Tibuana
Journal of applied Industrial Engineering-University of PGRI AdiBuana

p-ISSN 2622-2027
e-ISSN 2622-2035

Application of K-Means Clustering for Detection of Downy Mildew At Madura Corn Plant Using Digital Image Processing

Imron Rosyadi NR[1], Erwin Prasetyowati[2], Badar Said[3],Syaiful Arifin[4] and Mohammad Syafiir Ridoni[5]

[1,2,4,5]Department of Industrial Engineering, Faculty of Engineering, Universitas Madura
[3] Department of Informatics, Faculty of Engineering, Universitas Madura

imron.rosyadi@unira.ac.id

***Abstract*—The development and cultivation of corn are necessary in line with the increasing consumption of food ingredients and industrial needs, especially food products made from corn. In the development of maize in Indonesia, the main obstacle is the disturbance of Plant Pest Organisms (OPT), especially diseases, one of which is downy mildew. This disease can be identified by a change in color, so we need a way to find out the difference between the color of healthy leaves and the color of leaves that have changed due to downy mildew. One solution that can be used is image processing. Therefore this study aimed to detect downy mildew based on leaf color in corn plants based on digital image processing, to produce precise and objective results. The algorithm used is the K-Means Clustering algorithm. This study uses 50 images of training data and 25 images of test data. Based on the simulation of downy mildew disease identification using K-Means Clustering it achieves an accuracy rate of 85%.**

Keywords: **Madura Corn, Image Processing, Segmentation, K-Mean Clustering.**

I. INTRODUCTION

Corn is one of the superior products in Pamekasan Regency in the agricultural sector. The development and cultivation of this commodity are necessary in line with the increasing consumption of food ingredients and industrial needs, especially food products made from corn. demand for this commodity tends to increase yearly for public consumption and the food industry (Rahman, 2015). Corn is widely used as a feed source and can also be processed into various food products that have economic value such as corn flour, corn oil, and corn snack products (Hermawati, 2016).

Regional mapping to determine the productivity of corn commodities, in general, has been carried out by the research team, especially in Pamekasan Regency. However, production results that tend to fluctuate will certainly disrupt the supply of this commodity. One of the contributing factors is the presence of corn disease, resulting in various growth disturbances such as crop failure which is a direct loss for farmers and for business owners or SMEs whose business is engaged in the processing of corn-based foods.

A common disease of corn that attacks is downy mildew. The emergence of downy mildew in corn plants is due to the fungi Peronosclero spores maydis and P. spores javanica and P. spores philippinensis, which triggers leaf discoloration (Budhi et al., 2019).

When exposed to disease, there will be changes in certain areas, either in the form of color changes or the emergence of different leaf pattern patterns (Budhi et al., 2019). Common symptoms that can be observed are a change in the color of the leaves to yellow and on the underside of the leaves there is white layer of fungal spores (Purwanto et al., 2016). If it continues, it can make the cob change shape and content. On mature plants, there are brown streaks on old leaves. Another disease is leaf blight, which is caused by the fungus Helminthosporium turcicum with symptoms that appear as elongated and regular spots that are yellow and surrounded by a brown tint. The leaves will then turn yellowish brown to dark brown (Ulhaq & Masnilah, 2019).

Based on these problems, this research will detect downy mildew in corn plants by utilizing image processing to find out which leaves are healthy and not from the color changes they experience. Image processing uses the

p-ISSN 2622-2027
e-ISSN 2622-2035

CIELab color space. The CIELab color space has been widely used in previous studies, including palm image recognition and image compression (Budhi et al., 2019), as well as image quality improvement based on luminance and saturation components (Giri et al., 2014).

The segmentation process to distinguish colors in the CIELab color space can utilize data mining techniques, namely K-Means which is often used for image segmentation (Zheng et al., 2018). K-Means is the most well-known clustering method and is widely used in various fields because it is simple, easy to implement, and has the ability to cluster large data. K-Means is a partitioning clustering method that separates data into different groups (Prasetyowati & Rofiq, 2016).

A digital image is defined as a two-dimensional image processing with a computer. The basic technology for creating and displaying color in digital images is based on research that color is a combination of three colors, namely red, green, and blue (Red, Green, Blue – RGB). Image segmentation is generally based on the nature of the discontinuity or similarity of the pixel intensity. The discontinuity approach is to partition the image when there is a sudden change in intensity (edge based). The similarity approach is to partition the image into regions that have certain characteristics in common (region-based) (Zheng et al., 2018).

In addition, the authors use K-Mean clustering which is a data mining method that performs the modeling process without supervision (unsupervised) and is one of the methods for grouping data with a partition system. The K-means method attempts to group existing data into several groups, where data in one group have the same characteristics as each other and have different characteristics from data in other groups. This method minimizes variation between data in a cluster and maximizes variation with data in other clusters (Prasetyowati et al, 2023; Lasena Y, 2020).

II. RESEARCH METHOD

This study uses two methods of collecting information and data including:

a. Observation, namely by observing directly, namely by looking at and taking the required data.

Figure 1:Diseased corn leaves

b. Interviews, this method is carried out by way of dialogue and discussion with Farmer Group sources and the Pamekasan Agricultural Service which handles Corn Plants.

Data processing will be divided into two stages, the first is the pre-processing stage, and the second is the processing stage. The output at the pre-processing stage functions to make input data for the second stage. In this system samples of input, BISI corn leaves are taken randomly, through the RGB to CIELab Color Transformation process. The steps to be able to identify downy mildew in bisi corn plants are:

1. Take a picture of a new corn leaf to be used as test data.
2. Enter test data in the form of an image of bisi corn leaves which will be compared with the corn leaf data which was used as the previous training data with a size of 200x200 pixels.
3. From the entered image, the a and b values will be obtained.
4. The results of the a and b values are segmented using K-Means, resulting in the average a and b values of each cluster.
5. The average cluster results will be compared with each data in the database using distance calculations.
6. The value of calculating the distance that is getting smaller indicates that the test image detected by downy mildew is close to the quality of the image in the training data.

Initial data processing begins with RGB (Red-Green-Blue) corn leaf image data, in this initial process there are two stages: 1) RGB image which will be converted to LAB, to simplify the intensity of the image color to facilitate and speed up computational calculations. After that the image will separate the L values (luminance or brightness level), A (green to red) and B (blue to yellow) in this process the L value is not used because it affects the value when shooting; 2) The images converted to the Lab are segmented. The segmentation results are calculated on average

p-ISSN 2622-2027
e-ISSN 2622-2035

per cluster, then the Euclidean distance is calculated and matched into the dataset.

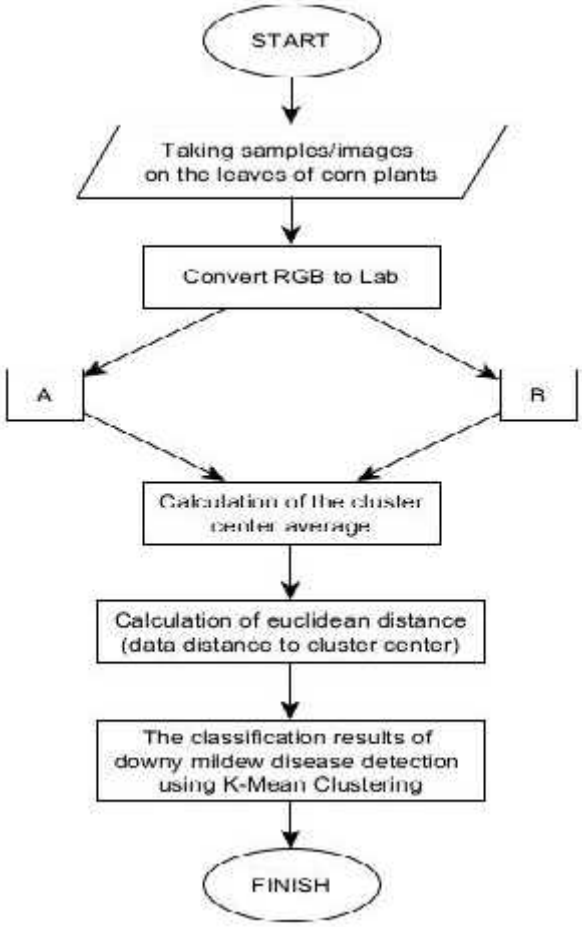

Figure 2:Image recognition process

III. RESULT AND DISCUSSION

The input data is a sample of BISI corn leaves. Where for the program in this study consisted of 25 samples of diseased corn leaves and 25 samples of leaves that were not diseased which would be used as dataset parameters. In this system samples of input, BISI corn leaves are taken randomly, through the RGB to CIELab Color Transformation process.

Before the segmentation process is carried out, the image is pre-processed. The first stage of the process is the Resizing process, which is changing the image size to 256 x 256 pixels. The resized image is then improved by enhancing the image quality by enhancing the contrast. The process of reducing the size of the image and improving the image is carried out to make the input data uniform because the convolution operation can only be performed on image data that has the same dimensions of length and width.

Figure 3: Resizing and enhancement process

p-ISSN 2622-2027
e-ISSN 2622-2035

In this study, K-Means clustering is used to separate regions in an image based on differences in image color. The preprocessed image which was originally in the RGB color space is converted to the L*a*b color space and then clustered using the a and b components. The number of clusters used is 3. Cluster 1 is represented by a blue object, cluster 2 is colored cyan, and cluster 3 is colored yellow. Of the three clusters, 1 cluster is selected based on the ROI (Region of Interest) or a certain area/part of the desired image.

From the input image, features have to be extracted. In this case, texture extraction is used in this study. Texture is one of the characteristics that can be extracted from a digital image. Texture can be used as a feature that distinguishes one image from another. The method used is GLCM (Gray-Level Co-Occurrence Matrix). In this study, the extraction was carried out on images that had been segmented. The extracted GLCM features are Contrast, Correlation, Energy, and Homogeneity extracted. And by using the MATLAB statistics command, other properties/formulas are found, namely Mean Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, and IDM. Then the extracted data is used as training data which is entered into a data set to later be used to determine image classification.

From the input image, features have to be extracted. In this case, texture extraction is used in this study. Texture is one of the characteristics that can be extracted from a digital image. Texture can be used as a feature that distinguishes one image from another. The method used is GLCM (Gray-Level Co-Occurrence Matrix). In this study, the extraction was carried out on images that had be segmented. The extracted GLCM features are Contst, Correlation, Energy, and Homogeneity extracted. And by using the MATLAB statistics command, other properties/formulas are found, namely Mean Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, and IDM. Then the extracted data is used as training data which is entered into a data set to later be used to determine image classification.$R-U = \begin{bmatrix} 147 & 147 & 147147 & 235 \\ 149 & 235 & 235235 & 149 \\ 235 & 149 & 155150 & 147 \\ 255 & 147 & 235147 & 235 \\ 147 & 235 & 149150 & 147 \end{bmatrix}$

$$G-U = \begin{bmatrix} 149 & 235 & 149235 & 149 \\ 255 & 149 & 147147 & 150 \\ 149 & 235 & 235147 & 235 \\ 235 & 155 & 150150 & 155 \\ 255 & 155 & 147235 & 155 \end{bmatrix}$$

$$B-U = \begin{bmatrix} 150 & 150 & 155149 & 150 \\ 147 & 235 & 149235 & 147 \\ 235 & 147 & 155149 & 235 \\ 147 & 149 & 235155 & 150 \\ 235 & 235 & 147235 & 149 \end{bmatrix}$$

The calculation steps are:

1) Transform the RGB color space into XYZ color space

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = \begin{bmatrix} 0{,}412553 & 0{,}357580 & 0{,}180423 \\ 0{,}212671 & 0{,}715160 & 0{,}072169 \\ 0{,}019334 & 0{,}119193 & 0{,}950227 \end{bmatrix} \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix}$$

$$= \begin{bmatrix} 0{,}412553 & 0{,}357580 & 0{,}180423 \\ 0{,}212671 & 0{,}715160 & 0{,}072169 \\ 0{,}019334 & 0{,}119193 & 0{,}950227 \end{bmatrix} \begin{bmatrix} 147 \\ 149 \\ 150 \end{bmatrix}$$

$$X_1 = (0{,}412553 \times 147) + (0{,}357580 \times 149) + (0{,}180423 \times 150)$$
$$= 140{,}9733$$
$$Y_1 = (0{,}212671 \times 147) + (0{,}715160 \times 149) + (0{,}072169 \times 150)$$
$$= 148{,}6467$$
$$Z_1 = (0{,}019334 \times 147) + (0{,}119193 \times 149) + (0{,}950227 \times 150)$$
$$= 163{,}1357$$

2) Determine the value L*,a*, and b*
ause$Y_1 = 148{,}6467 > 0 L^* = L_1 = 116(148{,}6467)^{\frac{1}{3}} - 16 = 598{,}4828$

Next, the calculation obtainsa^*

$$a^* = a_1 = 500(140{,}9733)^{\frac{1}{3}} - (148{,}6467)^{\frac{1}{3}} = 2602{,}2500$$

r get$b^*$is:$b^* = b_1 = 200(148{,}6467)^{\frac{1}{3}} - (163{,}1357)^{\frac{1}{3}} = 33{,}3611$

p-ISSN 2622-2027
e-ISSN 2622-2035

This study, the values used are from the $L^*a^*b^*$, color space, namely the values of the a^*,andb^*color components only. The reason the L^* value or brightness value is not used is because the L^* value has no effect on the image segmentation process.

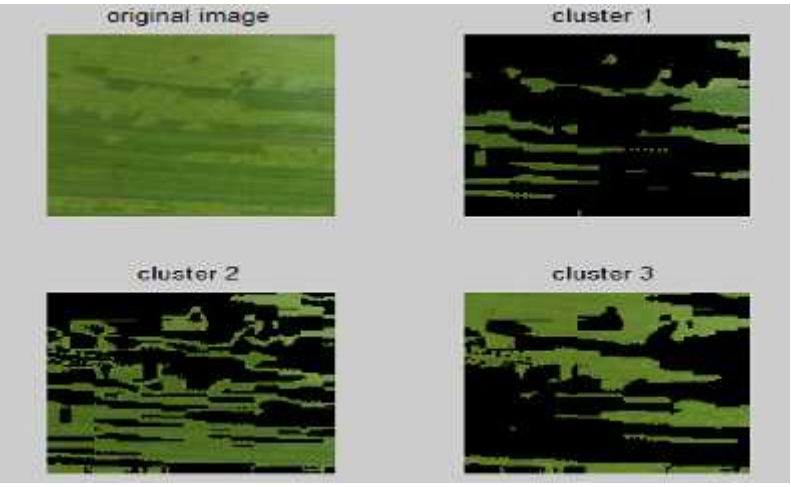

Figure 4:Analyzing process using K-Means

Next is the process of introducing test data which can be seen in the following table 1.

Table 1. Units for Magnetic Properties

Test image	Euclidean distance	Decision	Explanation
	0.191289	diseased	success
	0.190465	healty	success
	0.162174	diseased	success
	0.48225	diseased	success
	0,251525	healty	success
	0.263358	diseased	failed

The test aims to be able to evaluate the programs that have been implemented, especially to be a measure of the extent to which the program can detect corn leaf spot properly. Testing is carried out by measuring the accuracy of the image results from the program with manual images according to expert directions.

From the identification results, it was found that the success rate of the downy mildew disease identification program in corn plants using K-Means Clustering resulted in a total of 20 sample sheets of test data with a result of 85%.

IV. CONCLUSIONS

After testing the program and analyzing the simulation that has been made, the following conclusions can be drawn:

1. The results of program testing can be seen from the input type of corn leaf image, where the accuracy obtained reaches 85%.
2. The function of the K-Means algorithm here is as a clustering which plays a very, very important role in grouping data.

The suggestion for the development of this research is to create a system that does not only consist of color features, so that it can provide better accuracy in making decisions because of the many features involved. This is due to the many types of diseases in corn plants. In addition, it is necessary to add other data mining methods such as K-NN and SVM to support the grouping of images on leaves so that they have more precise accuracy.

p-ISSN 2622-2027
e-ISSN 2622-2035

REFERENCES

Budhi, R. K., Prayitno, A., & Elvina, S. (2019). Pengenalan Pola Daun untuk Pendeteksi Dini Penyakit Tanaman Jagung Menggunakan Deteksi Tepi Sobel. In *Seminar Nasional APTIKOM*. https://publikasi.dinus.ac.id/index.php/semnastik/article/download/2880/1758

Giri, K. J., Peer, M. A., & Nagabhushan, P. (2014). A Robust Color Image Watermarking Scheme Using Discrete Wavelet Transformation. *International Journal of Image, Graphics and Signal Processing*, *7*(1), 47–52. https://doi.org/10.5815/ijigsp.2015.01.06

Hermawati, D. T. (2016). Kajian Ekonomi antara Pola Tanam Monokultur dan Tumpangsari Tanaman Jagung, Kubis dan Bayam (Issue 1). https://journal.uwks.ac.id/index.php/inovasi/article/download/590/545

Kurniawan, A. F., Prasetyo, J., & Suharjo, R. (2017). Identifikasi Dan Tingkat Serangan Penyebab Penyakit Bulai Di Lampung Timur, Pesawaran, Dan Lampung Selatan (Vol. 5, Issue 3). https://jurnal.fp.unila.ac.id/index.php/JA/article/view/1824

Lasena Y, M. Y. (2020). Clustering Komoditi Unggulan Daerah Provinsi Gorontalo Menggunakan Algoritma K-Means. *Jambura Journal of Electrical and Electronics Engineering (JJEEE)*, *2*(1), 14–18. https://ejurnal.ung.ac.id/index.php/jjeee/article/view/4392/1734

Prasetyowati, E., & Rofiq, A. A. (2016). Penilaian Kinerja Keuangan Koperasi Pada Dinas Koperasi Dan UMKM Pamekasan Dengan K-Means. *Simantec*, *5*(2).

Prasetyowati, E., Rosyadi NR, I., &Rachmatullah, S. (2023). Penerapan K-Means Algorithm UntukMengidentifikasi Supplier Bahan Baku PadaKomoditasAgrikultur Di KabupatenPamekasan. Simantec, 5(2), 147-156.

Purwanto, D. S., Nirwanto, H., & Wiyatiningsih, S. (2016). Model Epidemi Penyakit Tanaman: Hubungan Faktor Lingkungan Terhadap Laju Infeksi Dan Pola Sebaran Penyakit Bulai (Peronosclerospora maydis) Pada Tanaman Jagung Di Kabupaten Jombang. *Berkala Ilmiah Agroteknologi - Plumula*, *5*(2), 138–152. http://ejournal.upnjatim.ac.id/index.php/plumula/article/view/764

Rahman, S. (2015). Analisis Nilai Tambah Agroindustri Chips Jagung. *Jurnal Aplikasi Teknologi Pangan*, *4*(3), 108–111. https://jatp.ift.or.id/index.php/jatp/article/view/136

Ulhaq, M. A., & Masnilah, R. (2019). Pengaruh Penggunaan Beberapa Varietas dan Aplikasi Pseudomonas fluorescens untuk Mengendalikan Penyakit Bulai (Peronosclerospora maydis) pada Tanaman Jagung (Zea mays L.). *Jurnal Pengendalian Hayati*, *2*(1), 1. https://doi.org/10.19184/jph.v2i1.17131

Zheng, X., Lei, Q., Yao, R., Gong, Y., & Yin, Q. (2018). Image segmentation based on adaptive K-means algorithm. *Eurasip Journal on Image and Video Processing*, *2018*(1). https://doi.org/10.1186/s13640-018-0309-3