Jurnal Mantik, 5 (4) (2022) 2152-2156

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Disease in Corn Leafe Using Gabor Wavelet and K-Means Clustering Algorithm

Mhd Furqan[1], Armansyah[2], Nurhasanah[3]

[123]Departmen Computer Science, Faculty of Science and Technology,
[123]Universitas Islam Negeri Sumatera Utara Medan, JL. wipe. Golf No. 120 Pancur Batu, Deli Serdang Regency, North Sumatra, Indonesia

Email: mfurqan@uinsu.ac.id[1], armansyah@uinsu.ac.id[2], hasanah221297@gmail.com[3]

ARTICLEINFO

Article history:

Received: Des 21, 2021
Revised: Jan 20, 2022
Accepted: Feb 30, 2022

Keywords:

Corn leaf disease,
Gabor Wavelet Filter,
K-means Clusterin

ABSTRACT

Abstract This study aims to develop a system to classify diseases that attack corn leaves. This study used four types of disease, namely: leaf blight (Helminthosporium turcicum), leaf spot (Bipolaris maydis syn), leaf rust (Puccinia polysora) and downy mildew (Peronosclerospora maydis). This study uses 52 data in the form of images. Every image is changed into vector data using Gabor wavelet filter. This study uses the K-Means Clustering method for disease grouping. The data in this study are vector data. This research process goes through the stages of preprocessing, clustering, and accuracy testing. Preprocessing includes Gabor wavelet filters to extract vector data from the original image. Clustering uses K-Means by determining the starting point manually and calculating similarity using Euclidean Distance. Independent testing of accuracy by comparing the system and manual. The highest accuracy is 98% of the 51 correct data using 52 data with 4 data cluster labels.

1. Introduction

Corn grows well in hot and cold areas with adequate rainfall and irrigation. However, during one life cycle from seed to seed, every part of corn is sensitive to a number of diseases, especially on corn leaves, so it can reduce the quantity of yield [1]. Plant disease is a condition in which plant cells and tissues do not function normally due to continuous interference by pathogenic agents or environmental factors and will result in the development of symptoms. Similarly, plant diseases that attack the leaves of corn plants, namely leaf *blight (Helminthosporium turcicum)*, leaf spot disease *(Bipolaris maydis syn)*, leaf rust (*Puccinia polysora)* and downy mildew (*Peronosclerospora maydis)* [2].

Knowledge of corn diseases and disease groupings is based on experience, counseling and not yet computerized which at any time can easily add new data about the latest diseases and pests by classifying new corn diseases entered in the system, so that farmers can know and understand disease group. The grouping of this disease is based on the type of disease and the shape of the spots on the corn leaves. In this case, a method is needed to formulate the problem. The method used is digital images by processing images of corn plant diseases on leaves using *Gabor Wavelet* and *K-Means Clustering* [3].

This study aims to create a system to classify diseases that attack corn leaves. And using 60 data in the form of images. Each image is converted into vector data using *Gabor Wavelet* andmethod *K-Means Clustering* to classify diseases. This research process goes through the stages of *preprocessing*, *clustering*, and accuracy testing. *Preprocessing* includes afilter *Gabor Wavelet* to extract vector data from the original image. *Clustering* is a K-Means by determining the starting point manually and calculating the similarity using *Euclidean Distance*. Independent testing of accuracy by comparing the system and manual [4]. By using *wavelets*, an ordinary vector space is decomposed into a set of nested vector spaces with different resolutions, thus enabling the analysis of functions in both the time and frequency domains at different resolutions [5].

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

K-means clustering is an algorithm used in partitioning the data to separate data into different groups. This algorithm is able to minimize the distance between the data to *cluster*its. The *K-Means algorithm* only takes part of the number of components obtained to be used as thecenter *cluster* initial, in determining the center *ofcluster, it* thisis chosen randomly from the data population. Then thealgorithm *k-means* will test each of every *ster* that has been defined previously dependent on the minimum distance between komponekomponen in the population data and mark these components into one central *clu*n with eachcenter. *cluster* Furthermore, the position of the center of the *cluster* will be recalculated until all data components are classified into each *cluster* and finallywill be formed *acluster* new[6]

2. Method

Methods This exploratory work methodology is expected to describe each stage of the exercise carried out during the research so that it will be in accordance with the objectives which have been specified. The stages to be completed include the following:

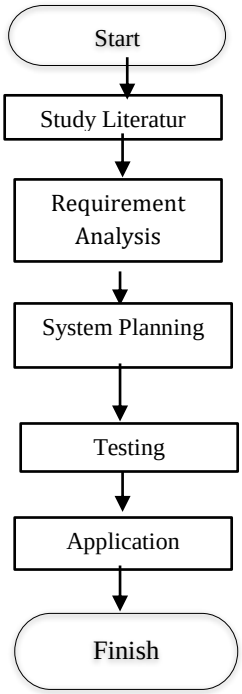

Fig 1 Flowchart of Work Procedures

The system design flow is in the stage of identifying corn plant disease patterns using Gabor wavelet and *k-means clustering*. The process begins with inputting the image of corn leaves, then preprocessing the RGB input will be converted to *grayscale.* Theresults *grayscale* obtained a size and extracted using Gabor wavelet the results from the Gabor wavelet were obtained vector data generated from the texture size by calculating the pattern from points, into lines then into a pattern, different sizes will distinguish each pattern so that the extraction process is obtained, namely the difference each pattern or depth is called a texture. The vector data will be input to k-means, then k-means will group adjacent vector data into one subgroup.

The workflow of *K-Means Clustering* for classifying corn diseases. This stage begins with deciding how many clusters are expected to group the existing data into how many groups. After that then take some data as much as the cluster as the centroid. Then compare the centroid data with other data in order to find out whether the first data with the centroid have similarities or not, if there are similarities it will be made in one group and vice versa by calculating applying *Euclidean Distance*. Then if after the calculation you get a part of a number of clusters that have been decided , then a new centroid will be obtained and compare the new centroid with other data and then repeat thecalculation stage *Euclidean Distance* until the cluster does not change or the last data is the same. And the final cluster is obtained from the grouping.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

3. Results and Analysis

Reaction To bring out the characteristics of each gray image, the Gabor Wavelet method is used to retrieve the vector values used in the image. At this stage, 4 variables are used, namely *u, v, m, n*. Each is the scale, orientation, column and row of the grayscale image. At this stage. An array value is used in the image which becomes the variable value (*u*,*v*). This process is called Gabor Filter which is used to extract images by taking scale and orientation values.

Then to output the vector values, the Gabor Features function is used which uses variable values in the gaborArray function and adds variables to the image and factors in each row and column in each image. These variables are img, gaborArray (*u, v, m, n*), d1 (row factor in the image), d2 (column factor in the image), d1 and d2 must have the same value so that the image size must be square and have a value. the same matrix. To get the vector value, the formula:

$$\frac{m \times n \times u \times v}{d1 \times d2}$$

Then, to get the vector value, several parameters are used using theparameter *Gabor wavelet*which is used to extract texture features with the parameters stage, orientation, freq, flags and values. fourier. To get the vector value, thefunction is used *Gabor wavelet*.

In the gabor process on the vector data that has been obtained, the next step is the process of determining the*cluster* of 52 data with 4 labels. Thestage *Clustering* uses *K-Means Clustering* to form *clusters* on the 52 data. The following is a list of sample data and sample points in *K-Means Clustering*.

Table 1
List of

Sample Data	No	X	Y
Bulai (1)	1	8	5
Spots (2)	2	4	6
Blight (3)	3	3	5
Carat (4)	4	2	3

Then to determine thepoint *centroid* , the cluster consists of 4 with the example as following:

Table 2
List of Centroids 1 and 2

No	Centroid	X	Y
1	D (1)	8	5
2	D (2)	4	6

a. The next step is to calculate the distance of the data to the *centroid* using theformula *Euclidean Distance* to determine the group of each *cluster*.

Table 3
Results of Grouping 1 and 2

Cluster	Centroid		Group
	X	Y	
C_1	0	4.1	1
C_2	4.1	0	2

b. The next step is to continue the calculation to data 3, namely (3.5). The step is to calculate the distance of data 3 to centroid 1 and centroid 2, to find out the 3rd data goes to centroid 1 or 2. Then it is entered into the table and determines the closest point to data 1 and 2 with a minimum distance:

Table 4
Results of Grouping Data 3

Data	Distance		Group
	X	Y	
Data 3 (3.5)	5	1.41	2

c. Then the data above will update the centroid value. So that the Cluster Data becomes:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Table 5
Data Update Results 3

No	Centroid	X	Y
1	D (1)	8	5
2	D (2)	3.5	8.5

3.1 The update table will be used for the next calculation with the same steps (iterations). This continues until all clusters and data get their respective clusters.

The number of data is 52 consisting of 4 clusters with 1 cluster each containing 13 data. The K-Means function uses the Distance parameter and the Euclidean distance value.

To display the results of the *cluster* and *centroid* , the results of thecalculation are used *Euclidean distance* as in the example above by following the results of the *centroid* that has been determined with the *centroid of the* results. With details: The

a. center of the centroid is the center of the cluster marked with the symbol **X**.
b. For cluster 1 is marked dot with awith acolor cyan.
c. For cluster 2 are marked with the color yellowdot.
d. For cluster 3 is marked with a magentadot.
e. For cluster 4 is marked dot with awith acolor black.

Based on the cluster calculation, the cluster results are obtained as follows:

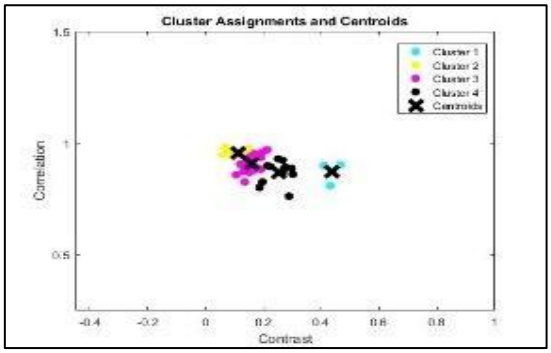

Fig 2 Centroid Cluster Results

By using K-Means Clustering, the centroid value will occupy each *cluster* that is close to its centroid value.

This process gets the results of a cluster that has a total of 52 data to 51 correct data from the 52 data cluster in question. The 52 data consists of 4 Clusters with each having 13 centroids. So it is accumulated to 96% accuracy in the second test and 98% in the first test.

Total accuracy of the K-Means Clustering Method. Of the 52 points Centroid has 52 clusters, when the centroid belonging to Cluster Results results matched with Cluster Label then didapatlah accuracy using the equation:

$$Accuracy = \frac{jumlah\ correct\ data}{jumlah\ Data} \times 100$$

So accuracy is obtained is 98% of 51 the correct data by using 52 data with Cluster label as much as 4 data.

4. Conclusions

This study obtained several conclusions in accordance with the problem formulations that have been compiled in the initial chapter. By using K-Means Clustering , the diseases found in corn leaves in the form of Bipolaris maydis syn, helminthosporium turcicum, puccinia polysora, peronosclerospora maydis can be grouped with experiments that can produce the highest accuracy of 98%. In this study, vector features of corn leaf disease in the form of Bipolaris maydis syn, helminthosporium turcicum, Puccinia polysora, Peronosclerospora maydis disease can be created and show separate vector results as a texture feature of the disease. Suggestions for this research are, add some other diseases in corn plants and improve the accuracy obtained. Added other features to the Matlab GUI design such as descriptions and brief knowledge about corn farming.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

References

[1] Delfana, U., Rahmad, C., Alifia, M., & Tupamahu, F. (2020). Segmentation Based on K-Means on Image Detection of Corn Leaf Diseases. JIP (Polinema Informatics Journal) Vol 6 , 38.

[2] Febrinanto, FG, Dewi, C., Wiratno, AT, 2018, Implementation of theAlgorithm K-Meansas an Image Segmentation Method in Identification of Citrus Leaf Disease.Journal of Information Technology and Computer Science Development, Vol. 2,5375-5383.

[3] Furqan, M., Sriani., Harahap, LS, 2020, Classification of bougainvillea leaves using Gray Level Co-Occurrence Matrix and K-Nearest Neighbor. Journal of CoreIT, Vol.6.

[4] Furqan, M., Sriani., Sari, IEY, 2020, Application of the Otsu Method in PerformingImage Segmentation on Imagery in Figs b. Journal of Management, Informatics and Computer Engineering, Vol. 20, 59-72.

[5] Gonzalez, RC, Woods, RE (2008). Digital Image Processing Third Edition. New Jersey: Prentice Hall.

[6] Hutagalung, SN, 2018, Learning Basic Physics and Basic Electronics Using the Simulink Method Matlab Application. Journal of Science and Social Research, Vol 4307, 30-35.

[7] Ide, Base., 2011. Health Secret Of Turmeric. Jakarta : PT Elex Media Komputindo

[8] Kusumanto, R., & Tompunu, N. (2011). Digital Image Processing To Detect Objects Using RGB Normalization Model Color Processing. National Seminar on Applied Information and Communication Technology .

[9] Ma'rifatin, U. (2020). Implementation of the K-Means Algorithm for Grouping Patients' Diseases at the Warujayeng Public Health Center. National Seminar on Technology Innovation UN PGRI Kediri , 286-287.

[10] Munir, R. (2004). Digital Image Processing with an Algorithmic Approach. Bandung: Informatics

[11] Padmo, A., & Murinto. (2016). Batik Image Segmentation Based on Texture Features Using Gabor Filter Method and K-Means Clustering. Journal of Informatics Vol 10 , 1175.

[12] Setiawan, A. (2016). Application of the Gabor Wavalet Algorithm as Home Security By Identifying Faces Based on Webcams. Explora Informatics Vol 5.