Classification of Orchid Types Based on Leaf Images Using the Gray Level Co-occurrence Matrix (GLCM) and K-Nearest Neighbor (KNN) Methods

[1]Rosa Prahasti, [2]Sriani Sriani
[12]Computer Science, Faculty of Sains and Technology
[12]State Islamic University of North Sumatra, North Sumatra, and Indonesian
[1]rosaprahasti423@gmail.com , [2]sriani@uinsu.ac.id

Orchids are popular ornamental plants with high economic value. Various types of orchids in Indonesia are difficult to distinguish when they are not in bloom, especially for laymen. This study aims to develop a classification system for orchid species based on leaf images using the Gray Level Co-occurrence Matrix (GLCM) and K-Nearest Neighbor (KNN) methods. GLCM texture feature extraction includes contrast, correlation, energy, and homogeneity, which are then classified using the KNN algorithm. The research dataset consists of 88 leaf images for training data and 22 images for test data from 11 orchid species, namely dendrobium, arachnis, cattleya, arundina, phalaenopsis, cymbidium bicolor, doritis polcherrima, appendiculata, flickingeria aggulata, phaius tankervilleae, and vanda douglas. The test results show that k=1 produces the highest accuracy of 77.2%, k=3 produces an accuracy of 72.7%, and k=5 produces an accuracy of 59.0% on the test data. The selection of the optimal k parameter was performed using Grid Search with the cross-validation method to ensure better model generalization. This study shows that the combination of the GLCM and KNN methods is effective for classifying orchid species based on leaf image characteristics with good accuracy.

Keywords— Leaves, Orchids, Gray Level Co-occurrence Matrix (GLCM), K-Nearest Neighbor (KNN)

I. INTRODUCTION

Orchids are one of the most popular ornamental plants in the world. The beauty of the flowers, the variety of colors, and their unique shapes make them particularly attractive[1]. According to data from the Central Statistics Agency (BPS) in 2024, there were approximately 4,429,032 orchids produced in Indonesia, with North Sumatra having around 3,126 species. Generally, orchids can be identified and distinguished by the shape of their flowers, but leaf characteristics can also be used as a primary reference for recognizing the diverse types of orchids[2][3].

The diversity of orchid species available at local plant nurseries often makes it difficult for consumers to distinguish between them, especially for those who identify plants based solely on the shape of their flowers. This is because orchid leaves generally have an elongated oval shape and a stiff texture, making them appear similar across species.

With advances in technology, the process of plant identification, which was previously done manually, can now be supported by technology-based approaches. Through feature extraction techniques on orchid leaf images, the unique characteristics of each type of orchid can be revealed.

Plant variety classification can be performed by analyzing leaf images. This study proposes a feature extraction method, specifically for the texture of orchid leaf images. To obtain texture values, the Gray Level Co-occurrence Matrix (GLCM) method is used.[4]. One algorithm that is often used in image classification is K-Nearest Neighbor (KNN). K-Nearest Neighbors is performed by searching for a group of K objects in the training data that are most similar to the objects in the testing data. The distance used is Euclidean Distance. The K-NN method is performed by comparing the test data with the training data. In this research, the Python programming language will be used and integrated with Excel to calculate the value of k as the number of closest neighbors [5].

This study developed a classification of eleven orchid leaf varieties using the GLCM and KNN methods combined with Grid Search to systematically determine the optimal K value. The application of Grid Search in this study served as objective validation to prove that the manual selection of K= [1,3,5] based on the characteristics of the dataset and literature was optimal. In addition, this study uses Python as the programming language because it is more flexible, has extensive image processing libraries, and supports the development of artificial intelligence-based systems. With this innovation, the study contributes to the development of a more accurate, efficient, and widely applicable orchid leaf image classification method.

II. METHOD

This study uses quantitative research methodology, where quantitative research is a research method that focuses on collecting and analyzing data in numerical form to test predetermined hypotheses or theories[6].

A. Literature Review

Image processing is the process of manipulating and analyzing images that are closely related to visual perception. This process has a key characteristic, namely image-based input and output data. In general, digital image processing is defined as the processing of two-dimensional images using a computer. Image processing aims to improve image quality, extract information, and process visual data to make it easier for humans and computer systems to analyze.[7][8].

Feature extraction is a process that aims to reduce data dimensions, whereby complex raw data sets are simplified into a set of information that is easier to manage and analyze. Feature extraction serves as a method for selecting or combining certain variables into more concise features, without losing the essence of the original information in the data[9].

Gray Level Co-occurrence Matrix (GLCM) is one of the methods used in texture analysis and feature extraction in digital images[10]. This method works by forming a matrix that represents the frequency of occurrence of pairs of pixels with a certain gray level in an image, based on a specific distance and direction. In this study, there are four main components of GLCM that are used, namely correlation, energy, homogeneity, and contrast. These four components are considered sufficient for analyzing and processing images effectively[11].

K-Nearest Neighbor (K-NN) is a classification method that belongs to supervised learning algorithms, where new test data will be classified based on the majority category of the K nearest neighbors. The main purpose of this algorithm is to group new data based on attribute similarities with existing training data[12]. K-NN method uses distance calculations to classify new data. The learning (training) and classification or testing processes are the two main stages in the KNN (K-Nearest Neighbor) technique[13][14].

Grid search is a validation and optimization method that works by determining grid points that are equidistant from each other, then calculating the accuracy value for each parameter so that the most optimal parameter point (the point with the best accuracy) can be found[15]

B. Observation

The observation activity was carried out at Madirsan Flower Garden, located in Bangun Sari Village, Tanjung Morawa District, Deli Serdang Regency.This location was chosen based on the diversity of orchid species from various genera growing in the area.Through direct observation, data was obtained on the morphology of orchid leaves, their shape, and differences in characteristics between species. The results of these observations formed the basis for the selection of 11 orchid species used as research objects.

C. Interview

An interview was conducted with a local orchid seller named Mrs. Toeety at the research location to obtain additional information about orchid types, physical characteristics of leaves, and other details. This information assisted researchers in the process of selecting samples and grouping image data according to orchid leaf characteristics.

D. Data Collection

The data used consisted of orchid leaf images from 11 species :
Dendrobium, Arachnis, Cattleya, Arundina, Phalaenopsis, Cymbidium Bicolor, Doritis Polcherrima, Appendiculata, Flickingeria Aggulata, Phaius Tankervilleae, and Vanda Douglas.

The dataset consisted of 110 leaf images, consisting of 88 training images and 22 test images. All image were captured with a smartphone camera under natural lighting conditions and then converted to a grayscale image of 512 x 512 pixels to prepare them for the feature extraction stage.

E. Research Design

The following image shows a flowchart of orchid leaf classification using Gray Level Co-occurrence Matrix and the K-Nearest Neighbor method.

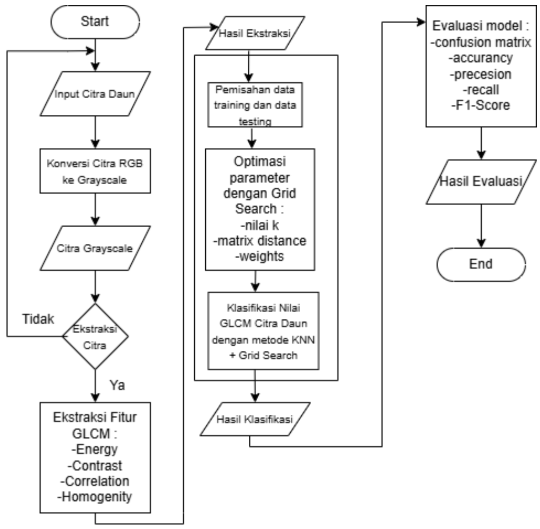

Figure 1. Flowchart of Orchid Leaf Type Classification

The process of classifying orchid leaf images consists of several stages. The first stage involves converting RGB images into grayscale images measuring 512x512 pixels in PNG/JPG format. The next stage involves extracting orchid leaf texture features using GLCM by calculating four features, namely energy, contrast, correlation, and homogeneity with 4 angles, namely 0°, 45°, 90°, and 135°. The following are the formulas for the 4 GLCM features:

1. Energy

$$Energy = \sum i,j\, P(i,j)^2$$

2. Contrast

$$Contrast = \sum i,j\, (i,j)^2 P(i,j)$$

3. Correlation

$$Correlatio = \sum i,j \frac{(i-\mu i)(j-\mu j)P(i,j)}{\sigma i \sigma j}$$

4. Homogenity

$$\text{Homogeneity} = \sum \text{i,j} \frac{\text{P(i,j)}}{1+|\text{i-j}|}$$

After the feature extraction process is complete, the data is separated into training and testing data. Next, classification is performed using the K-Nearest Neighbor (KNN) method by manually determining the k value, namely k=[1,3,5]. To improve classification performance, optimization is performed using Grid Search, which functions to find the best parameters in the KNN model.

F. Testing

The test was conducted to evaluate the method's ability to classify orchid leaf types. The image data was divided into two parts, namely training data and test data. The testing process used grayscale images in *bitmap format with a size of 512x512 pixels. Feature extraction was performed using the GLCM method and classification using K-Nearest Neighbor (KNN), which was implemented using the Python programming language in Visual Studio Code (VSCode)[16].

G. Evaluation

The evaluation was conducted by calculating the accuracy level of the system from the classification results of orchid leaf images. Accuracy was calculated using the following formula:

$$Accuracy = \frac{\sum Correct\ Predictions}{\sum Test\ Data} \times 100\%$$

III. RESULTS AND DISCUSSION

A. Data Analysis

The data for analysis in this study are samples of grayscale images of orchid leaves measuring 4x4 pixels, which will be classified by applying the Gray Level Co-occurrence Matrix (GLCM) method. Furthermore, to test the K-Nearest Neighbor (KNN) method, sample testing was carried out in accordance with the designed system. The data used consisted of four types of orchid leaves classified into the following types: Dendrobium, Arachnis, Cattleya, Arundina, Phalaenopsis, Cymbidium Bicolor, Doritis Polcherrima, Appendiculata, Flickingeria Aggulata, Phaius Tankervilleae, and Vanda Douglas.

B. System Flowchart

A system flowchart is a flowchart that serves to show the process flow in the system to be built. In this system, the flowcharts that will be needed consist of a main menu flowchart, a flowchart about the author, and a classification process flowchart. The following are all the flowcharts that will be built.

1. Main Menu Flowchart

This main menu flowchart shows the process flow when the user is in the main menu.

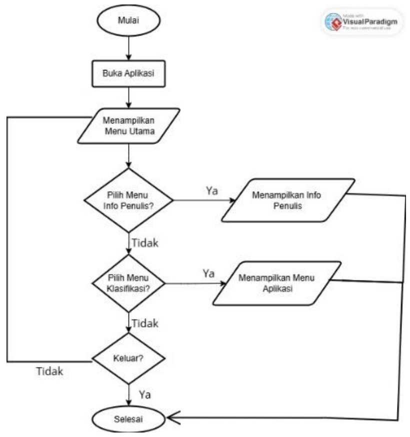

Figure 2. Main Menu Flowchart

Based on the flowchart in the main menu, it can be explained that when users run the system, they will be presented with two menu options, namely the author info menu, which contains information about the author, and the classification process menu, which is used to process the classification of orchid leaf images.

2. Author Info Flowchart

The author info flowchart is a flowchart that appears when the user opens the author info menu.

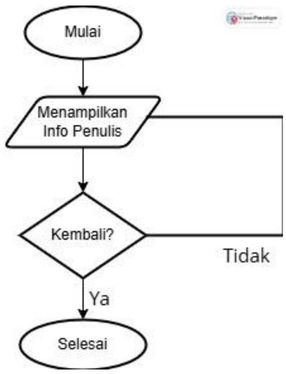

Figure 3. Author Information Flowchart

Based on the author information flowchart, it can be explained that users are presented with information about the researcher

3. Classification Flowchart

The classification flowchart is a diagram of the process flow when users classify orchid leaf images.

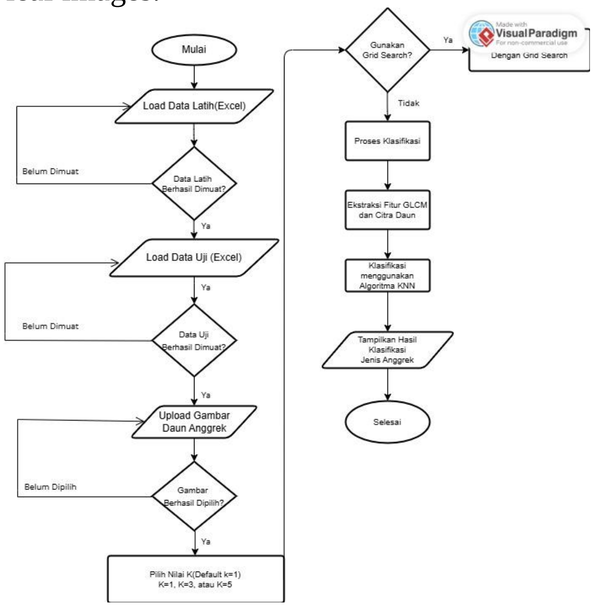

Figure 4. Classification Process Menu Flowchart

Based on the classification process flowchart above, it is explained that the first step is for the user to load training data in Excel format. then the user selects the orchid leaf image to be tested. For the classification process, we can choose which K value to test, then process it using the GLCM method. The extracted results are then searched for the closest value according to our preferences using the training data with the KNN method, resulting in the classification of orchid types that match the training data.

C. System Testing

After completing data analysis and system model design, the next step is to test the system model developed using Python. The following is a preview of the program that has been built.

1) Main Menu Program Preview: The following is a preview of the main menu.

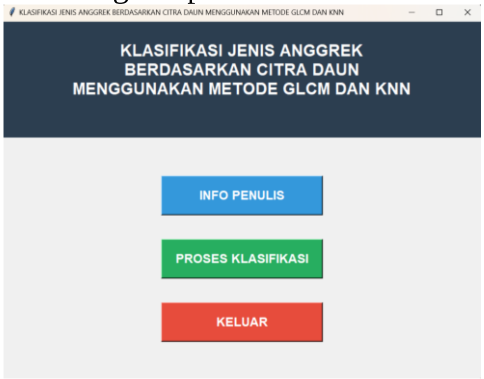

Figure 5. Main Menu Display

2) Author Info Menu Display: The following is the author info menu display..

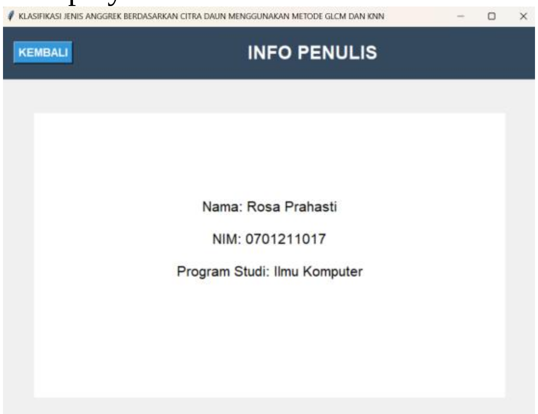

Figure 6. Author Info Menu

3) Classification Process Menu Program Display: The following is the classification process menu display. It has been built into the system model using GUI in Python..

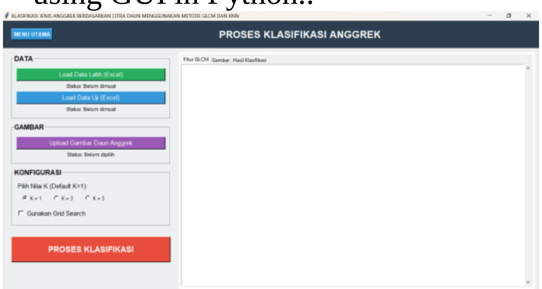

Figure 7. Classification Process GUI Display

The stages in testing this system involved preparing training data and test data in the form of 88 orchid leaf images as training data and 22 orchid leaf images as test data. The following is the complete data from the orchid leaf images presented in tabular form:

Tabel 1. Training Data and Test Data Information

Data Citra	**Jumlah**	**Jenis Anggrek**	**Jumlah**
Data Latih	88 Citra	Dendrobium	8
		Arachnis	8
		Arundina	8
		Cattleya	8
		Cymbidium Bicolor	8
		Doritis Polcherrima	8
		Phalaenopsis	8
		Appendiculata	8
		Flickingeria Aggulata	8
		Phaius Tankervilleae	8
		Vanda Daouglas	8
Data Uji	22 Data	Dendrobium	2
		Arachnis	2
		Arundina	2
		Cattleya	2
		Cymbidium Bicolor	2
		Doritis Polcherrima	2
		Phalaenopsis	2
		Appendiculata	2
		Flickingeria Aggulata	2
		Phaius Tankervilleae	2
		Vanda Daouglas	2

Based on the orchid leaf image information table, the next step is to prepare the training leaf image data and test leaf image data, which are entered into the training data and test data folders as follows :

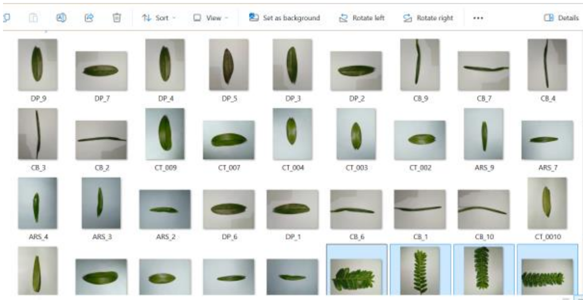

Figure 8. Image of Orchid Leaves Training

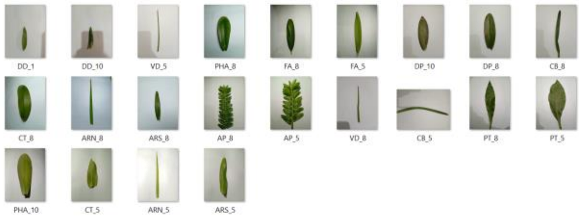

Figure 9. Orchid Leaf Image Test

The next step is to classify based on the prepared test leaf images. To start the orchid classification process, users can press the "load training data (Excel)" button. After pressing this button, a search menu will appear that is directly connected to the Excel folder containing the extracted training data. Then, after successfully entering the training data, the user presses the "Upload Orchid Leaf Image" button, and an image search menu will appear, allowing the user to select the images to be classified.

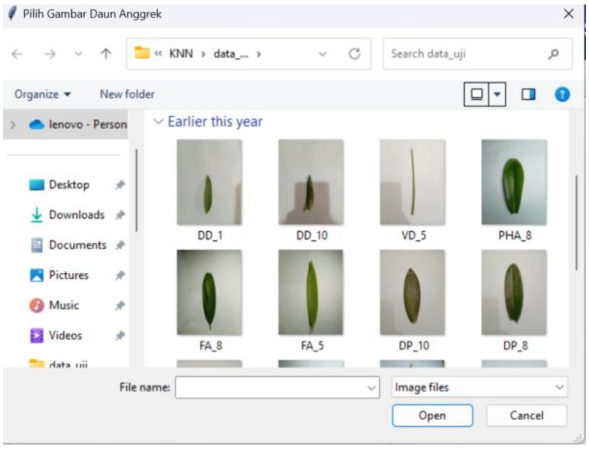

Figure 10. Image Search Menu

Next, after the user selects the image they want, they can determine the k value they want to predict.

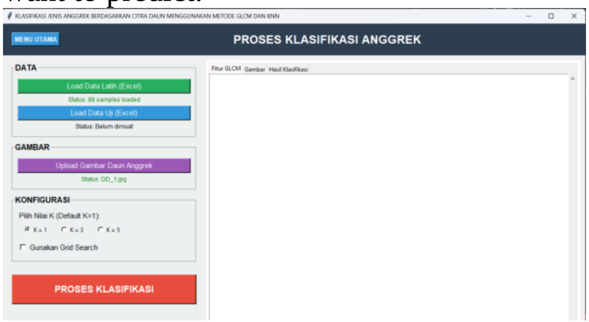

Figure 11. Selecting the K Value

Based on the image above, the user selects the "DD1" image, which is an image of a Dendrobium orchid leaf. After selecting the k value to be tested, the user presses the "classification process" button, which will display the GLCM feature extraction, the grayscale orchid image, and the classification results and percentages.

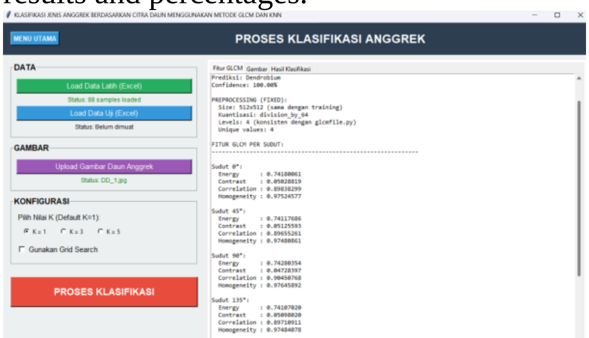

Figure 12. GLCM feature results

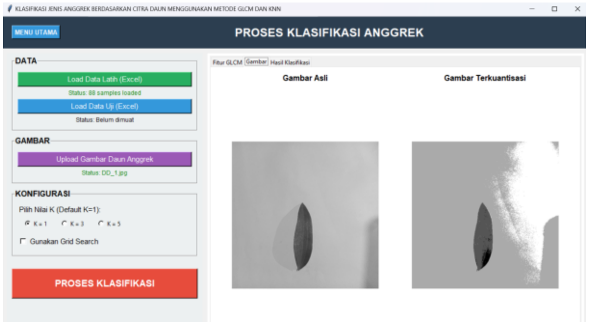

Figure 13. Hasil Grayscale

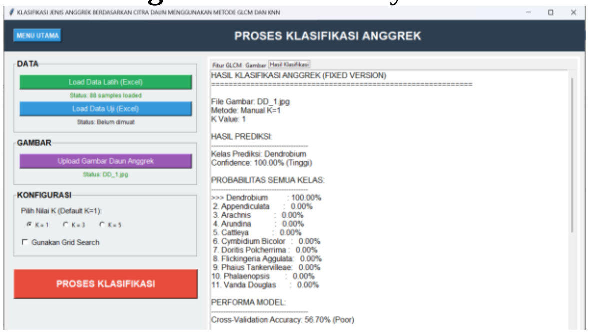

Figure 14. Leaf Image System Classification Results with k=1

Figure 15. Leaf Image System Classification Results with k=3

Figure 16. Leaf Image System Classification Results with k=1

Based on the image above, the classification process was successfully applied to the application system, where DD_1 leaves for k=1, k=3, and k=5 were correctly classified as "Dendrobium". These results indicate that the classification system is functioning correctly.

Figure 17. Leaf Classification Results with Grid Search

Based on the image above, the classification process uses the Grid Search method, where the results show that the best parameter is k=3 with the prediction class Dendrobium.

Testing of 22 test images from 11 types of orchids showed the following results:

Tabel 2. Test Results

Nama File	Jenis Anggrek	Hasil Klasifikasi		
		K=1	K=3	K=5
DD_1	Dendrobium	Dendrobium	Dendrobium	Dendrobium
DD_10	Dendrobium	Cattleya	Cattleya	Cattleya
ARS_8	Arachnis	Arachnis	Arachnis	Arachnis
ARS_5	Arachnis	Phaius Tankervilleae	Doritis Polcherrima	Doritis Polcherrima
ARN_8	Arundina	Arundina	Arundina	Arundina
ARN_5	Arundina	Arundina	Arundina	Dendrobium
CT_8	Cattleya	Cattleya	Cattleya	Cattleya
CT_5	Cattleya	Dendrobium	Dendrobium	Dendrobium
CB_8	Cymbidium Bicolor	Cymbidium Bicolor	Cymbidium Bicolor	Cymbidium Bicolor
CB_5	Cymbidium Bicolor	Cymbidium Bicolor	Cymbidium Bicolor	Cymbidium Bicolor
DP_8	Doritis Polcherrima	Doritis Polcherrima	Cattleya	Cattleya
DP_5	Doritis Polcherrima	Cymbidium Bicolor	Cymbidium Bicolor	Cymbidium Bicolor
PHA_8	Phalaenopsis	Phalaenopsis	Phalaenopsis	Phaius Tankervilleae
PHA_10	Phalaenopsis	Phalaenopsis	Phalaenopsis	Phalaenopsis
AP_8	Appendiculata	Appendiculata	Appendiculata	Appendiculata
AP_5	Appendiculata	Appendiculata	Appendiculata	Appendiculata
FA_8	Flickingeria Aggulata	Flickingeria Aggulata	Flickingeria Aggulata	Flickingeria Aggulata.
FA_5	Flickingeria Aggulata.	Flickingeria Aggulata.	Flickingeria Aggulata.	Flickingeria Aggulata.
PT_8	Phaius Tankervilleae	Phaius Tankervilleae	Phaius Tankervilleae	Doritis Polcherrima
PT_5	Phaius Tankervilleae	Flickingeria Aggulata.	Flickingeria Aggulata.	Flickingeria Aggulata.
VD_8	Vanda Daouglas	Vanda Daouglas	Vanda Daouglas	Vanda Daouglas
VD_5	Vanda Daouglas	Vanda Daouglas	Vanda Daouglas	Vanda Daouglas

Based on the table above, for k=1, 17 test images were correctly classified and 5 test images were classified incorrectly. For k=3, 16 test images were correctly classified and 6 test images were classified incorrectly. For k=5, there were 13 test images that were correctly classified and 9 test images that were incorrectly classified. The next step is to calculate the accuracy rate of the system based on the results of the test image classification. To perform this calculation, the following formula will be used:

$$Accuracy = \frac{\sum Correct\ Predictions}{\sum Test\ Data} \times 100\%$$

Thus, the highest accuracy of the orchid classification system based on leaf images is k=1 at 77.2%. This result was obtained from testing 88 orchid leaf images for training data and 22 orchid leaf images for test data from 11 orchid species.

IV. CONCLUSION

From the results of the analysis and testing that has been carried out, the process of classifying orchid types based on leaf images using the Gray Level Co-occurrence Matrix (GLCM) and K-Nearest Neighbor (KNN) methods shows good results. The system is capable of extracting leaf texture features and performing classification with an accuracy of 77.2% at k=1, 72.7% at k=3, and 59.0% at k=5. Based on validation using the Grid Search method with cross-validation, the value k=3 was selected as the best parameter because it provides more consistent results and is not easily affected by data variations. Therefore, the application of the GLCM and KNN methods optimized with Grid Search is considered effective in classifying orchid types through leaf images.

REFERENCES

[1] H. N. Syafira, A. Komariah, R. Nurhayatini, and R. Romiyadi, “Respon Pertumbuhan Tanaman Anggrek (Phalaenopsis fimbriata JJ. Smith) Akibat Perlakuan Berbagai Media Tanam Di Pembenihan,” *OrchidAgro*, vol. 2, no. 1, p. 1, 2022, doi: 10.35138/orchidagro.v2i1.368.

[2] A. faiza Al farikhi, Farrady Alif Fiolana, and Fajar Yumono, “Klasifikasi Bunga Anggrek Bulan Berdasarkan Warna Dan Teksturnya Menggunakan Metode Jst,” *J. Ilm. Sist. Inf.*, vol. 1, no. 3, pp. 25–37, 2022, doi: 10.51903/juisi.v1i3.417.

[3] S. F. Risdiana, S. A. Azharia, and A. Supriyatna, “Inventarisasi Dan Analisis Jenis Anggrek (Orchidaceae) Di Kampung Nambo, Desa Batukarut, Kecamatan Arjasari, Kabupaten Bandung,” *J. Ilmu Pertan. dan Perkeb.*, vol. 5, no. 2, pp. 41–50, 2023, doi: 10.55542/jipp.v5i2.713.

[4] C. Anggraini, “KLASIFIKASI DAUN KELENGKENG MENGGUNAKAN METODE GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) DAN K- NEAREST NEIGHBOR (KNN),” vol. 11, no. 2, pp. 72–78, 2024, doi: 10.30656/jsii.v11i2.9157.

[5] I. P. Arisanti and Y. Yamasari, “Mengenali Jenis Tanaman Obat Berbasis Pola Citra Daun dengan Algoritma K-Nearest Neighbors,” *J. Informatics Comput. Sci.*, vol. 3, no. 02, pp. 95–103, 2021, doi: 10.26740/jinacs.v3n02.p95-103.

[6] D. F. Berlianti, A. Al Abid, and A. C. Ruby, “Metode Penelitian Kuantitatif Pendekatan Ilmiah untuk Analisis Data,” *J. Rev. Pendidik. dan Pengajaran*, vol. 7, no. 3, pp. 1861–1864, 2024.

[7] A. D. Goenawan, M. Bakhara, A. Rachman, and M. P. Pulungan, “Identifikasi Warna Pada Objek Citra Digital Secara Real Time Menggunakan Pengolahan Model Warna HSV,” *J. Tek. Inform. dan Elektro*, vol. 4, no. 1, pp. 68–74, 2022, [Online]. Available: https://jurnal.ugp.ac.id/index.php/jutei

[8] Tri Laksono Aditiya, Endryansyah, Wanarti Rusmamto Puput, and Syariffuddien Zuhrie Muhammad, “Pengolahan Citra Digital Buah Murbei Dengan Algoritma LDA (Linear Discriminant Analysis),” *Indones. J. Eng. Technol.*, vol. 4, no. 2, pp. 71–78, 2022, [Online]. Available: https://journal.unesa.ac.id/index.php/inajet

[9] Amna *et al.*, *Data Mining Data mining*, vol. 2, no. January 2023. 2023. [Online]. Available: https://repository.uinjkt.ac.id/dspace/bitstream/123456789/71093/1/DATA MINING.pdf

[10] A. Nurjulianty, P. Purnawansyah, and H. Darwis, “Perbandingan Metode Naïve Bayes dan K-NN dengan Ekstraksi Fitur GLCM pada Klasifikasi Daun Herbal,” *J. Media Inform. Budidarma*, vol. 7, no. 4, p. 1740, 2023, doi: 10.30865/mib.v7i4.6262.

[11] H. P. Hadi and E. H. Rachmawanto, “Ekstraksi Fitur Warna Dan Glcm Pada Algoritma Knn Untuk Klasifikasi Kematangan Rambutan,” *J. Inform. Polinema*, vol. 8, no. 3, pp. 63–68, 2022, doi: 10.33795/jip.v8i3.949.

[12] Yahya and W. P. Hidayanti, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok El ektrik) pada ‘ Lombok Vape On ’ Pendahuluan dihasilkan tidak stabil dan tidak mampu diprediksi Dari penelitian yang dilakukan , berusaha untuk mengklasifikasik,” *J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 104–114, 2020, [Online]. Available: https://e-journal.hamzanwadi.ac.id/index.php/infotek/article/view/2279/pdf_23

[13] Sriani, Supriyandi, M. Furqan, and W.

Fadilla Rischa, "Pengenalan Pola Penyakit Daun Jambu Air Menggunakan Metode PCA dan KNN," *J. Jar. Sist. Inf. Robot.*, vol. 7, no. 2, pp. 158–163, 2023, [Online]. Available: http://ojsamik.amikmitragama.ac.id

[14] M. Meiriyama, S. Devella, and S. M. Adelfi, "Klasifikasi Daun Herbal Berdasarkan Fitur Bentuk dan Tekstur Menggunakan KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2573–2584, 2022, doi: 10.35957/jatisi.v9i3.2974.

[15] S. T. Kusuma and T. B. Sasongko, "Optimasi K-Nearest Neighbor dengan Grid Search CV pada Prediksi Kanker Paru-Paru," *Indones. J. Comput. Sci.*, vol. 12, no. 4, pp. 2162–2171, 2023, doi: 10.33022/ijcs.v12i4.3267.

[16] M. B. Nendya, B. Susanto, G. I. W. Tamtama, and T. J. Wijaya, "Desain Level Berbasis Storyboard Pada Perancangan Game Edukasi Augmented Reality Tap The Trash," *Fountain Informatics J.*, vol. 8, no. 1, pp. 1–6, 2023, doi: 10.21111/fij.v8i1.8836.