Classification of Chili Leaf Disease Using the Gray Level Co-occurrence Matrix (GLCM) and the Support Vector Machine (SVM) Methods

by Yuslena Sari

Submission date: 13-Feb-2022 11:11AM (UTC+0700)
Submission ID: 1761009635
File name: rence_Matrix_GLCM_and_the_Support_Vector_Machine_SVM_Methods.pdf (868.43K)
Word count: 2548
Character count: 13343

Classification of Chili Leaf Disease Using the Gray Level Co-occurrence Matrix (GLCM) and the Support Vector Machine (SVM) Methods

Yuslena Sari
Department of Information Technology
Universitas Lambung Mangkurat
Banjarmasin, Indonesia
yuzlena@ulm.ac.id

Andreyan Rizky Baskara
Department of Information Technology
Universitas Lambung Mangkurat
Banjarmasin, Indonesia
andreyan.baskara@ulm.ac.id

Rika Wahyuni
Department of Information Technology
Universitas Lambung Mangkurat
Banjarmasin, Indonesia
ikha.rikawhy@gmail.com

2021 Sixth International Conference on Informatics and Computing (ICIC) | 978-1-6654-2155-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIC54025.2021.9632920

***Abstract*— Chili is a type of vegetable that has a very high economic value. The problem that often occurs in chili plants is that many agricultural losses are caused by disease. Plant diseases are always considered a very serious problem in all countries because economic growth is largely dependent on the agricultural sector in developing countries. In some plant, diseases sometimes caused by bacteria, viruses and fungi. To anticipate this problem, a method designed into a classification system for diagnosing chili leaf disease by applying the Gray Level Cooccurrence Matrix (GLCM) feature extraction method. Then classified using the Support Vector Machine (SVM) method. The output classification of disease diagnoses in chili obtained an overall accuracy level of 88%. The results obtained prove that the method of extracting the features of Gray Level Co-occurrence Matrix (GLCM) and Support Vector Machine (SVM) can be applied to diagnosing chili plants disease.**

Keywords*— *Chili Plant, Classification, Digital Image Processing, Gray Level Co-occurrence Matrix (GLCM), Support Vector Machine (SVM)

I. Introduction

Chili is a plant and fruit that can be classified as a vegetable, and has been bred since 3000 BC and is used as a spicy seasoning. Chili plant is a type of vegetable that has a very high economic value [1]. Chili is included in Horticultural crops based on garden cultivation.

Diseases of plants have always been considered a very serious problem in all countries because economic growth in most developing countries is highly dependent on agriculture. In some plant, diseases sometimes caused by bacteria, viruses and fungi. Efforts to anticipate this problem has been conducted such as by doing researches in the field of agriculture. One of the researches conducted is the development of a systematic system for the detection or diagnosis of plant disease [2].

The diagnosis of chili disease must be carried out as quickly and accurately as possible, because it can quickly attack and spread if not treated as soon as possible. Based on data from interviews with the Head of Agricultural Extension at the Government Office of Food, Crops and Horticulture Banjar District in Indonesia, almost every planting season problem with chili plants often occur in Tambak Anyar area in Banjar District and the most severe problems occurred in 2018, which resulted in considerable losses. The disease identified on chili often attack its leaves. There are several types of diseases in chili leaves such as:

1. Fusarium wilt.
2. Leaf Curl.
3. Leaf Spots
4. Ralstonia Bacterial Wilt
5. Yellow Virus.

Previous research has raised cases of computer vision-based plant diseases identification, entitled "Classification of Potato Leaf Diseases Based on Texture Features and Color Features Using a Support Vector Machine" whose purpose is to diagnose potato plant diseases using two methods, namely Gray Level Co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) [3]. Research that is currently developing on the identification of chili plants is carried out in the field of remote sensing, while there is no identification of chili diseases using leaf imagery that analyzes diseases on chili leaves [4]. therefore, this study discusses features that are suitable and are expected to be suitable for diagnosing diseases in chili leaves.

The first stage in this research is to collect image data of plant leaves affected by the disease in a field survey. The next step is image pre-processing so that it is processed at the feature extraction stage. The image taken by the camera is in RGB format [5]. This study uses two methods to extract objects on the image of chili plants. GLCM is used to extract features from an image, to classify images into vector engines. GLCM determines the probability of occurrence of gray combinations at a certain distance. Then the characteristics of the chili plants are classified using the SVM method so that it can be known what diseases exist in chili plants.

II. Related Work

Some of previous researchers have conducted a study about identification of plant diseases using digital image processing. In [3], They identify diseases in a potato leaf. The methods used are Gray Level Co-Occurrence Matrix (GLCM), Color moment and Support Vector Machine (SVM). The results obtained are 87% for the average accuracy of 3 types of diseases on potato plant leaves.

Another research about identification of plant disease had been conducted in [11]. They identify diseases in sugarcane plant leaves. The method used are Gray Level Co-Occurrence Matrix, Color moment and Support Vector Machine. The difference with previous research is they used four combination of GLCM features while the previous one used seven combination of features. The result obtained was

slightly better in term of accuracy which is 90.33%. In [14], a research was conducted to identify a disease in rice leaf. The method used are Gray Level Co-Occurrence Matrix and Backpropagation. Five combination of GLCM features also used in this research namely, Energy, Entropy, Contrast, Homogeneity, and Correlation. The result obtained was 80% in term of accuracy.

From the mentioned related works above, it showed that finding the best combination of feature in GLCM is still a problem that need to be solved. Our study proposed to use a six combination of GLCM features namely, Contrast, Energy, Entropy, Dissimilarity, Inverse Different Moment, and Correlation and then observe how those features affect the accuracy on identifying disease on plant leaves, especially on chili plant leaves. Support Vector Machine is used to classify the five chili plants disease which are trying to be identified.

III. Proposed Method

This study proposes a method with data processing steps, feature extraction, training and testing in order to obtain research results. The flow of the method proposed in the study is as shown in Fig. 1.

A. Data Acquisition

In this step, images of chili leaf are taken from chili plant in Tambak Anyar area, Banjar District. The images taken using smartphone camera with minimal specification of 8 Megapixel. The image data taken is an image of disease in chili plants consisting of Fusarium Wilt, Leaf Curl, Leaf Spots, Ralstonia Bacterial Wilt, and Yellow Virus. The image data taken only the chili leaves [6].

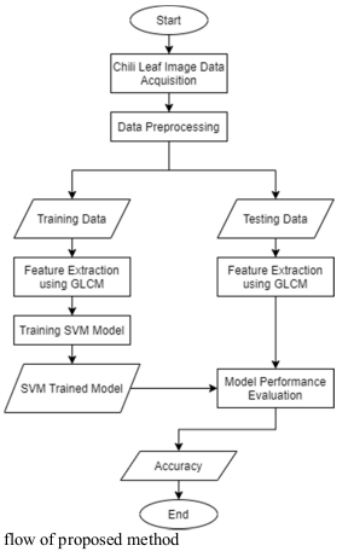


Fig. 1. The flow of proposed method

B. Data Preprocessing

From the image that has been obtained, it will go through the data processing stage. This stage will be carried out using the Matlab application. This data processing will be divided into two parts, namely processing the training data first and then processing the test data [7]. Image data is divided according to the category of the type of disease that occurs. Then change the pixel size (resize) on all chili leaf image data, in this process the original image is resized to 256x256 pixels.

The sample image of Fusarium Wilt, Leaf Spots, Leaf Curl, Ralstonia Bacterial Wilt, and Yellow Virus can be seen on Fig. 2, Fig. 3, Fig. 4, Fig. 5 and Fig. 6 respectively.

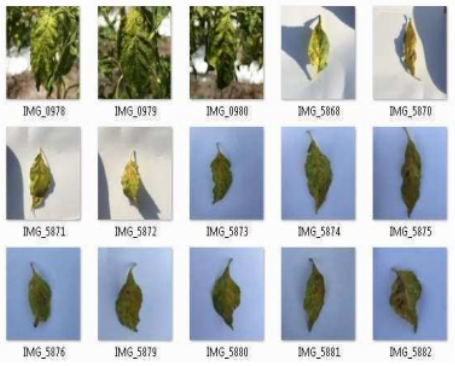


Fig. 2. Sample image data of Fusarium Wilt disease

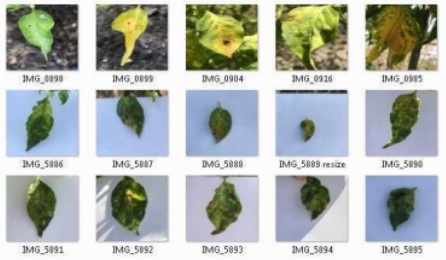


Fig. 3. Sample image data of Leaf Spot Disease

Fig. 4. Sample image data of leaf curl disease

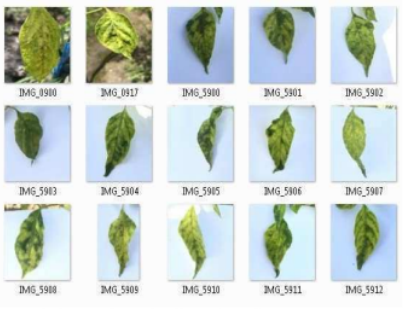


Fig. 5. Sample image data of yellow virus disease

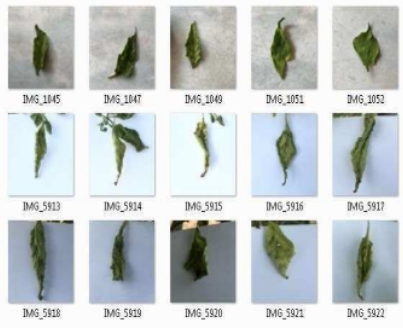


Fig. 6. Sample image data of Ralstonia Bacterial Wilt disease

C. Feature Extraction Using Gray Level Co-Occurrence Matrix (GLCM)

GLCM is a method of analyzing pixels in an image using texture calculations to determine the level of gray that occurs in the second order [8]. The image analysis can be done by extracting the texture features based on the statistical distribution of the pixel intensity. Characteristics obtained from the pixel matrix value from the GLCM method form an orientation pattern angle [9]. There are 6 features used, namely:

1. Contrast, is the result of measuring the intensity contrast between pixels and their neighbors in the entire image. Contrast value can be obtained using Eq. (1).

$$\boldsymbol{Contrast} = \sum_{n=1}^{L} n^2 \left\{ \sum_{|i-j|=n} GLCM(i,j) \right\} \quad (1)$$

2. Energy, the higher the similarity, the higher the Energy value. Energy value can be obtained using Eq. (2).

$$\boldsymbol{Energy} = \sum_{i,j=0}^{N-1} \left(P_{ij}\right)^2 \quad (2)$$

3. Entropy, when the image is not uniform in texture, it has a very small value, which implies that the entropy is very large. Therefore, entropy is inversely proportional to energy. Entropy value can be obtained using Eq. (3).

$$\boldsymbol{Entropy} = \sum_i \sum_j P(i,j) \log(P(i,j)) \quad (3)$$

4. Dissimilarity, Calculating the dissimilarity of a texture, it will be resulting in high value if random and low value if similar. This value can be obtained using Eq. (4).

$$\boldsymbol{Dissimilarity} = \sum_{i,j=0}^{N-1} P_{ij} |i-j| \quad (4)$$

5. Inverse Different Moment (IDM), measure the homogeneity of the image. This parameter reaches the greatest value when the gray levels are similar. This value can be obtained using Eq. (5).

$$\boldsymbol{IDM} = \sum_i \sum_j \frac{1}{1+(i-j)^2} P(i,j) \quad (5)$$

6. Correlation is a measurement of how pixels correlate with their neighbors in the image. Correlation value can be obtained using Eq. (6).

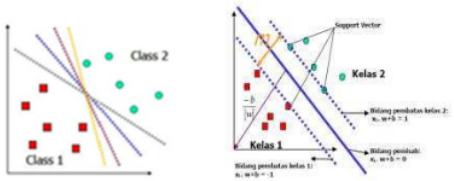
$$\boldsymbol{Correlation} = \sum_{i,j=0}^{N-1} P_{ij} \left| \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right| \quad (6)$$

D. Generate Model using Support Vector Machine

SVM is a learning algorithm that analyzes data for classification and regression analysis. This statistical learning theory is based on the optimization of the implementation of learning bias. SVM is a machine learning method based on Vladmir Vapnik's statistical theory [10].

The SVM method is a method used to classify data and select variables. SVM can perform a generalization process to determine certain patterns based on test data and training data. SVM has a simple structure and can solve existing problems easily [11].

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen [12]. The objective



(a) (b)

Fig. 7. (a) illustration of alternative hyperplane (b) illustration of a hyperplane with the best margin separate two data class

is to find a plane that has the maximum margin, i.e., the maximum distance between data points of both classes.

Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence. The illustration of hyperplane in SVM is shown in Fig. 7.

E. Model Performance Evaluation

After the model of Support Vector Machine (SVM) method generated using training data, then performance evaluation step carried out. Accuracy used as metric to measure the performance of model [13]. Equation (7) used to calculate accuracy of the model.

$$Accuracy = \frac{rightly\ classified\ data}{Total\ data} \times 100\% \quad (7)$$

IV. Result and Discussion

From all testing results, the classification obtained for the accuracy testing process with each chili plant disease

diagnosis is detailed in Table I. The image data used for testing is 25 image data, 5 image data for each chili leaf disease classified in this study.

TABLE I. TESTING RESULT

	Classification Result					
Real Data		**A**	**B**	**C**	**D**	**E**
	A	5	0	0	0	0
	B	0	3	0	1	1
	C	0	0	4	0	1
	D	0	0	0	5	0
	E	0	0	0	0	5

Where,
A = Fusarium Wilt Disease
B = Leaf Spot Disease
C = Leaf Curl Disease
D = Yellow Virus Disease
E = Ralstonia Bacterial Wilt Disease

Next, Accuracy value can be calculated based on classification result from testing the proposed method using Eq. (7).

$$Accuracy = \frac{22}{25} \times 100\% = 88\ \%$$

So from the above calculation, the accuracy of the trained model for classifying chili leaf diseases is 88%. Some misclassification occurred because some leaves image taken on a bright light and it affected the result of GLCM such as Leaf spot disease mistakenly classified as Yellow Virus Disease. It can be seen on the sample data that leaf spot disease have a bit of yellow color similar to yellow virus disease.

V. CONCLUSION

This study facilitates the diagnosis of diseases in chili plants with the following stages: feature extraction using Gray Level Co-Occurrence Matrix (GLCM) and classification using Support Vector Machine (SVM). The performance of proposed method evaluated using accuracy metric and obtained value of 88%. From the method that has been proposed, it can simplify the process of classifying the diagnosis of chili leaf disease using images and from the results of the classification system, it can later be used as supporting information to overcome the disease.

REFERENCES

[1] F. Roziq, I. R. Sastrahidayat, S. Djauhari, and U. B. M. Program Studi Agroekoteknologi, Jurusan Hama dan Penyakit Tumbuhan, "Kejadian Hama Dan Penyakit Tanaman Cabai Kecil Yang Dibudidayakan Secara Vertikultur Di Sidoarjo," *J. HPT*, vol. 1, pp. 30–36, 2013.

[2] P. K. Sethy, B. Negi, S. K. Behera, N. K. Barpanda, and A. K. Rath, "An Image Processing Approach for Detection, Quantification, and Identification of Plant Leaf Diseases -A Review," *Int. J. Eng. Technol.*, vol. 9, no. 2, pp. 635–648, 2017.

[3] P. U. Rakhmawati, Y. M. Pranoto, and E. Setyati, "Klasifikasi Penyakit Daun Kentang Berdasarkan Fitur Tekstur Dan Fitur Warna Menggunakan Support Vector Machine," pp. 1–8, 2018.

[4] S. Saifudin and A. Fadlil, "Sistem Identifikasi Citra Kayu Berdasarkan Tekstur Menggunakan Gray Level Coocurrence Matrix (Glcm) Dengan Klasifikasi Jarak Euclidean," *Sinergi*, vol. 19, no. 3, p. 181, 2015.

[5] R. Munir, *Pengolahan Citra Digital*. Bandung: Informatika Bandung, 2002.

[6] A. Andoko, Budidaya Cabai Merah Secara Verkultur Organik. 2004.

[7] A. Sindar and R. M. Sinaga, "Implementasi Teknik Thresholding Pada Segmentasi Citra Digital," vol. 1, no. 2, pp. 48–51, 2017.

[8] S. R. Andhika, F. Abdul, and Y. Anton, "Ekstraksi Ciri Metode Gray Level Co-Occurrence Matrix (GLCM) dan Filter Gabor untuk Klasifikasi citra Batik Pekalongan," *J. Inform. J. Pengemb. IT*, vol. 2, no. 2, pp. 23–26, 2017.

[9] H. B. Kartal and F. Cebi, "Support Vector Machines for Multi-Attribute ABC Analysis," *Int. J. Mach. Learn. Comput.*, vol. 3, no. 1, pp. 154–157, 2013.

[10] N. A. Fourina and P. W. Santi, "Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi," *J. Sains dan Seni ITS*, vol. 1, no. 1, 2012.

[11] R. K. Dewi and R. V. H. Ginardi, "Identifikasi Penyakit pada Daun Tebu dengan Gray Level CoOccurrence Matrix dan Color Moments," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 2, p. 70, 2014.

[12] F. Y. Manik, "Ciri Morpologi Menggunakan Support Vector Machine (SVM)," 2015.

[13] R. P. Putra and O. Setyawati, "Klasifikasi Penyakit Tanaman Kedelai Melalui Tekstur Daun dengan Metode Gabor Filter," vol. 12, no. 1, pp. 40–46, 2018.

[14] J. Kusanti and N. A. Haris, "Klasifikasi Penyakit Daun Padi Berdasarkan Hasil Ekstraksi Fitur GLCM Interval 4 sudut," vol. 03, no. 1, pp. 1–6, 2018.

Classification of Chili Leaf Disease Using the Gray Level Co-occurrence Matrix (GLCM) and the Support Vector Machine (SVM) Methods

ORIGINALITY REPORT

18%	15%	15%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

4%

★ 1library.net

Internet Source

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		