International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 51 No. 2 July 2025, pp. 333-338

Implementation K-Nearest Neighbor (KNN) Algorithm For Identifying Pattern Differences Between Betel Leaves And Pepper Leaves

[1]Zulham Sitorus, [2]Ade May Luky Harefa, [3]Indra Marto Silaban, [4]Maimuddin Noor, [5]Ica safrila, [6]Donas Putra

Master of Information Technology
Postgraduate Program, Pembangunan Panca Budi University
Medan, Indonesia
Corresponding Author: Ade May Luky Harefa. E-mail: ademayluky@gmail.com

Abstract–This research aims to identify the pattern differences between betel leaf (Piper betle) and black pepper leaf (Piper nigrum) using the K-Nearest Neighbor (KNN) algorithm. The method is applied to classify the two types of leaves based on image features such as texture, shape, and color. A total of 200 leaf images were used, divided into 70% training data and 30% testing data. Feature extraction was conducted to obtain the most relevant characteristics from each image. The classification process was performed with various K values, and the highest accuracy was achieved when K = 3. The results showed an accuracy of 91.5%, with a precision of 90.8%, recall of 92.3%, and F1-score of 91.5%. These findings indicate that the KNN algorithm is effective in distinguishing betel and pepper leaves using digital image processing. Texture and color features contributed the most to the classification performance. This study shows the potential of KNN-based leaf pattern recognition for practical applications in agriculture, herbal identification, and plant classification systems.

Key words: K-Nearest Neighbor; Leaf Identification; Texture Features; Classification.

I. INTRODUCTION

Indonesia has a very rich biodiversity, including various types of plants with high economic and cultural value [1]. Among the plants, betel leaves (Piper betle) and pepper leaves (Piper nigrum) are two types of plants from the Piperaceae family that are often found in plantations and in the yards of Indonesian people's houses[2][3]. These two types of plants have similar leaf morphological shapes that are often difficult to distinguish visually, especially by the general public who lack a clear understanding of the specific differences between the two. In fact, both have different functions and benefits in the health, culinary, and herbal industries[4] [5].

As technology has developed, many efforts have been made to make it easier to identify plants using computational approaches [6]. One of the popular and effective methods for object classification is the algorithm K-Nearest Neighbor (KNN) [7]. This algorithm is well-known in the field of pattern recognition and Machine Learning Because of its good ability to group data based on the proximity of the distance between the features that the data has [8] [9]. The application of the KNN algorithm in plant leaf identification is interesting because it allows automatic classification based on leaf image features, such as shape, color, and texture. This automatic identification can help the public and industry players to classify leaves with high accuracy, reduce misidentification, and support herbal plant-based business processes.

Some previous studies have implemented the kNN method in leaf identification or plant classification. Research [10] This study aims to classify three types of bananas in large quantities using 150 photo data, with 120 data as test data that achieves an accuracy of 99.1667%. In addition, research conducted by [11] This study developed a corn variety classification system using the k-NN method on 250 corn images, resulting in the highest accuracy of 93.24% at k=3 value with 91.89% recall, 94.44% accuracy, and f-measure 0.9315. However, research that specifically uses KNN to distinguish betel leaves and pepper leaves is still rare. In fact, these two plants have high economic value and are important to be identified appropriately for industrial and research purposes.

Based on the background of the existing problems, this study is designed to answer how the implementation of the K-Nearest Neighbor (KNN) algorithm can be used to accurately distinguish betel leaf and pepper leaf patterns. With the challenge of identifying these two types of leaves that have morphological similarities, this study aims to explore the ability of the KNN method in classifying the two types of leaves based on image features. In addition, this study also aims to evaluate the level of accuracy that can be achieved by the KNN algorithm in the leaf pattern identification process, so that it is expected to be able to provide practical solutions for the public and industry players in classifying betel leaves and pepper leaves quickly and accurately.

II. PURPOSE AND METHODS

The researchers used a quantitative approach with experiments to prove the effectiveness of the K-Nearest Neighbor (K-NN) algorithm in identifying differences in betel leaf and pepper leaf patterns based on digital imagery. The data used in this study included 200 leaf images consisting of 100 betel leaf images and 100 pepper leaf images. These images were obtained through shooting using a digital camera on a white background to reduce *noise*. To ensure data quality, all images are converted to a standard resolution of 256x256 pixels to standardize dimensions and reduce computational complexity. The dataset is then divided into two parts, namely 70% training data and 30% test data, according to common practices in machine learning.

The preprocessing step is carried out to prepare the image data so that it is ready to be used in the feature extraction process [12]. Images that are RGB formatted are first converted to *grayscale* to simplify texture analysis. Next, image segmentation is carried out using the *Thresholding Otsu*, which automatically separates the leaf object from the background based on the difference in pixel intensity [13]. In addition, the data is normalized using the *min-max scaling* to ensure all features are in the same range, thus avoiding bias in the classification process. Feature extraction aims to extract the main characteristics of the leaf image. The extracted features consist of texture, shape, and color [14]. Texture features are obtained using *Gray Level Co-occurrence Matrix* (GLCM) with values such as contrast, correlation, energy, and homogeneity. Shape features are calculated through parameters such as area, perimeter, and aspect ratio to describe the dimensions and proportions of the leaves. Finally, color features are extracted from the color histogram in the HSV (Hue, Saturation, Value) color space to analyze the typical color distribution of each leaf type [15].

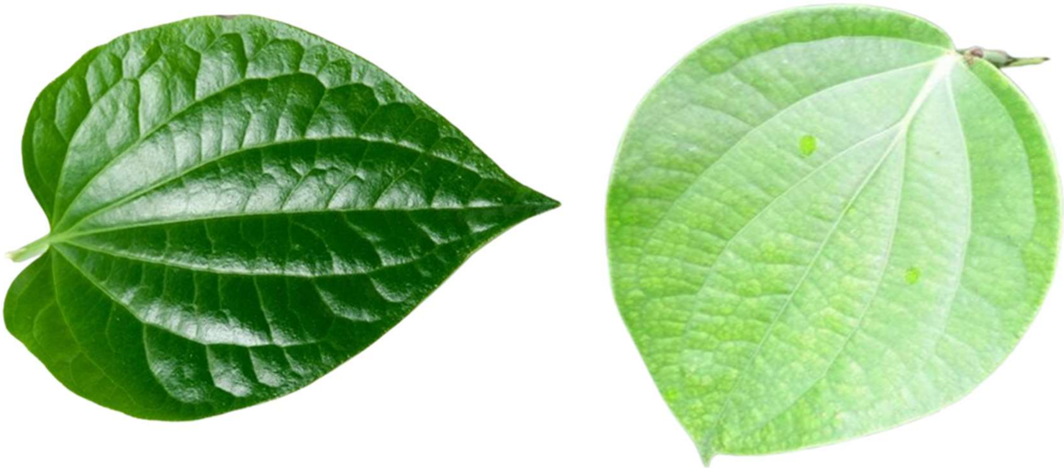
The K-Nearest Neighbor (KNN) method is used to classify data based on its proximity to known data [16]. The K values tested in this study were 3, 5, and 7. The distance between samples was calculated using Euclidean formulas to determine the proximity between the data. The KNN model was trained using training data, then tested with test data to predict labels. The prediction is determined by the majority of labels from the nearest neighbor K [17]. This implementation is done using MATLAB, which provides built-in functions to build KNN models efficiently.

Evaluation of model performance is carried out using accuracy, precision, recall, and F1-score metrics. Accuracy measures the percentage of correct predictions against the total test data. Precision calculates the proportion of positive predictions that are correct, while recall assesses the model's ability to find all available positive data. The F1-score, which is a harmonious average of precision and recall, provides an idea of the balance of the model's performance. Evaluation values for various K parameters showed that K=3 scores resulted in the highest accuracy of 91.5%, with precision, recall, and F1-scores also being optimal. The implementation results are visualized through scatter graphs to show the distribution of betel leaf and pepper leaf feature data. These graphs help in understanding the visual patterns of the data. In addition, the results of the model evaluation at various K-values are visualized in the form of line graphs, which show trends in accuracy, precision, recall, and F1-score. This visualization shows that a smaller K-value, such as K=3, performs better than a larger K-value. The collection of digital image data of betel leaf and pepper

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 51 No. 2 July 2025, pp. 333-338

leaf patterns uses the Xiaomi Redmi 11 Pro mobile phone camera with ISO 800 setting and uses a tripod to maintain the same distance and position.



(a) (b)

Fig 1. Digital image data of (a) betel leaf and (b) pepper leaf

III. RESULTS AND DISCUSSIONS

This study used the K-Nearest Neighbor (KNN) algorithm to identify the difference in patterns between betel leaves and pepper leaves based on texture, shape, and color features. The K-values tested were 3, 5, and 7, with 70% training data and 30% test data. The results of the model performance evaluation are shown in Table 1. Based on the accuracy, precision, recall, and F1-score metrics, the K=3 score gave the best results with an accuracy of 91.5%, followed by K=5 with an accuracy of 89.2%, and K=7 with an accuracy of 87.0%.

Table 1. KNN Model Performance Evaluation

K Value	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
3	91.5	90.8	92.3	91.5
5	89.2	88.5	89.7	89.1
7	87.0	86.2	87.5	86.8

Graphs of model performance at various K-values are visualized in Figure 3. This graph shows a comparison of accuracy, precision, recall, and F1-score for grades K = 3, 5, and 7.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 51 No. 2 July 2025, pp. 333-338

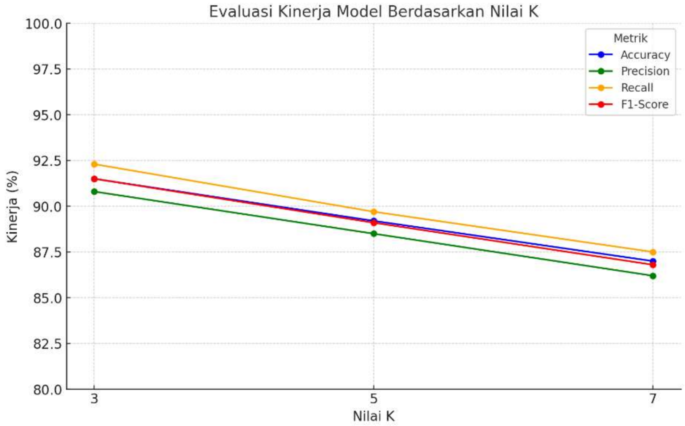


Fig 2. Graph of model performance on various K-values

Figure 2 illustrates the performance evaluation of the KNN model across different K values (K=3, 5, and 7) based on four metrics: Accuracy, Precision, Recall, and F1-Score. The graph shows a clear downward trend in all metrics as the K value increases. The highest performance is achieved at K=3, with all metrics above 90%, indicating optimal classification at this value. As K increases to 5 and 7, the model becomes less sensitive to local variations, resulting in a decline in accuracy and overall performance. This trend highlights the importance of selecting an appropriate K value to achieve optimal classification results.

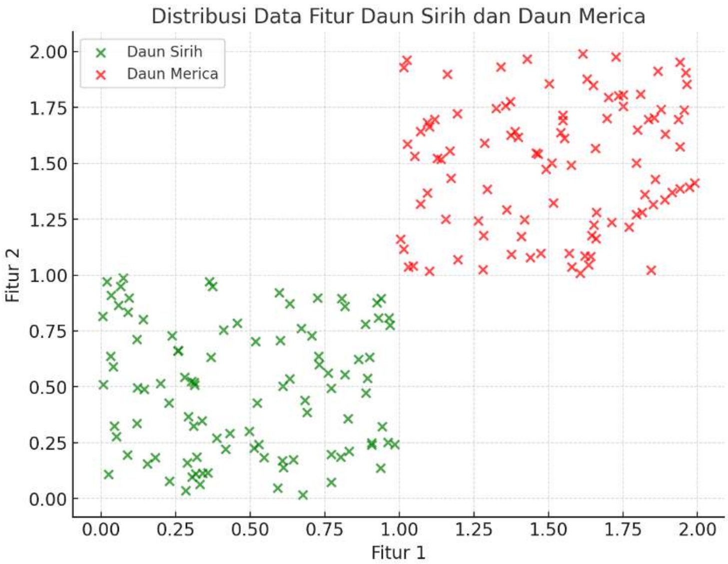


Fig 3. Scatter Plot

Figure 3 above illustrates the feature distribution of betel leaves and pepper leaves based on two main features (Feature 1 and Feature 2). Green points represent betel leaf data, while red points represent pepper leaf data. The two groups form distinct clusters, indicating significant differences in feature characteristics between the two types of leaves. This clear separation provides a strong foundation for the K-Nearest Neighbor (KNN) algorithm to accurately classify the data based on feature proximity.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 51 No. 2 July 2025, pp. 333-338

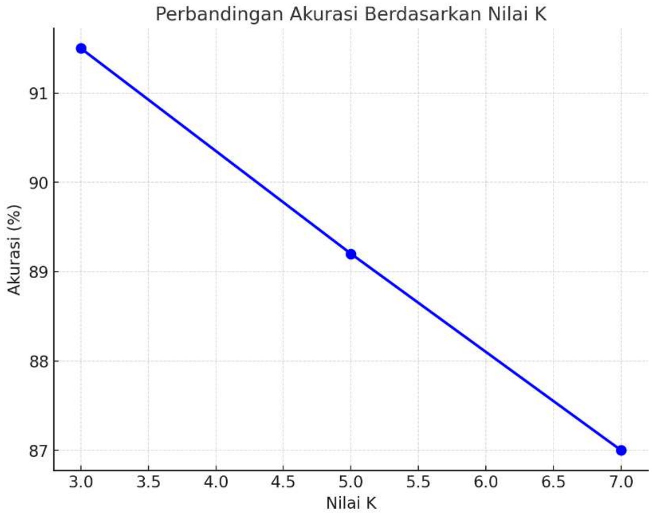


Fig 4. Accuracy Graph Based on K Value

Figure 4 above shows the comparison of classification accuracy across different K values in the K-Nearest Neighbor(KNN) algorithm. It can be observed that the highest accuracy, approximately 91.5%, is achieved when K=3. As the K value increases to 5 and 7, the accuracy decreases progressively. This indicates that a smaller K value is more effective in capturing local patterns in the data, while larger K values tend to include more irrelevant neighbors, leading to less accurate classification results.

The results showed that a smaller K value, specifically K = 3, provided the best performance in classifying betel leaf and pepper leaf patterns. This outcome is attributed to the lower K's sensitivity to local variations, allowing the algorithm to focus on the most relevant neighboring data points. In contrast, higher K values such as K = 7 resulted in decreased accuracy, as the model became more generalized and included less relevant neighbors in the classification process, thereby reducing its ability to differentiate between the leaf types effectively.

Texture features contributed significantly to distinguishing the leaf patterns, particularly through contrast and homogeneity values, which reflect surface differences between betel and pepper leaves. Additionally, shape features such as aspect ratio, and color features in the HSV color space, also played a vital role in classification by highlighting distinct differences in form and pigmentation. Overall, the KNN algorithm demonstrated high effectiveness in identifying leaf patterns, especially when supported by appropriate preprocessing and feature extraction techniques. However, the model's performance could be further enhanced by increasing the size of the training dataset or integrating alternative classification algorithms for comparison and hybrid approaches.

IV. CONCLUSION

This study successfully implemented the K-Nearest Neighbor (KNN) algorithm to identify the difference in patterns between betel leaves and pepper leaves. Based on the model performance evaluation, the K=3 value gave the best results with an accuracy of 91.5%, precision of 90.8%, recall of 92.3%, and an F1-score of 91.5%. These results show that the KNN algorithm is effective in distinguishing the patterns of the two types of leaves using a combination of texture, shape, and color features. This method shows good potential for application in the recognition of other image-based patterns. However, future research may consider increasing the number of datasets, the use of data augmentation techniques, or comparisons with other classification algorithms to obtain better performance.

REFERENCES

[1] M. P. Akmal, *Keanekaragaman Hayati (Biodiversitas): Modul Ajar IPA Biologi Kelas X, Fase E*. Akmal’s Library, 2022.

[2] A. F. Al Mutahhar, “IDENTIFIKASI KARAKTERISTIK FISIK MINYAK ATSIRI PADA LADA HITAM (Piper nigrum

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 51 No. 2 July 2025, pp. 333-338

L.) DAN SIRIH HIJAU (Piper betle L.).” Akademi Farmasi Surabaya, 2021.

[3] L. Yuliana, “Studi morfologi genus Piper dan variasinya,” *Biocaster J. Kaji. Biol.*, vol. 3, no. 1, pp. 11–19, 2023.

[4] A. N. B. R. Tarigan, E. Nurtjahya, and T. A. Aththorick, “Obat Tradisional Peningkat Imunitas Tubuh Etnis Karo, Kecamatan Sinembah Tanjung Muda Hulu,” *Biocelebes*, vol. 16, no. 1, pp. 21–29, 2022.

[5] M. N. Ramdhani, A. Firdaus, H. F. Reine, and A. Supriyatna, “Analisis Morfo-Anatomi Daun Sirih dari Famili Piperaceae dan Araceae di Kampung Warung Peuteuy, Kecamatan Cicalengka,” *Polyg. J. Ilmu Komput. dan Ilmu Pengetah. Alam*, vol. 2, no. 4, pp. 70–82, 2024.

[6] E. Erwin *et al.*, *Pengantar & Penerapan Internet Of Things: Konsep Dasar & Penerapan IoT di berbagai Sektor*. PT. Sonpedia Publishing Indonesia, 2023.

[7] J. Saputra, Y. Sa’adati, V. Y. P. Ardhana, and M. Afriansyah, “Klasifikasi Kematangan Buah Alpukat Mentega Menggunakan Metode K-Nearest Neighbor Berdasarkan Warna Kulit Buah,” *Resolusi Rekayasa Tek. Inform. dan Inf.*, vol. 3, no. 5, pp. 196–203, 2023.

[8] G. Rahmawati, S. A. Sanmas, E. Nudyawati, N. D. Syaharani, M. R. Annas, and F. Fauzi, “Studi Perbandingan Performa: Prediksi Status Stunting Pada Anak Berdasarkan Data Antropometri Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN),” in *PROSIDING SEMINAR NASIONAL SAINS DATA*, 2024, pp. 782–790.

[9] R. F. Putra *et al.*, *Algoritma Pembelajaran Mesin: Dasar, Teaknik, dan Aplikasi*. PT. Sonpedia Publishing Indonesia, 2024.

[10] S. Sugiyono and M. Ruswandi, “Pemodelan Pengolahan Citra untuk Klasifikasi Jenis Buah Pisang Menggunakan Metode KNN,” *J. Pendidik. dan Konseling*, vol. 4, no. 5, pp. 823–833, 2022.

[11] V. Chandra, J. H. Jaman, and G. Garno, “Identifikasi Varietas Jagung Mutiara Berdasarkan Data Citra Digital Menggunakan Algoritma K-Nearest Neighbor,” *NUANSA Inform.*, vol. 16, no. 2, pp. 1–14, 2022.

[12] N. Saefulloh, J. Indra, R. Rahmat, and A. R. Juwita, “Implementasi Algoritma Convolutional Neural Network (CNN) Untuk Klasifikasi Kecacatan Pada Proses Welding di Perusahaan Manufacturing,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, pp. 387–394, 2024.

[13] R. N. J. S. Intam, A. Raihan, M. Alfajri, A. B. Kaswar, and D. D. Andayani, “Sistem Klasifikasi Jenis Sampah Berdasarkan Kombinasi Fitur Warnac Tekstur Menggunakan Artifical Neural Network Berbasis Pengolahan Citra Digital,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 2, pp. 411–420, 2024.

[14] A. H. S. Hadi and S. Agustin, “PENENTUAN KALORI MAKANAN BERDASARKAN FITUR WARNA DAN BENTUK MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOUR) BERBASIS ANDROID,” *Kohesi J. Sains dan Teknol.*, vol. 4, no. 9, pp. 61–70, 2024.

[15] A. A. Paturrahman, “Analisis pengenalan pola daun berdasarkan fitur canny edge detection dan fitur GLCM menggunakan metode klasifikasi K-Nearest Neighbor (K-NN),” *Publ. Tugas Akhir S-1 PSTI FT-UNRAM*, 2020.

[16] I. Yolanda and H. Fahmi, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT. Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 3, no. 1.1, pp. 9–15, 2020.

[17] R. Rina, P. M. Hasan, N. Ayu, and R. A. Saputra, “KLASIFIKASI KERINGANAN UKT MAHASISWA UHO MENGGUNAKAN K-NEAREST NEIGHBOR (KNN),” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 6, pp. 11939–11945, 2024.