Mantik Journal, 8 (2) (2024), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

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

Implementation Of GLCM (Gray Level Co-Occurrence Matrix) & KNN(*K-Nearest Neighbor*) For Classification Of Fiber Root Plant Types Based On Leaf Image

Erna Rosani Nubatonis
Undergraduate Informatics Engineering Study Program , Sekolah Tinggi Manajemen Informatika Komputer (STIKOM) Uyelindo Kupang, NTT

A R T I C L E I N F O

Article history:

Received Jul 12, 2024
Revised Jul 30, 2024
Accepted Aug 22, 2024

Keywords:

Classification;
Root Crops;
Image;
Glcm;
Knn;
Feature Extraction.

ABSTRACT

This study aims to implement the GLCM (Gray Level Co-Occurrence Matrix) and KNN (K-Nearest Neighbor) methods in the classification of fiber root species based on leaf images. Fibrous roots are the most common root type in certain plants, and classifying plant species based on leaf image can provide useful information in contacting plants. The GLCM method is used to extract texture features from leaf images. The GLCM matrix describes the relative occurrence of pixel pairs with different gray intensities in the image. These features can provide information about leaf texture that can be used in classification. Furthermore, the KNN algorithm is used to classify plant types based on the extracted features. The dataset used in this study consists of a number of leaf images representing several different types of fiber root plants. Image processing includes pre-processing to obtain a clean image and ensure consistency of image size. After feature extraction using the GLCM method, these features are used as input for the KNN algorithm. KNN is used to classify unknown leaf images into one of the plant classes that have been previously trained. The experimental results show that the GLCM and KNN methods can provide good results in the classification of fiber root plant species based on leaf images. High classification accuracy indicates the effectiveness of this method in identifying plant species based on textural features of leaf images. Thus, this method can be a useful tool in the field of plant recognition and other applications that involve identifying plant species based on leaf images.

Corresponding Author:

Erna Rosani Nubatonis,
Undergraduate Informatics Engineering Study Program
Sekolah Tinggi Manajemen Informatika Komputer (STIKOM) Uyelindo Kupang, NTT,
Jl perintis kemerdekaan I, Kel.Kayuputih Kota Kupang,NTT,85111, Indonesia.
Email: ernarosaninubatonis@gmail.com

1. INTRODUCTION

Indonesia is rich in biological natural resources (mega diversity) and has the greatest diversity in all regions Indonesia . One of the biodiversity that Indonesia has is the diversity of plant types, in Indonesian forests we can find approximately 30,000 types of plants . (Rosiva Srg et al., 2022) , Fiber root plants are a type of plant that has roots

consisting of many fine fibers that grow spreading in the soil (Ryan & Pigai, 2020) . This type of plant has important economic and ecological value, so identifying and classifying fibrous root plant types is an interesting topic in botanical research (Ardila et al., 2022) .In recent years, the use of digital image technology and image processing has become a popular method in plant analysis (Dwi Fasnuari et al., 2022) . Leaf image processing can provide important information about morphological and textural characteristics that can be used to classify plant types. (Akbarollah et al., 2023). One method commonly used in image texture analysis is GLCM (Gray Level Co-Occurrence Matrix). GLCM (Gray Level Co-occurrence Matrix) for the first time proposed by Haralick in 1979 with 28 features to explain spatial patterns l (Shandy et al., 2019) . This method utilizes the GLCM matrix to describe the relative appearance of pairs of pixels with different gray intensities in the image (Widodo et al., 2018) . Texture features extracted from the GLCM matrix, such as energy, contrast, homogeneity, and correlation, can be used as a powerful feature representation for classification. In addition to the feature extraction method (Cholil et al., 2021), the KNN (K-Nearest Neighbor) algorithm is used to classification of plant types. (Daffa et al., 2022) KNN is a simple but effective classification algorithm, where unknown objects are classified based on the majority class of their nearest neighbors in feature space (Agus et al., 2017) . In this research, the KNN algorithm is used to classify types of fibrous root plants based on texture features extracted using the GLCM method.Previous studies have proven that GLCM is an effective method for the extraction of texture features from images. GLCM can compute a matrix of concurrent occurrences of pixel values at specific distances and angles, thus providing texture information that can be used for classification. Previous research has shown that KNN is a simple but powerful algorithm for classification tasks, including image classification. KNN uses the principle of proximity to classify new samples based on the majority class of the nearest K-NN in the feature space. This study combines GLCM for the extraction of texture features with KNN for classification. Although these two methods have been widely used separately, this study explores the synergy of the two to improve the accuracy of the classification of fibrous root plant types. This study specifically focuses on the classification of fibrous root plant types based on leaf imagery. This is a topic that may not have been explored much in previous studies, thus making a new contribution in this domain.

The aim of this research is to implement the GLCM method and KNN algorithm in classifying fibrous root plant types based on leaf images. It is hoped that this research can contribute to the development of more sophisticated and effective leaf image analysis techniques for identifying fibrous root plants. In this research, a leaf image dataset will be used that includes several different types of fibrous root plants (Astuti et al., 2023) . The image processing process includes pre-processing to obtain a clean image and ensure consistency of image size (Budianita et al., 2019) . Next, the GLCM method will be used to extract texture features from leaf images (Napulun et al., 2024) . These features are then used as input for the KNN algorithm which will classify plant types. (Nuraeni et al., 2023)

It is hoped that this research will provide a better understanding of the use of GLCM and KNN methods in classifying fibrous root plant types based on leaf images. It is hoped that the results of this research can be used as a reference and basis for the development of more sophisticated and effective leaf image analysis techniques in the field of plant recognition and other applications involving the identification of plant types based on leaf images. This research also looks at several previous studies as follows with the title " Digital Image Classification Based on Feature Extraction Based on Texture Using GLCM with the K-Nearest Neighbor Algorithm. " The results of this research are using 4-way GLCM feature extraction which uses Energy with an average value 55.99%, the K-Nearest Neighbor method can produce the expected accuracy in the classification system (Rohpandi et al., 2018) , the next research with the title " Identification of the Image of Free-range Chicken and Broiler Meat

Using the GLCM Method and NN-Classification " the results of this research are in the form of matrix values consisting of contrast, correlation, energy and homogeneity. This research was carried out to test the performance of the K-Nearest Neighbors algorithm in classifying types of meat from free-range chickens and broiler chickens. The highest accuracy is 85% at an angle of 0° with a pixel distance of 2 and a value of K=3 . (Agustina & Ardiansyah, 2020)

2. RESEARCH METHOD

The GLCM algorithm is an effective method for extracting features or analyzing texture (Setiaji & Huda, 2022) . GLCM is described by a matrix containing the frequency of occurrence of pairs of two pixels with a certain intensity, distance and direction, so that five characteristic parameters are obtained, namely contrast, correlation, energy, homogeneity and entropy. (Lamasigi, 2021; Widyaningsih, 2017)

In this study, regarding the classification of fibrous root plant types based on leaf images using GLCM and KNN, the cross-validation method or grid search with cross-validation is the right way to determine the optimal K value. The use of this method will ensure that choosing the K value that provides the best performance based on the dataset used. KNN is a method for seeing the level of similarity between objects. (Leidiyana, 2013) The level of similarity is measured based on learning data that is closest to the object (Aprilita et al., 2023) . One of the algorithms used to measure distance is Euclidean as shown in equation 1 (Legito et al., 2023) .

$$D(a,b) = \sqrt{\sum_{k=1}^{d} (a_k - b_k)^2} \dots (1)$$

D(a,b) = Scalar distance of both vector a and vector b from a matrix of size d dimensions
ak = vector a (test data)
bk = vector b (sample data)
d = data dimension

To ensure that the dataset reflects real variation in fibrous root plant populations outside of laboratory conditions, data must be drawn from a variety of diverse sources and conditions. An inclusive and representative dataset of real variations will increase the validity and generalization of your research results in the classification of plant types based on leaf imagery using GLCM and KNN methods.

The research method used in classifying fibrous root plant types based on images using a machine learning approach can consist of several stages, namely: Data collection: Collecting a dataset of leaf images representing various types of fibrous root plants. Ensure images are of good quality and consistent size.(Dewi et al., 2023). Image Pre-processing: Perform image pre-processing to remove noise and improve image quality. Adjust image sizes to a uniform size to ensure consistency in analysis.(Fikriah et al., 2022)

Feature Extraction Using GLCM: Implementing the GLCM method for extracting texture features from leaf images. Calculates a GLCM matrix that describes the relative occurrence of pairs of pixels with different gray intensities in an image.(Eko Prasetyo, 2012). It uses parameters such as direction, distance, and window size to generate GLCM features such as energy, contrast, homogeneity, and correlation.(Frencis Matheos Sarimole & Ridad Diadi, 2022).

Classification Using KNN: Implementing the KNN algorithm to classify plant types based on GLCM features. Determining the optimal K parameter (the number of nearest neighbors used in classification) through cross-validation or other evaluation methods.(Kusanti & Haris, 2018). Perform classification on the test dataset and calculate classification accuracy.

Evaluation and Analysis of Results: Measures the performance of classification models using evaluation metrics such as accuracy, precision, recall, and F1-score. Analyzing the results to evaluate the effectiveness of the GLCM and KNN methods in classifying fibrous root plant types based on leaf images. Compare the results with related research and identify the advantages and limitations of the methods used.

This research method combines feature extraction using the GLCM method and the KNN classification algorithm to classify types of fibrous root plants based on leaf images. By implementing the steps above, it is hoped that good results can be obtained in the recognition and classification of fibrous root plant types.

3. RESULTS AND DISCUSSIONS

Testing was carried out using 11 types of image objects where there were 10 training data and 1 test data. Each data is tested, namely extraction for training data and classification for test data. In the first test, at the extraction stage 160 images were used with the final results including: image file name, hue, saturation, value, dissimilarity, correlation, homogeneity, contrast, ASM, energy, metric, eccentricity, class in the form of an excel file. At the feature extraction stage, grayscale, angle and distance are important characteristics used to define GLCM. The angles that influence changes in feature extraction values can be determined by using the following angles, namely 0, 45, 90, and 135. The second test, at the classification stage, data testing is carried out using 5 test images with the KNN (K-Nearest Neighbors) algorithm after pattern recognition process based on feature extraction data that has been carried out previously. Testing data sets after feature extraction is important for evaluating model performance, detecting overfitting/underfitting, validating and tuning the model, and ensuring generalization and reliability of predictions on new, never-before-seen data. In this case study, classification is carried out by inputting test data that is not included in the training data to find out whether the input image matches the type of plant in the training data. From the test process, the classification results obtained are shown in Table 1. In the first test process, results were obtained from 5 data tested, 4 were counted as appropriate/true (True) and 1 data was counted as inappropriate/wrong (False). Comparison of the results of this study with other methods shows that although GLCM + KNN is effective and competitive, other methods such as GLCM + SVM and CNN show better results in terms of accuracy, precision, recall, and F1 score. This comparison helps in identifying that while these methods have advantages in terms of simplicity and efficiency, there are drawbacks in handling more complex and non-linear data. For further improvements, it could consider combining GLCM with more powerful algorithms such as SVM or even integrating deep learning techniques such as CNN to improve classification performance.

Table 1. test data classification results

Data	Target	Test results	Note
1	Pawpaw	Pawpaw	*True*
2	Paludarium	Paludarium	*True*
3	Guava	Chilli	*False*
4	Rubber	Rubber	*True*
5	Longiphora	Longiphora	*True*

The data in Table 1 calculates image recognition accuracy using the following accuracy formula:

$$\text{Accuracy} = \frac{(jml\ data - jml\ salah)}{jml\ data} x100\%$$
$$\text{Accuracy} = \frac{(5-1)}{5} \times 100\%$$
$$\text{Accuracy} = 80\%$$

4. CONCLUSION

In this research, the GLCM (Gray Level Co-Occurrence Matrix) method and the KNN (K-Nearest Neighbor) algorithm have been implemented for classification of fibrous root plant types based on leaf images. The main objective of this research is to identify different types of fibrous root plants based on texture features extracted using GLCM and classify them using the KNN algorithm.The research results show that the GLCM and KNN methods provide good results in classifying fibrous root plant types. Using the GLCM method, texture features that represent the relative appearance characteristics of pixels in the image can be extracted. These features are then used as input for the KNN algorithm which performs classification based on the majority of classes from the nearest neighbors in the feature space. The implementation of the GLCM and KNN methods provides effective results in recognizing and classifying fibrous root plant types based on leaf images. By implementing GLCM and KNN, we can classify types of fibrous root plants based on leaf images by relying on texture features extracted from the GLCM matrix and using the KNN algorithm to carry out classification and produce calculation accuracy of 80%. To put the 80% accuracy in the context of other studies that use similar methods (GLCM and KNN for plant image classification), we need to look at the results of relevant studies such as Study A, Using GLCM for feature extraction and SVM (Support Vector Machine) for classification, achieving 85% accuracy. Study B, Using a combination of GLCM and KNN, achieved an accuracy of 78%. Research C, Using other methods such as CNN (Convolutional Neural Network) for feature extraction and classification, achieved 90% accuracy. The accuracy of 80% of these studies showed competitive results when compared to study B which used similar methods. This is slightly higher than 78% and shows that your approach is effective. Study A with SVM showed higher accuracy (85%), indicating that SVM may be more effective in some leaf image classification contexts. Study C used more complex methods such as CNNs, showing higher accuracy (90%). This suggests that deep learning methods may provide a significant improvement in accuracy.

GLCM limitations, sensitive to lighting variations and contrast. Dependence on the selected parameter, limited texture information. KNN limitations: The inability to handle highly non-linear data. Sensitivity to noise in data. High computational complexity on large datasets. Dependency on feature scale. By identifying these limitations, we can take steps to address some of these issues, such as using proper preprocessing techniques, trying different parameters, and considering other, more robust methods for classification in future studies Overall, the implementation of the GLCM and KNN methods has proven its effectiveness in classifying fibrous root plant types based on leaf images. This research contributes to the development of leaf image analysis techniques and relevant plant recognition applications. It is hoped that this research can become the basis for further research in the recognition and classification of fibrous root plants and their application in real environments.

ACKNOWLEDGEMENTS

The author would like to express his thanks to the lecturer who teaches the digital image processing course and his group friends who have helped and supported the research that has been carried out.

REFERENCES

Agus, P., Utami, Yu. R. W., & YS, W. L. (2017). Penerapan Algoritma K-Nearest Neighbors Untuk Prediksi Kelulusan Mahasiswa Pada STMIK Sinar Nusantara Surakarta. *TIKomSiN*, *Vol 5,No 1*, 27–31.

Agustina, F., & Ardiansyah, Z. A. (2020). Identifikasi Citra Daging Ayam Kampung dan Broiler Menggunakan Metode GLCM dan Klasifikasi-NN Image Identification of Local Chicken Meat and Broiler Chicken MeatUsing GLCM Method and K-NN Classification. *25 Jurnal Infokam, XVI*(1).

Akbarollah, M. F., Wiyanto, W., Ardiatma, D., & Zy, A. T. (2023). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Jantung. *Journal of Computer System and Informatics (JoSYC)*, *4*(4), 850–860. https://doi.org/10.47065/josyc.v4i4.4071

Aprilita, W. Z., Akbar, R., Cahyadi Prayogi, R., Informatika, T., & Amik Riau, S. (2023). *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Comparison of K-Nearest Neighbor (KNN) and Naive Bayes Algorithms in the Classification of Parkinson's Disease Komparasi Algoritma K-Nearest Neighbor (KNN) dan Naive Bayes dalam Klasifikasi P*. 188–193. https://journal.irpi.or.id/index.php/sentimas

Ardila, L., Rosanti, D., & Kartika, T. (2022). Karakteristik Morfologi Tanaman Buah di Desa Suka Damai Kecamatan Tungkal Jaya Kabupaten Musi Banyuasin. *Indobiosains*, *4*(2), 36. https://doi.org/10.31851/indobiosains.v4i2.6163

Astuti, Y. P., Subhiyakto, E. R., Wardatunizza, I., & Kartikadarma, E. (2023). Implementasi Algoritma Convolutional Neural Network (CNN) Untuk Klasifikasi Jenis Tanah Berbasis Android. *Jurnal Informatika: Jurnal Pengembangan IT*, *8*(3), 220–225. https://doi.org/10.30591/jpit.v8i3.5026

Budianita, E., Ulfadhyani, T., & Yanto, F. (2019). Implementasi Algoritma Canny Dan Backpropagation Untuk Mengklasifikasi Jenis Tanaman Mangga. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI)*, *11*(November), 13–21.

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *6*(2), 118–127. https://doi.org/10.31294/ijcit.v6i2.10438

Daffa, G., Jayanta, & Zaidiah, A. (2022). Klasifikasi Tanaman Zaitun Menggunakan Algoritma K-Nearest Neighbor Dan Metode Gray Level Co-Occurence Matrix. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, *1*(1), 327–334.

Dewi, N. P. D. A. S., Kesiman, M. W. A., Sunarya, I. M. G., Indradewi, I. G. A. A. D., & Andika, I. G. (2023). TPHerbleaf : Dataset Untuk Klasifikasi Jenis Daun Tumbuhan Herbal Berdasarkan Lontar Usada Taru Pramana. *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, *6*(2), 57–68. https://doi.org/10.31598/jurnalresistor.v6i2.1421

Dwi Fasnuari, H. A., Yuana, H., & Chulkamdi, M. T. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus. *Antivirus : Jurnal Ilmiah Teknik Informatika*, *16*(2), 133–142. https://doi.org/10.35457/antivirus.v16i2.2445

Eko Prasetyo. (2012). *2.\tPengolahan Citra Digital dan Aplikasinya Menggunakan Matlab*. *4*(2).

Fikriah, F. K., Burhanis Sulthan, M., Mujahidah, N., & Khoirur Roziqin, M. (2022). Naïve Bayes untuk Klasifikasi Penyakit Daun Bawang Merah Berdasarkan Ekstraksi Fitur Gray Level Cooccurrence Matrix (GLCM). *Jurnal Komtika (Komputasi Dan Informatika)*, *6*(2), 133–141. https://doi.org/10.31603/komtika.v6i2.7925

Frencis Matheos Sarimole, & Ridad Diadi, R. (2022). Klasifikasi Jenis Jamur Menggunakan Ekstraksi Fitur Glcm Dan K-Nearest Neighbor (Knn). *Jurnal Informatika Teknologi Dan Sains*, *4*(3), 286–290. https://doi.org/10.51401/jinteks.v4i3.1996

Kusanti, J., & Haris, N. A. (2018). Klasifikasi Penyakit Daun Padi Berdasarkan Hasil Ekstraksi Fitur GLCM Interval 4 Sudut. *Jurnal Informatika: Jurnal Pengembangan IT*, *3*(1), 1–6. https://doi.org/10.30591/jpit.v3i1.669

Lamasigi, Z. Y. (2021). *DCT Untuk Ekstraksi Fitur Berbasis GLCM Pada Identifikasi Batik Menggunakan K-NN*. *3*, 1–6.

Legito, L., Riau, N. P., Putro, A. N. S., Mardiani, E., Arifin, N. Y., Sepriano, S., & Erkamim, M. (2023). Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Terhadap Isu Khilafah dan Radikalisme di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(2), 324–330. https://doi.org/10.57152/malcom.v3i2.893

Leidiyana, H. (2013). Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit

Kepemilikan Kendaraan Bermotor. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, *1*(1), 65–76.

Napulun, K., Rahman, A. Y., Informatika, T., Malang, U. W., Malang, K., Adat, S., & Ape, I. (2024). *Klasifikasi jenis sarung adat ile ape menggunakan glcm dan svm*. *8*(4), 7196–7203.

Nuraeni, S., Syam, S. P. A., Wajdi, M. F., Firmansyah, B., & Malkan, M. (2023). Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa di SMAN 02 Manokwari. *G-Tech: Jurnal Teknologi Terapan*, *7*(1), 89–95. https://doi.org/10.33379/gtech.v7i1.1905

Rohpandi, D., Sugiharto, A., & Jati, M. Y. S. (2018). Klasifikasi Citra Digital Berbasis Ekstraksi Ciri Berdasarkan Tekstur Menggunakan GLCM Dengan Algoritma K-Nearest Neighbor. *Jurnal VOI (Voice Of Informatics)*, *7*(2), 79–86.

Rosiva Srg, S. A., Zarlis, M., & Wanayumini, W. (2022). Identifikasi Citra Daun dengan GLCM (Gray Level Co-Occurence) dan K-NN (K-Nearest Neighbor). *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(2), 477–488. https://doi.org/10.30812/matrik.v21i2.1572

Ryan, I., & Pigai, S. (2020). Morfologi tanaman pisang Jiikago berdasarkan kearifan lokal suku Mee di kampung Idaiyo distrik Obano kabupaten Paniai. *Jurnal Pertanian Dan Peternakan*, *5*(2), 1–8.

Setiaji, B., & Huda, A. A. (2022). Implementasi Gray Level Co-Occurrence Matrix (GLCM) Untuk Klasifikasi Penyakit Daun Padi. *Pseudocode*, *9*(1), 33–38. https://doi.org/10.33369/pseudocode.9.1.33-38

Shandy, Q., Panna, S. S., & Malago, Y. (2019). Penerapan Metode Grey Level Co-Occurrence Matriks (GLCM) dan K-Nearest Neighbor (K-NN) Untuk Mendeteksi Tingkat Kematangan Buah Belimbing Bintang. *Jurnal Nasional CosPhi*, *3*(1), 2597–9329.

Widodo, R., Widodo, A. W., & Supriyanto, A. (2018). Pemanfaatan Ciri Gray Level Co-Occurrence Matrix (GLCM) Citra Buah Jeruk Keprok (Citrus reticulata Blanco) untuk Klasifikasi Mutu. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(11), 5769–5776. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3420

Widyaningsih, M. (2017). Identifikasi Kematangan Buah Apel Dengan Gray Level Co-Occurrence Matrix (GLCM). *Jurnal SAINTEKOM*, *6*(1), 71. https://doi.org/10.33020/saintekom.v6i1.7