggunakan Metode Thresholding Otsu Untuk Identifikasi Osteoporosis," POSITRON, vol. 3, no. 1.

[18] 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, 2009, pp. 79–83.

[19] S. Kusumadewi, "Artificial intelligence (teknik dan aplikasinya)," Yogyak. Graha Ilmu, 2003.