Classification of Tomato Plant Diseases through Leaf using Gray-Level Co-Occurrence Matrix and Color Moment with Convolutional Neural Network Methods

Anton[1][1-2] Supriadi Rustad [2][1] Guruh Fajar Shidik [3][1] Abdul Syukur [4][1]

[1] Universitas Dian Nuswantoro, Semarang
[2] STMIK Nusa Mandiri, Jakarta

Abstract. Tomato is a fruit vegetable source of vitamins and minerals, in addition to being consumed as fresh fruit it can also be processed into food industry raw materials such as fruit juices and sauces. However, due to various causes such as diseases, pest attacks, and unstable weather conditions cause a decrease in the quality and quantity of production. In order to contribute to maintaining the productivity of tomato plants, the use of technology can be an alternative to be applied to the cultivation of tomato plants. This study applies image processing techniques to detect the texture of affected leaf using Gray-level Co-occurence Matrix (GLCM) extraction and Color Moment using Convolutional Neural Network (CNN) method. Among the diseases that often occur in tomato leaf are late blight, Septoria spot, bacterial spot, Target Spot, Early blight, leaf curl, Spider mites Two spotted spider mites, and Leaf Mold. In this study a combination of GLCM-Color Moment and CNN method was chosen because of its reliability in identifying and classifying plant diseases compared to only using CNN. In this study, using a data set from Plant Village totaling 16.012 images.

Keywords: Tomato Plant diseases, Color Moment, GLCM, CNN, deep learning

1 Introduction

Tomato is a fruit vegetable that is very potential to be cultivated because the fruit is a source of vitamins and minerals. Besides being consumed as fresh fruit and for cooking spices, it can also be further processed as raw materials for food industries such as fruit juices and sauces [1]. Tomatoes become one of the horticultural commodities that have high economic value and still require serious handling, especially in terms of increasing yields and fruit quality [2]. Plant diseases, pest attacks, and unstable weather conditions can cause a decrease in the quality and quantity of production. Diseases that attack tomato plants can be caused by fungi, bacteria and viruses. The symptoms can be seen from changes in the shape and color of leaves. Most farmers do is recognize the symptoms with the naked eye and take immediate action without knowing how to overcome them. Therefore, we need the help of image processing technology that can recognize diseases in tomato plants. Some types of diseases found in tomatoes include yellow leaf curl, potato virus y, tomato mosaic virus, and leaf spot [3].

Developing research on the identification of tomato plant diseases before [4], the number of classes used was six classes, including mushrooms: 2, bacteria: 2, viruses: 1 and healthy (normal) 1. Therefore, this research class discusses the right features, so it is expected to be suitable for disease identification in tomato leaf. Research on the identification of plant diseases in terms of leaf digital images [5] suggests the use of shape, texture and color features. This study did not use the shape features because the form of rust disease is not patterned, other than that the shape features are more suitable for classifying of plant species with different leaf patterns [6] and identification of plant diseases with symptoms of hollow or deformed leaves at the edge of the leaf, for example in disease mosaic on cassava leaves [7]. In addition to leaf disease, rain and temperature factors also affect the quality and quantity of production. Therefore, the management of image pattern recognition to detect disease in leaves is a major problem in agriculture [8].

2 Related Work

Deep Learning has made remarkable progress in identifying images in large numbers. Several previous studies used one of the Deep Learning architectures namely Convolutional Neural Network (CNN) [9], [10], [11]. Deep learning uses an approach that can learn many layers of features and automatically makes representations of input data. Other studies have modified CNN using feature extraction to improve the performance of CNN [12].

Table 1. Comparison between all classifications

Classification Problem	Method	Accuracy (%)
Leaf recognition for plant	GLCM & PCA	GLCM (78%); PCA (98.46%)
Classification of maize leaf disease	CNN (GoogLeNet)	97.89%
Recognition of Maize Leaf Diseases	CNN	92,85%
Identification of Corn leaf disease	CNN (GoogLeNet & Cifar10)	GoogLeNet (98.9%) Cifar10 (98.8%)
Cow Race Classification	GLCM-CNN	93.763%
classification of a variety of plant species	CNN	97.47%
Tomato Plant Leaf Disease Detection	CNN	91.67 %

3 Method

There are texture differences between normal, diseased and mosaic leaves [9]. This study chose the gray-level co-occurrence matrix to extract textures because GLCM as an excellent method is used to determine the texture of the leaves and detect plant types and have high accuracy [5], the complexity of image textures is difficult to define and quantify, but GLCM can be used to quantify and compare various aspects of image texture [10]. The use of color moments for color because there are differences in color on healthy and disease leaf [5] and [9]. Color Moment is an effective image extraction method for analyzing images with color due to the presence of vector dimensions with the lowest computational features and complexity compared to other methods such as color correlation, histogram and color structure descriptors [8]. An innovative solution is made to maintain the quality of tomato production by utilizing automatic motor technology with a camera that can capture the four sides of the plant to detect and recognize leaf disease. Using 4.923 images of diseased and healthy plant leaves, these image data are trained with in a convolutional neural network a deep F-RCNN anomaly detection model to identify whether or not the disease exists. This research is expected to facilitate the identification of tomato plant diseases quickly [11]. Research on the analysis of plant diseases through leaf images using SVM for classiflying of features including linear, quadratic, gaussian, and cubic kernels produces an accuracy of 98.3% [12].

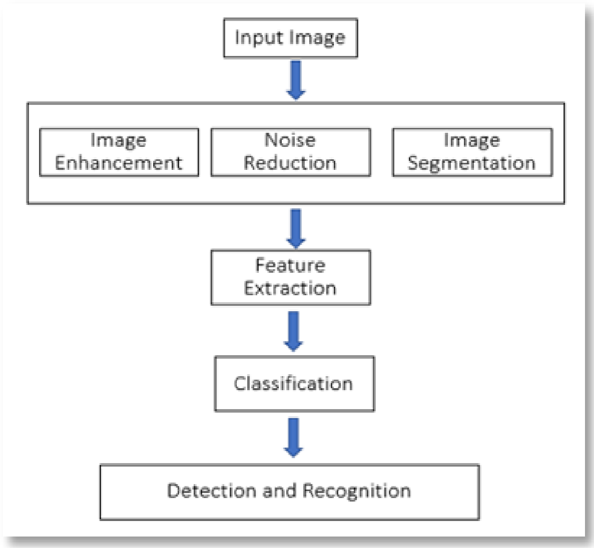

Fig 1. Proposed framework for the detection and classification of plant leaf diseases

4 Result and Discussion

4.1 Result

This research uses computer specifications with Intel (R) Core (TM) i5-8250U @ 1.60 GHz 1.80 GHz Processor, 8.00 GB Memory, 256 GB SSD and HDD for 1 TB data and Matlab. The dataset used for this study was sourced from plantvillage (https://plantvillage.org), divided into training data and test data used to solve small sample size problems in modeling. Training data from ten classes of tomato leaf is

70% and test data is 30%. Vector features a combined GLCM Texture feature called Contrast, Correlation, Energy and homogeneity. The amount of training data is 11.208 images and 4.804 images of test data.

Table 2. Data on tomato plant leaf

Leaf Type	**Total leaf**
T. Spot	1.404
Mosaic	373
YLCV	3.209
B. Spot	2.127
Early B	1.000
Late B	1.909
Leaf Mold	952
SL. Spot	1.771
SMT. Spotted	1.676
Healty	1.591
Total all leaf	16.012

4.1.1 Gray-Level Co-Occurrence Matrix (GLCM)

The improvement of the proposed GLCM-Color moment-CNN method uses three approaches, namely: Comparing GLCM-Color moment-CNN with previous studies. Second, use ten classes to show increased accuracy. Third, reduce the background of the image to be imported and speed up the process of extracting the texture features of the image in computing. Feature extraction used, namely:

$$\text{Contrast}=\sum |\text{i-j}|\ p(i,j)^2 \qquad (1)$$

$$\text{Correlation}=\sum \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma i \sigma j} \qquad (2)$$

$$\text{Energy}=\sum p(i,j)^2 \qquad (3)$$

$$\text{Homogenity}=\sum \frac{p(i,j)}{1+|i-j|} \qquad (4)$$

Some GLCM Extracted textural features are illustrated in Table 3 and Table 4 for two different leaf images.

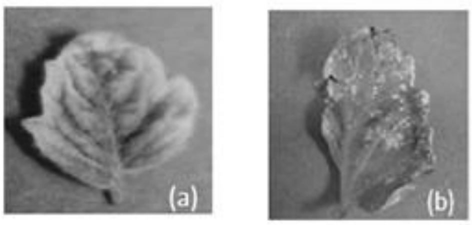

Fig 2. Images textures extracted from leaf images (a) (b)

Table 3. Some image textures extracted from leaf images (a)

Angle	Contrast	Correlation	Energy	Homogeneity
0^0	0.10669	0.95639	0.24762	0.94733
45^0	0.14825	0.93958	0.23102	0.92801
90^0	0.11157	0.95441	0.24484	0.94491
135^0	0.14885	0.93931	0.23178	0.92824
Average	0.12884	0.94742	0.23882	0.93712

Table 4. Some image textures extracted from leaf images (b)

Angle	Contrast	Correlation	Energy	Homogeneity
0^0	0.23151	0.82838	0.24639	0.8987
45^0	0.29727	0.78	0.23092	0.87992
90^0	0.23148	0.82861	0.24946	0.90208
135^0	0.30544	0.77393	0.22816	0.87725
Average	0.26642	0.80273	0.23873	0.88949

At degrees 0° and 90° the GLCM method provides the same accuracy and results. Poor results are obtained at 45° because any change in neighboring degrees will change the value of the extracted leaf features. For any changes to the image using the GLCM method is very sensitive such as changes in rotation, scale, etc. This can be seen in Tables 3 and 4, differences in features extracted. The introduction of the GLCM method is very fast and the computing time for the GLCM method is less.

4.1.2 Color Moment

Color Moment is used to distinguish images based on color features and measure the similarity of colors between images. The color distribution of an image is defined as the probability distribution based on the Color Moment assumption. The color moment feature selection is based on the accuracy results per color moment feature in Table 5.

Table 5. Testing Accuracy per Color Moments Feature

Color moments (Cm)	**accuracy**
cm 1	52,33 %
cm 2	93,67 %
cm 3	67,33 %

4.1.3 Convolutional Neural Networks (CNN)

Convolutional Neural Network (CNN) is a type of neural network commonly used to analyze visual images, detect and recognize objects in images. Input data performed by CNN uses raw images (RGB images), but in the GLCM-Color Moment-CNN method, the input data is extraction from the image texture using GLCM-Color Moment.

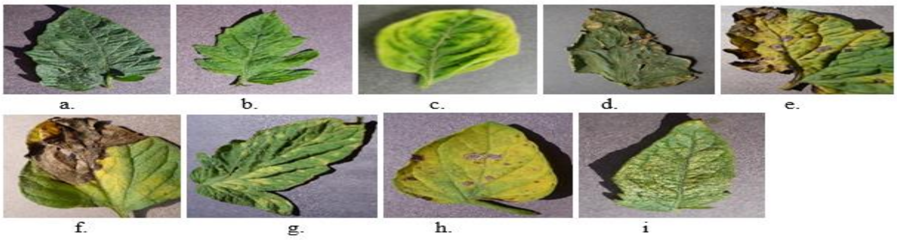

Fig 3. Example of the data used in this research; (a) Target Spot, (b) mosaic virus, (c) Yellow Leaf Curl Virus, (d) Bacterial spot, (e) Early blight, (f) Late blight, (g) Leaf Mold, (h) Septoria leaf spot, (i) Spider mites Two spotted spider mites

Grayscalling on the training data image and test data performed on MATLAB, using the function: I = rgb2gray ("image_variable")

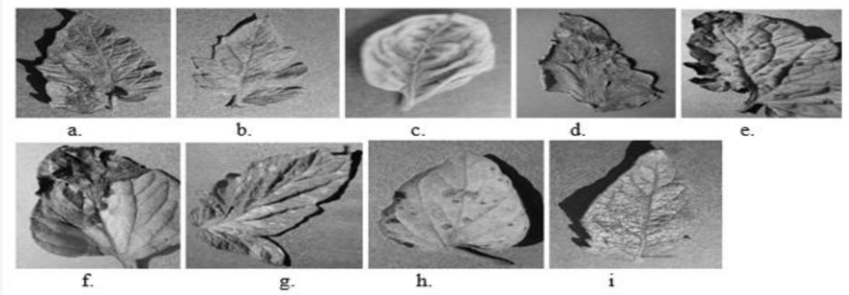

Fig 4. Grayscale image of Tomato leaf

Fig 5. Tomato leaf color segmentation

The results of this experiment can also be applied in other crops (mangoes, potatoes, rice or sugar cane) with different dataset sizes. Factor similarity in texture and type of disease from each leaf will be a challenge. Modifying the parameters of the model being trained depends on the size of the dataset used.

4.2 Discussion

Based on the strengths and weaknesses of previous studies, there is a research room related to the use of GLCM, Color Moment and CNN that can maximize the value of classification accuracy in leaf texture in tomato plant diseases. Three stages of our research were carried out to evaluate the effectiveness of the proposed method. The first stage compares previous research, the second stage we used ten classes of tomato leaf disease, because previous studies were used below. The third stage we do image enhancement, Image reduction and Image segmentation so that the extraction process on the GLCM feature does not require a long time.

4. Conclusion

The conclusion that can be obtained from this research is the image identification system of diseases of tomato leaf based on the extraction of color moment features and texture features of GLCM with angular orientations of 0˚, 45˚, 90˚ and 135˚ and pixel spacing ranging from 1 to 10 pixels with CNN classification succeeded in identifying the types of diseases found in tomato leaf. The results of the identification there are 1.404 images of Target Spot diseased images, 373 images of mosaic virus and 3.209 images of Yellow Leaf Curl Virus, 2.127 Bacterial spot images, 1.000 Early blight images, 1.909 Late blight images, 952 Leaf Mold images, 1.771 Septoria images leaf spot, 1.676 SMT Spot images and 1.591 healthy images. The GLCM-Color moment-CNN method can be implemented to detect diseases in tomato leaf plants with 99% accuracy in detecting tomato leaf plant diseases.

References

[1] C. Wasonowati, “Meningkatkan Pertumbuhan Tanaman Tomat (Lycopersicon esculentum) Ddengan Sistem Budidaya Hidroponik,” *Agrovigor*, vol. 4, no. 1, pp. 21–28, 2011.

[2] D. Homepage, “Tingkatkan konsumsi sayur dan buah nusantara menuju masyarakat hidup sehat,” *last accessed 2019/12/10.*, 2017.

[3] F. Alviansyah, I. Ruslianto, and M. Diponegoro, “Identifikasi Penyakit Pada Tanaman Tomat Berdasarkan Warna Dan Bentuk Daun Dengan Metode Naive Bayes Classifier Berbasis Web,” *J. Coding Sist. Komput. Untan*, vol. 05, no. 1, pp. 23–32, 2017.

[4] H. Sabrol and K. Satish, “Tomato Plant Disease Classification in Digital Images using Classification Tree,” pp. 1242–1246, 2016.

[5] A. Camargo and J. S. Smith, “Image pattern classification for the identification of disease causing agents in plants,” *Comput. Electron. Agric.*, vol. 66, no. 2, pp. 121–125, 2009.

[6] M. Shabanzade, M. Zahedi, and S. Amin Aghvami, “Combination of Local Descriptors and Global Features for Leaf Recognition,” *Signal Image Process. An Int. J.*, vol. 2, no. 3, pp. 23–31, 2011.

[7] J. R. Aduwo, E. Mwebaze, and J. A. Quinn, “Automated Vision-Based Diagnosis of Cassava Mosaic Disease,” *Proc. ICDM Work. Data Min. Agric.*, pp. 114–122, 2010.

[8] J. K. Patil, “Color Feature Extraction of Tomato Leaf Diseases,” *Int. J. Eng. Trends Technol.*, vol. 2, no. 2, pp. 72–74, 2011.

[9] M. M. Santoni, D. I. Sensuse, A. M. Arymurthy, and M. I. Fanany, “Cattle Race Classification Using Gray Level Co-occurrence Matrix Convolutional Neural Networks,” *Procedia Computer Science*, vol. 59. ICCSCI 2015, pp. 493–502, 2015.

[10] H. Yalcin and S. Razavi, “Plant Classification using Convolutional Neural Networks Hulya,” *Adv. Intell. Syst. Comput.*, vol. 898, pp. 357–365, 2019.

[11] R. A. Priyadharshini, “Maize leaf disease classification using deep convolutional neural networks,” *Neural Comput. Appl.*, vol. 3, 2019.

[12] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, “Identification of maize leaf diseases using improved deep convolutional neural networks,” *IEEE Access*, vol. 6, pp. 30370–30377, 2018.

[13] J. C. Comstock and R. S. Lentini, “Sugarcane Mosaic Virus Disease,” *Univ. Florida*, pp. 1–3, 2005.

[14] R. G. De Luna, E. P. Dadios, and A. A. Bandala, “Automated Image Capturing System for Deep Learning-based Tomato Plant Leaf Disease Detection and Recognition,” *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2018-Octob, no. October, pp. 1414–1419, 2018.

[15] W. Zaka-ud-din, Muhammad Aziz, Sumair Rashid, Junaid Ismail, “Classification of Disease in Tomato Plants ’ Leaf Using Image Segmentation and SVM,” *Tech. Journal, Univ. Eng. Technol.*, vol. 23, no. 2, pp. 81–88, 2018.

[16] M. A. N. Patil and M. V. Pawar, “Detection and Classification of Plant Leaf Disease,” *Iarjset*, vol. 4, no. 4, pp. 72–75, 2017.

[17] Y. Sharath Kumar, “Leaf recognition for plant classification using GLCM and PCA methods,” *Orient. J. Comput. Sci. Technol.*, vol. 3, no. 1, pp. 31–36, 2010.

[18] D.N.D.Harini and Dr.D.Lalitha Bhaskari, “Identification of Leaf Diseases in TomatoPlant Based on Wavelets and PCA,” *2011 World Congr. Inf. Commun. Technol.*, pp. 1398–1403, 2011.

[19] M. Sibiya and M. Sumbwanyambe, “A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks,” *AgriEngineering*, vol. 1, no. 1, pp. 119–131, 2019.

[20] C. E. Honeycutt and R. Plotnick, “Image analysis techniques and gray-level co-occurrence matrices (GLCM) for calculating bioturbation indices and characterizing biogenic sedimentary structures,” *Comput. Geosci.*, vol. 34, no. 11, pp. 1461–1472, 2008.