OPEN ACCESS
Volume 9 (3), July-September 2025, 1016-1023
E-ISSN:2580-1643
Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)
DOI: https://doi.org/10.35870/jtik.v9i3.3826

Identification of Flower Type Images Using KNN Algorithm with HSV Color Extraction and GLCM Texture

Edhy Poerwandono [1], M. Endang Taufik [2*]
[1,2*] *Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.*

article info

Article history:
Received 10 January 2025
Received in revised form 10 February 2025
Accepted 10 March 2025
Available online July 2025.

Keywords:
Identification; HSV; GLCM; K-Nearest Neighbor.

Kata Kunci:
Identifikasi; HSV; GLCM; K-Nearest Neighbor.

abstract

Due to the variety of types of flowers that exist and having and tracking each variety, making plant lovers and cultivators difficult to distinguish in determining the type of flower, it takes a very long time to find out the type of flower if you only rely on the five senses. With the application of the K-Nearest Neighbor algorithm and feature extraction of color and texture, it is very helpful in image processing to identify flowers more easily and shorten the time, with the greatest accuracy of 71% using the K-7 value, the flower was successfully carried out.

abstrak

Karena beraneka ragamnya jenis bunga yang ada dan memiliki serta melacak setiap varietasnya, membuat pecinta dan pembudidaya tanaman kesulitan membedakan dalam menentukan jenis bunga, butuh waktu yang sangat lama untuk mengetahui jenis bunga jika hanya mengandalkan panca indera. Dengan penerapan algoritma K-Nearest Neighbor dan ekstraksi ciri warna dan tekstur sangat membantu dalam pengolahan citra untuk mengidentifikasi bunga dengan lebih mudah dan mempersingkat waktu, dengan akurasi terbesar sebesar 71% dengan menggunakan nilai K-7 bunga berhasil di identifikasi.

ACM Computing Classification System (CCS)
EBSCOhost
Communication and Mass Media Complete (CMMC)

Corresponding Author. Email: endangtaufik@gmail.com [2].

1. Introduction

Flowers consist of leaves and twigs that are around them and have undergone many changes. These changes are due to the results of a number of enzymes that are stimulated by certain phytohormones (Amasino, 2010). The formation of a flower is tightly controlled genetically, and many types are induced by certain environmental changes, such as low temperatures, lack of sunlight, and available water. Flowers are almost always equilateral, which is often used as a characteristic of a taxon. There are two forms of flowers based on the symmetry of their shape: actinomorph and zygomorph. The actinomorphic form is very easy to find (Gogul & Kumar, 2017). With the beauty of the flower, it causes an increase in the number of plant lovers and cultivators, and flower production continues to grow and develop (Yuan *et al.*, 2018). Therefore, computer technology is developing rapidly as a tool to obtain information that can be used to identify types of flowers (Sharma *et al.*, 2021). Computers can process data quickly, precisely, and accurately with a high level of precision, making them useful tools for obtaining necessary information in various fields (Huang & Aizawa, 1993). The development of image processing technology today provides the opportunity to create systems that can recognize objects in digital images (Poerwandono & Taufik, 2025).

Image processing is also a technology used to solve problems related to image manipulation (Parker, 2010). In image processing, images are processed in a way that allows them to be used for further applications (Vasconcelos *et al.*, 2009). Therefore, the demand for information regarding types of flowers encourages the creation of a system for flower identification based on digital image processing. Previous studies have successfully integrated digital image processing with artificial intelligence for various applications, such as face recognition in schools (Shafiyyatul *et al.*, 2023), fish classification (Apriyanti *et al.*, 2013), detection of bacteria during the COVID-19 pandemic based on HSV color (Gogul & Kumar, 2017), and batik detection through texture extraction (Joly *et al.*, 2014). Moreover, the KNN algorithm has been successfully applied in chronic kidney disease classification (Poerwandono & Taufik, 2025). Based on these previous studies, the idea of identifying flowers was developed due to the increasing number of plant lovers and the growing cultivation of ornamental plants.

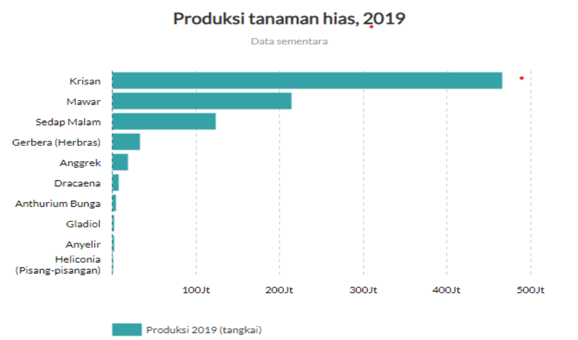


Figure 1. Ornamental Plant Production 2019

2. Research Methodology

This research methodology is based on the stages of digital image processing.

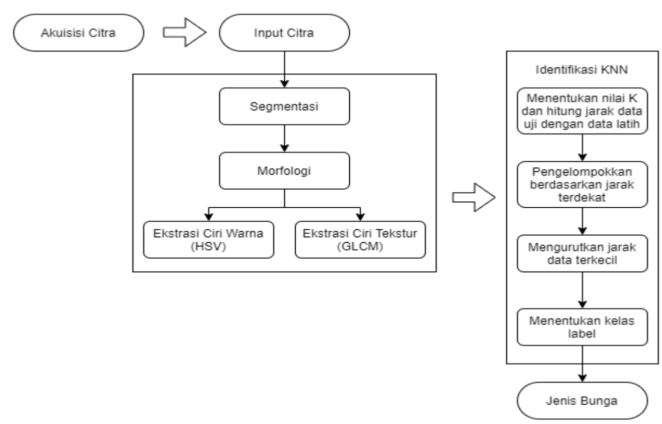


Figure 2. Research Flow

The data collection phase involves determining the necessary data and selecting an appropriate digital image processing method. In this stage, image retrieval is carried out from the Kaggle website, specifically from the "Flower Color Image" dataset, which consists of 10 flower types used for identification based on color and texture. Each flower type has 100 images, averaging 500x500 pixels in .png format. The pre-processing phase follows, which includes cropping, resizing, segmentation, and noise removal to prepare the images for further analysis. Cropping is performed to extract specific regions of the image containing the flower, and resizing reduces the image size to enhance processing speed. Image segmentation separates the objects from the background, enabling individual analysis of each flower. Otsu thresholding is applied to convert the RGB images into grayscale, making it easier to identify homogeneous areas for further processing. This is followed by morphological operations, such as erosion, which help remove noise. After pre-processing, feature extraction is performed, beginning with HSV color features, where the Hue, Saturation, and Value attributes are used to describe the flower's color. Additionally, GLCM texture features, such as contrast, correlation, homogeneity, and energy, are extracted to characterize the texture of the flowers. Finally, the K-Nearest Neighbor (KNN) algorithm is employed for classification. In KNN, the flower images are mapped into a multi-dimensional space where each dimension represents a feature, and classification is based on the proximity of the data points. The KNN algorithm uses Euclidean distance to determine the closest neighbors and assigns a class label based on the majority class among the nearest neighbors.

$$Dij = \overset{n}{\underset{K=1}{\%\&}}(i-j)^2$$

Information:
Dij : Euclidean distance between i and j.
i : data on the ith x for the calculation stage.
j : data on the jth y for the calculation stage.

3. Results and Discussion

Results

The results of this study are how to identify flower type images by applying HSV color feature extraction and GLCM texture K-Nearest Neighbor (K-NN) algorithm to identify flower type images into ten classes, and obtain the accuracy of the evaluation results in identifying flower type images for each type of flower as many as 10 images, with a dataset of 100.

1) Data Collection: At this stage, the type of flower will be imaged by visiting the website https://www.kaggle.com/with the title Flower Color Image then download the entire public data.

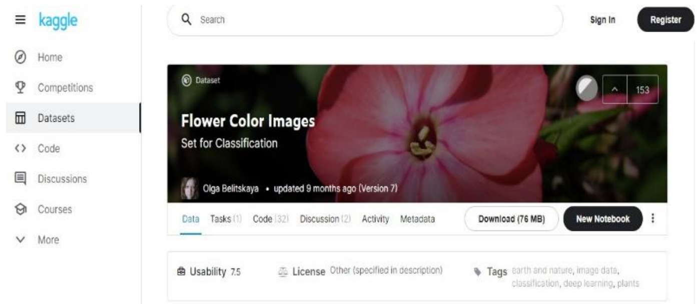
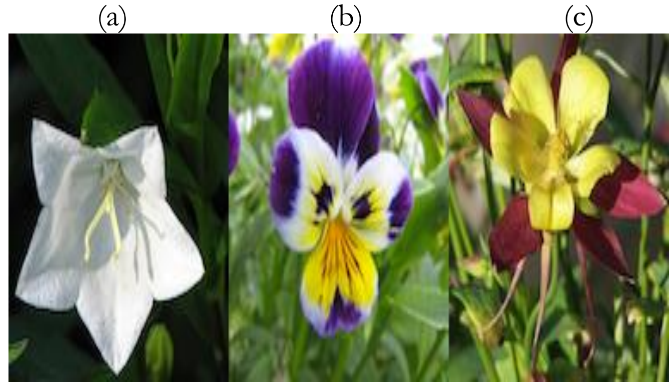
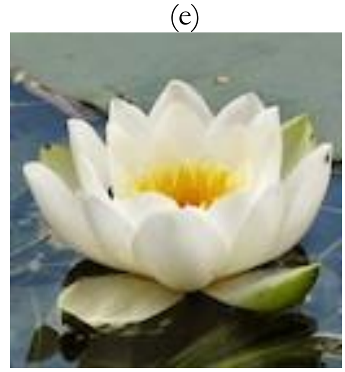


Figure 3. Kaggle Web Page Here are the

(a) (b) (c)



(d) (e) (f)



(g)

Figure 4. .(a) Rose flower (b) Calendula flower (c) Iris flower (d) Bellflower flower (e) Viola flower (f) Lilium flower (g) Water Lily flower

2) Pre Processing: Pre Processingused to simplify theprocess of identifying images. This stage consists of cropping, compressing image size, image segmentation and removing noise from the image. Cropping and compressing image size are done using third-party applications with an image size of 500 x 500 pixels and a ratio of 1:1.

Figure 5. Original Image

Figure 6. Grayscale Image

Then image segmentation using MATLAB software with Otsu thresholding to facilitate the image processing process to the next stage with images that have been converted from RGB images to grayscale images, then morphological segmentation is carried out to remove noise.

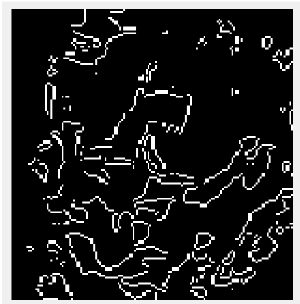


Figure 7. Binary Image

3) Feature Extraction: In this study, the feature extraction used is color feature extraction because the factor that affects the difference in flower types is color. The colors included for feature extraction are hue, saturation, and value as well as GLCM texture feature extraction which includes contrast, correlation, homogeneity and energy. Feature extraction has various variations, the more feature extractions that are relevant to the image are used, the better the accuracy results. The following is an HSV feature extraction image while for GLCM features using a grayscale image.

Figure 8. HSV Image

4) K-Nearest Neighbor: The stage of identifying flower type image data using the K-Nearest Neighbor (KNN) algorithm. The data from the previously standardized feature extraction stage has been obtained, then it will be divided into two in the identification process, namely training data and testing data. In dividing the data, the percentage split technique is applied where there are three data division scenarios. The following are details of the three scenarios of the percentage split technique and the results of class division on flower type images

Table 1. Percentage Split

No	*Split Data*	*Training Data*	*Testing Data*
1	70% and 30%	70	21
2	60% and 40%	60	28

Table 2. Types of Flowers

No	ClassFlower	TypeFlower
1	FlowerA	Rose Flower
2	FlowerB	Calendula Flower
3	FlowerC	Iris
4	BungaD	Bellflower
5	FlowerE	Viola Flower
6	FlowerF	Lilium Flower
7	FlowerG	Water Lily Flower

Steps in implementing the data sharing scenario that has been set, the identification process is carried out using the K-Nearest Neighbor algorithm with the k values used being 3, 5 and 7. The following are the results of the identification test.

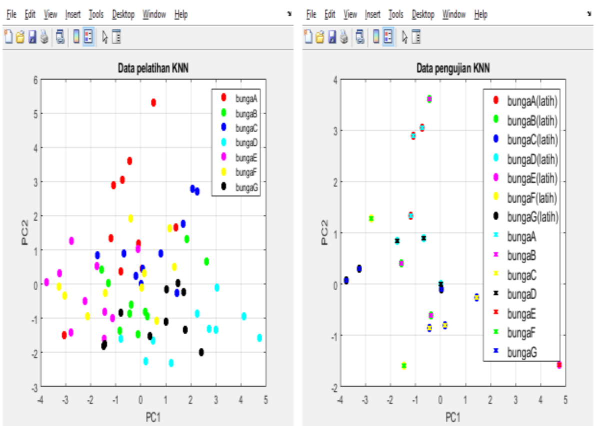


Figure 9. Visualization of KNN training and testing data results

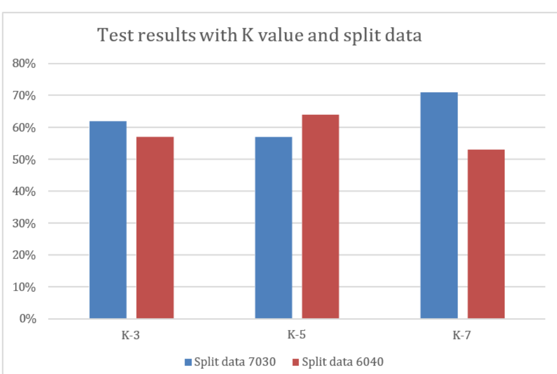


Figure 10. Test Results With K Value and Split Data

The first test involved dividing the dataset using a 70% training data and 30% testing data split. With a total of 100 images, consisting of 10 images per flower type, this resulted in 70 images for training and 21 images for testing. The identification accuracy for this split was found to be 62% for K-3, 57% for K-5, and 71% for K-7. In the second test, a 60% training and 40% testing data split was used. This resulted in 60 training images and 28 test images. The accuracy for this split was 57% for K-3, 64% for K-5, and 53% for K-7. The purpose of testing with different K values and data splits was to evaluate how varying ratios of training to testing data affect the accuracy of the KNN method (Poerwandono & Taufik, 2025). After completing these tests, the optimal K value and corresponding accuracy were selected for use in the graphical user interface (GUI) in MATLAB. In the GUI, several buttons are provided, each serving a specific function. The first button, labeled "Image Input," allows users to upload an image for testing, as illustrated in the following image.

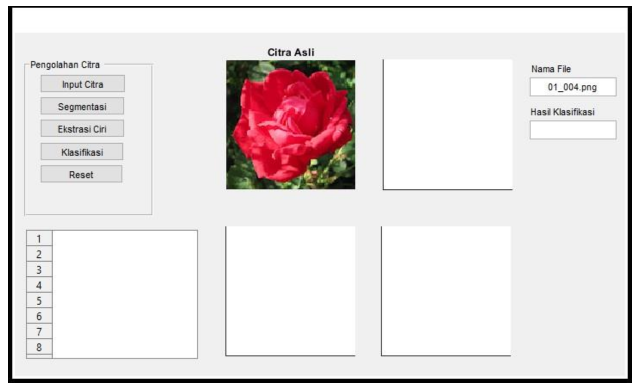


Figure 11. Image Input

After the image is successfully input, the image will appear on axes 1 or as the original image and on the right side there is the name of the file that has been input. Next there is a segmentation button that functions to segment the image with morphology to help eliminate noise, as seen in the following figure.

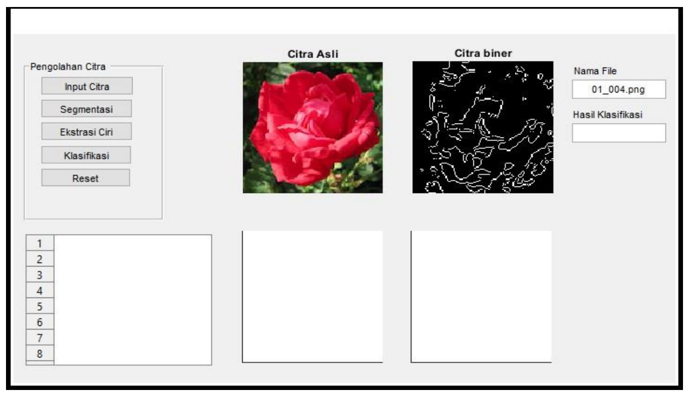


Figure 12. Image Segmentation

After carrying out the preprocessing stage, a feature extraction search will be carried out on the image, as can be seen in the following image.

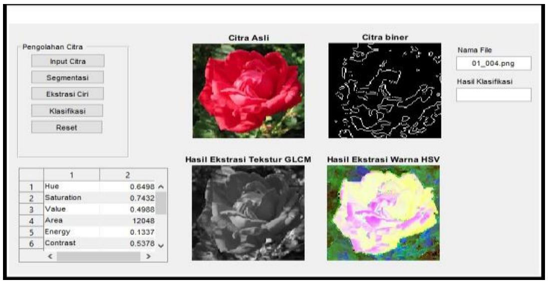


Figure 13. Feature Extraction

Because it uses GLCM texture features and HSV color features, there are grayscale image results for HSV texture and color features, in the table on the left side there are feature extraction values from an image. After getting the extraction value results on the image, an identification will be carried out on the image. Can be seen in the following Figure.

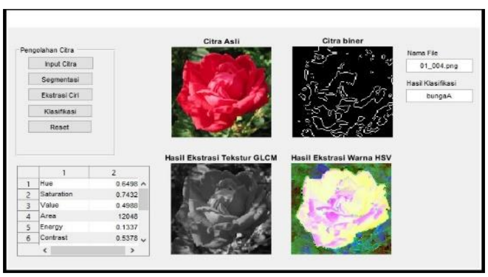


Figure 14. Identification Results

Based on the test graph above, there are two scenarios of division between training data and testing data where the division details have been explained in Table 1. In each of these scenarios, each experiment was carried out with various K, namely K-3, K-5 and K-7. From the results shown in Figure 10, it can be seen that the ratio of 70:30 and K-7 is a fairly good result compared to other tests. And testing with GUI succeeded in identifying the image of the type of flower according to its class.

Table 3. Test Results with K ValueNo

No	Split Data	K Value	Accuracy
1		K-3	62%
2	700	K-5	57%
3		K-7	71&
4		K-3	57%
5	60 40	K-5	64%
6		K-7	53%

Discussion

This study aims to identify flower species using the K-Nearest Neighbor (KNN) algorithm with HSV color feature extraction and GLCM texture analysis. The research process began with data collection, followed by image pre-processing, feature extraction, and finally, the application of the KNN algorithm for classification. In the data collection phase, flower images were obtained from the "Flower Color Image" dataset available on Kaggle, which includes 10 different flower species. Each species is represented by 100 images, with an average size of 500x500 pixels in PNG format. This dataset was used to identify flower species based on their color and texture features. The image acquisition from Kaggle aligns with previous studies that used content-based image retrieval for plant species identification (Apriyanti *et al.*, 2013). The pre-processing phase is crucial for preparing images to be used in the identification system. This phase includes cropping, resizing, image segmentation, and noise removal. Cropping is performed to select relevant parts of the image, while resizing reduces the image size to accelerate processing time. Image segmentation is done using Otsu thresholding, which converts RGB images to grayscale, facilitating the recognition of objects and backgrounds. This approach is similar to techniques used in image processing for object recognition tasks (Pavlidis, 2012). Morphological operations, such as erosion, are applied to remove noise, which has been shown to improve image quality for further analysis (Parker, 2010).

After pre-processing, feature extraction is performed. HSV color feature extraction is used to describe the flower's color, with three main components: Hue (basic color), Saturation (color intensity), and Value (brightness) (Huang & Aizawa, 1993). These features provide essential information about the flower's color, which is a crucial characteristic in identifying plant species. Additionally, texture features are extracted using the Gray Level Co-occurrence Matrix (GLCM), which provides second-order statistical values such as contrast, correlation, homogeneity, and energy. This technique has been widely used in image texture analysis for object classification (Acharya & Ray, 2005; Sharma *et al.*, 2021). Following feature extraction, the K-Nearest Neighbor (KNN) algorithm is applied to classify flower species. The KNN algorithm works by measuring the proximity of data points in a multi-dimensional space, where each dimension represents a feature of the flower. The class of the flower is determined based on the majority class of its nearest neighbors. In this study, three different values for K (3, 5, and 7) were tested to assess the impact on classification accuracy. The results showed that a K value of 7 yielded the highest accuracy, reaching 71% with a 70% training and 30% testing data split. This indicates that increasing the number of nearest neighbors can enhance the classification results, which is consistent with previous research on KNN for flower species recognition (Gogul & Kumar, 2017). However, the results of this study are limited by the small dataset used. Although an accuracy of 71% was achieved, the classification performance could be improved by increasing the dataset size. Further research with a larger dataset and the implementation of additional algorithms or feature extraction methods could potentially improve the accuracy, as suggested in similar studies in the field of digital image processing (Joly *et al*., 2014). In conclusion, this study successfully identified flower species using HSV color and GLCM texture feature extraction combined with the KNN algorithm. The achieved accuracy demonstrates the potential of using image processing for plant species identification. Further research, including the use of a larger dataset and the exploration of alternative feature extraction methods, could enhance classification accuracy and expand the application of digital image-based flower identification systems.

4. Conclusion

The conclusion obtained in the study of Identification of Flower Type Images using HSV Color and GLCM Texture is that digital image identification of flower types using the feature extraction method for color and texture was successfully carried out with data divisions of 70:30 and 60:40 and testing on K-3, K-5, K-7 values. Getting the best accuracy in data division 70:30 using a K-7 value of 71%, the results obtained were considered lacking because the data obtained was very limited to use.

5. References

Acharya, T., & Ray, A. K. (2005). *Image processing: principles and applications*. John Wiley & Sons.

Amasino, R. (2010). Seasonal and developmental timing of flowering. *The Plant Journal*, *61*(6), 1001-1013. https://doi.org/10.1242/dev.063511.

Apriyanti, D. H., Arymurthy, A. M., & Handoko, L. T. (2013, November). Identification of orchid species using content-based flower image retrieval. In *2013 International conference on computer, control, informatics and its applications (IC3INA)* (pp. 53-57). IEEE.

Gogul, I., & Kumar, V. S. (2017, March). Flower species recognition system using convolution neural networks and transfer learning. In *2017 fourth international conference on signal processing, communication and networking (ICSCN)* (pp. 1-6). IEEE.

Huang, T. S., & Aizawa, K. (1993). Image processing: some challenging problems. *Proceedings of the National Academy of Sciences*, *90*(21), 9766-9769.

Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., ... & Barthélémy, D. (2014). Interactive plant identification based on social image data. *Ecological Informatics*, *23*, 22-34.

Mouradov, A., Cremer, F., & Coupland, G. (2002). Control of flowering time: interacting pathways as a basis for diversity. *The plant cell*, *14*(suppl_1), S111-S130.

Parker, J. R. (2010). *Algorithms for image processing and computer vision*. John Wiley & Sons.

Pavlidis, T. (2012). *Algorithms for graphics and image processing*. Springer Science & Business Media.

Poerwandono, E., & Taufik, M. E. (2025). Identification of Flower Type Images Using KNN Algorithm With HSV Color Extraction and GLCM Texture. *Router: Jurnal Teknik Informatika dan Terapan*, *3*(1), 01-14.

Sharma, A., Gupta, A., & Jaiswal, V. (2021). Solving image processing critical problems using machine learning. *Machine Learning for Intelligent Multimedia Analytics: Techniques and Applications*, 213-248.

Turner, B. M. (2009). Epigenetic responses to environmental change and their evolutionary implications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1534), 3403-3418.

Vasconcelos, M. C., Greven, M., Winefield, C. S., Trought, M. C., & Raw, V. (2009). The flowering process of Vitis vinifera: a review. *American journal of enology and viticulture*, *60*(4), 411-434.

Wood, A. (2023). *Leaves and Flowers*. BoD–Books on Demand.

Yuan, P., Li, W., Ren, S., & Xu, H. (2018). Recognition for flower type and variety of chrysanthemum with convolutional neural network. *Transactions of the Chinese Society of Agricultural Engineering*, *34*(5), 152-158.