**International Journal of Multilingual Education and Applied Linguistics**
**Vol.1, No.4 November 2024**
e-ISSN :3047-907X; p-ISSN :3047-9061, Page 25-37
DOI: https://doi.org/10.61132/ijmeal.v1i4.113
*Avalable online at:* https://international.aspirasi.or.id/index.php/IJMEAL

# Identification of Traditional Herbal Leaves and Their Benefits Using K-Nearest Neighbors (KNN)

**Nur Rahma Ditta Zahra [1*], Kanaya Sabila Azzahra [2], Nur Iman Nugraha [3], Muhammad Ilham Nurfajri [4], ,Nabil Malik Al Hapid [5], Endang Purnama Giri [6], Gema Parasti Mindara [7]**
[1-7] IPB University, Indonesia

Address: Jl. Kumbang No.14, RT.02/RW.06, Bogor - Jawa Barat 16128
*Corresponding author : rahmaditta12@gmail.com**

***Abstract***. *This study presents a web-based system for identifying traditional herbal leaves using K-Nearest Neighbors (KNN) and image processing techniques focused on analyzing leaf shape and color. The dataset used consists of images of various types of herbal leaves, providing a basis for classification and medicinal benefit information retrieval. The system was tested with multiple leaf samples to assess accuracy, speed, and effectiveness in identifying leaf types based on visual characteristics. Results show that the system can recognize different types of herbal leaves and display information on their medicinal properties in a user-friendly interface..*

***Keywords:*** *Herbal leaves, K-Nearest Neighbors, image processing, feature extraction, medicinal benefits.*

## 1. BACKGROUND

Indonesia, as a country blessed with fertile soil, has abundant flora diversity. Plants in Indonesia not only function as food chain producers, but also have potential as medicinal plants, especially from the leaves. Since ancient times, many plants have been used for medicine, and traditional herbal medicine is one form of treatment that is still practiced today (Sibero & Saleh, 2020).

Although previous generations relied heavily on traditional medicines, the development of the times has changed lifestyles, especially among the younger generation. Unhealthy lifestyles, such as smoking habits, fast food consumption, and lack of physical activity, can trigger various diseases at a young age. Therefore, awareness of the importance of recognizing and utilizing herbal plants as a natural treatment solution is becoming increasingly urgent.

However, identifying the type of herbal leaves is not easy. The simplest way to identify leaves is to look at their shape directly. However, not everyone has the knowledge or skills to distinguish one type of leaf from another, especially because of the similarities between plants (Jamaliah et al., 2017). To overcome this challenge, computer-based technology is needed that can assist in the process of automatic leaf identification.

In line with that, emphasizing the importance of digital image processing technology in recognizing leaves. By using image processing methods, the system can identify the characteristics of objects, namely herbal leaves, more accurately. The process of identifying features in leaf images is very important to distinguish one image from another (Shofrotun et

*Received: Oktober 17, 2024; Revised: Oktober 31, 2024; Accepted: November 08, 2024; Published : November 12, 2024*

al., 2018). Adding that feature extraction such as leaf shape and structure is a necessary initial step. The value of this extraction result is stored as comparative data between the input image and other images (Sibero & Saleh, 2020).

The method often used in measuring the distance between features is Euclidean Distance, and one of the relevant methods for classification based on this distance is K-Nearest Neighbors (KNN). With KNN, the system can group leaves based on the proximity of the extracted features, making it easier to recognize and classify herbal leaves.

With this background, this study aims to design a system that can be accessed via the web that is able to identify traditional herbal leaves using the KNN method and image processing techniques. It is hoped that this system can help the community in recognizing herbal leaves and their benefits effectively and efficiently.

## 2. THEORETICAL STUDY

### Traditional Herbal Leaves

Herbal plants are plants that, based on human experience and observation, are known to contain active compounds that are efficacious in preventing and treating various diseases or supporting certain biological functions. These traditional medicinal plants are often also known as "living pharmacies," where part of the land is used specifically to grow medicinal plants that can be used daily. These natural medicinal plants are widely known to be able to treat various types of health disorders. The main advantage of medicinal plants is their natural nature so they tend to have milder or even minimal side effects compared to synthetic drugs, making them a popular choice among the public.

Leaves from traditional herbal plants play a major role in nature-based medicine. For example, betel leaves, sambiloto, and basil have been used for years to treat various health problems, such as infections, digestive disorders, and to maintain physical fitness. In addition to being a source of medicine, these herbal leaves also play an important role in preserving cultural heritage because knowledge of their benefits is passed down from one generation to the next. By utilizing local herbs, people can obtain more environmentally friendly health solutions and respect the values of local wisdom (Kumontoy, Deeng, & Mulianti, 2023).

### Image Technology

An image is a visual representation or imitation of an object, which allows the object to be viewed in various formats. In a storage system, images can be presented as photos (optical), video signals such as images on a television screen (analog), or digital formats that can be

stored in electronic devices. Digital images themselves are types of images that can be processed by computers. Generally, images are two-dimensional, resulting from the conversion of continuous analog images into discrete forms (Raharja & Harsadi, 2018).

As part of multimedia, digital images or images are sources of visual information created through the digitization process. Digital images are divided into three main types, namely black and white images (monochrome), grayscale images, and full color images or true color (Zebua & Ndruru, 2017). The smallest unit in a digital image is a pixel or Picture Element, which is a small point formed from the meeting of horizontal and vertical lines. The appearance of color in an image depends on the value of each pixel (Sunandar, 2017).

**KNN (*K- Nearest Neural Network*)**

The K-Nearest Neighbors (KNN) method is a classification technique that determines the group of an object, such as wood fiber, by comparing it to similar data around it, which is a number of k nearest neighbors that have been determined. This method classifies new objects into certain categories based on similarity patterns with existing data (Masutani et al., 2012). KNN is an algorithm based on finding similarities in previous data, where training data is prepared and then the value of k is obtained from the closest data using the Euclidean distance calculation. In classification, the main features of the test data are calculated, and objects are retested for unidentified classifications or to verify the accuracy of the classification (Kim et al., 2019).

KNN focuses on data classification by considering the shortest distance between new objects and relevant training data. In the training stage, the algorithm only stores relevant data features for the clustering process. Then, when classifying, the distance between the test data and the training data vector is calculated, and several nearest k values of similar training data are taken as references. The test data is finally classified into groups based on the most common categories among the neighbors. This distance is usually calculated using the Euclidean formula; the larger the distance, the lower the similarity between the test and training data, and conversely, the smaller the distance, the higher the similarity between the two. This method has the advantage of simplicity, although it is less effective if the data is too densely distributed, but it is still widely used because it is easy to understand (Siregar et al., 2019).

The principle of KNN is to measure the distance between the test data and the k nearest neighbors, where k is a positive integer. K is usually chosen as an odd number so that the classification results do not have the same number in the two classes. The following are the general stages of KNN (Heryadi & Wahyono, 2020):

1. Determine the number k for the nearest neighbor.
2. Select the k nearest samples from the existing data.
3. Calculate the frequency of each category that appears.
4. Assign classes to the test data according to the category that appears most often in its neighbors.

## 3. RESEARCH METHODS

This research methodology is designed to develop an efficient traditional herbal leaf identification system using the K-Nearest Neighbors (KNN) method and image processing techniques. This study utilizes a leaf image dataset containing various types of herbal leaves as the basis for classification. Each step in this methodology is important to ensure the accuracy and effectiveness of the system in identifying leaf types and providing information about their properties.

**Dataset Preparation**

In this study, the dataset used is in the form of images with bmp format consisting of 250 images of 10 types of traditional herbal leaves, with each type of leaf having 25 sample images. The types of leaves contained in this dataset include:

**a. Jasmine Flower Leaves**

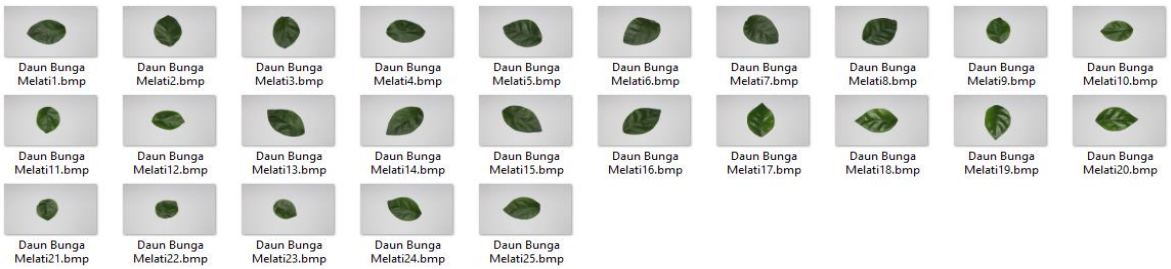



**Figure 1. Jasmine Flower Leaves**

**b. Avocado Leaves**


**Figure 2. Avocado leaves**

### c. Mango Leaves


**Figure 3. Mango leaves**

### d. Balakacida Leaves

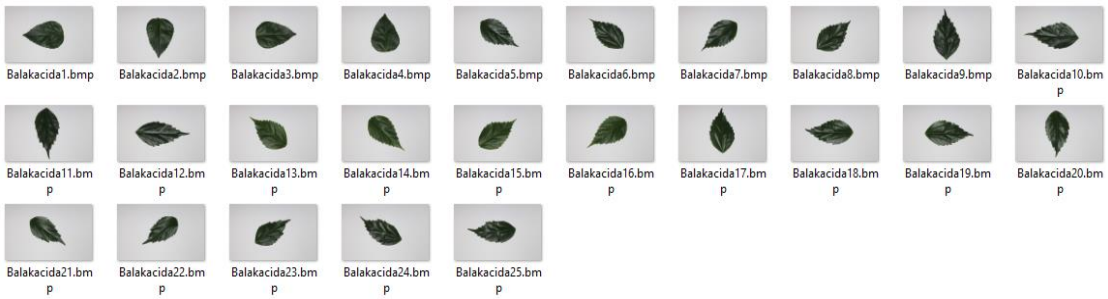



**Figure 4. Balakacida leaves**

### e. Gotu Kola Leaves

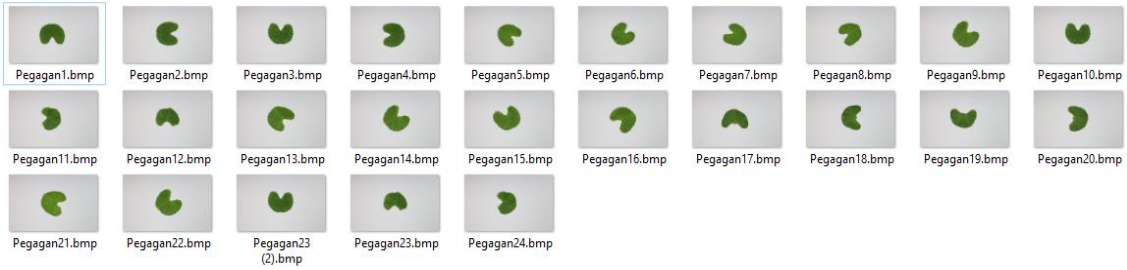



**Figure 5. Gotu Kola Leaves**

### f. Soursop leaf

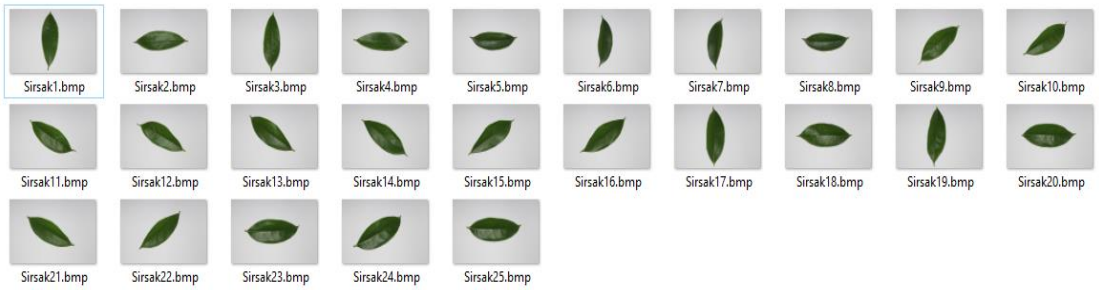



**Figure 6. Soursop leaf**

**g. Guava Leaves**

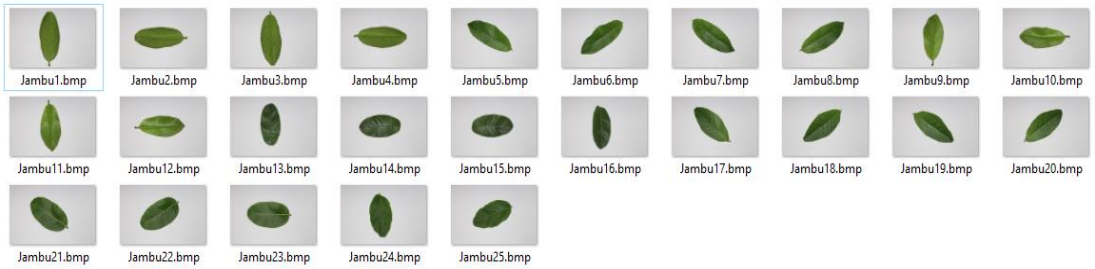



**Figure 7. Guava Leaves**

**h. Teak Leaf**

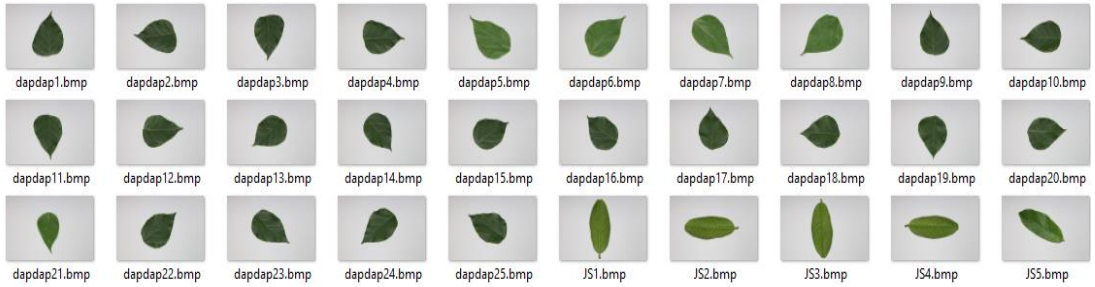



**Figure 8. Teak Leaf**

**i. Bay leaf**

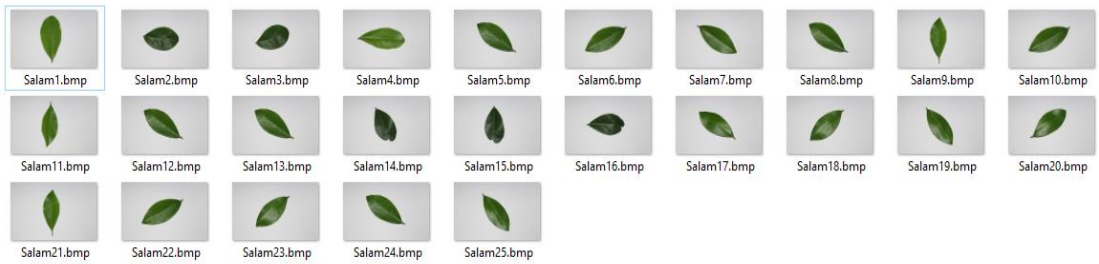



**Figure 9. Bay leaf**

**j. Ceremai Leaves**

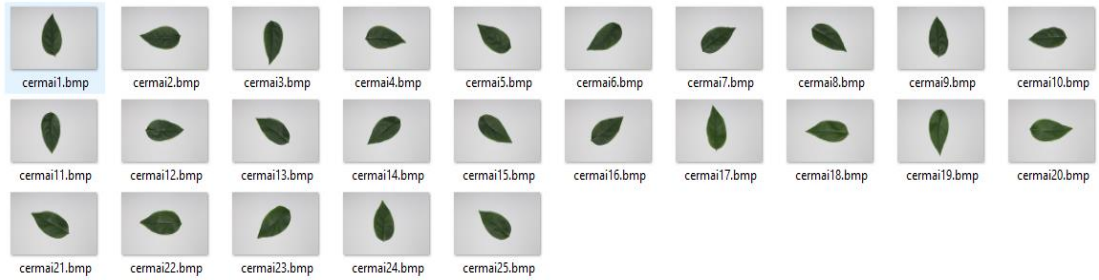



**Figure 10. Ceremai Leaves**

**k. Betel leaf**

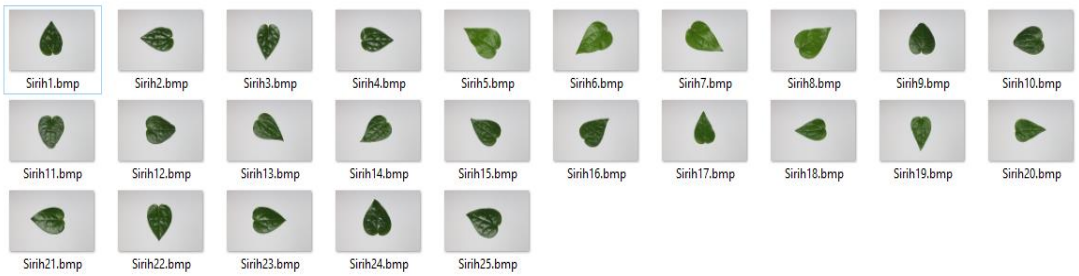



**Figure 11. Betel leaf**

***Preprocessing* Data**

The preprocessing stage is an important step that aims to improve the quality of the image before further analysis. This process includes several key steps:

- Reading Image: Using OpenCV to read leaf image from specified path.
- Convert to Grayscale: Color image is converted to grayscale for easy processing.
- Thresholding: Using thresholding technique to create binary image, separating object (leaf) from background.
- Morphological Operation: Dilation and erosion process is applied to binary image to clean noise and clarify object shape.

**Visual Feature Extraction**

After preprocessing, important features are extracted from the leaf images to obtain relevant information to support classification. This feature extraction includes:

- Color Dominance: Using KMeans, the dominant color of the leaf image is calculated to provide additional information in the classification.
- Gray Level Co-occurrence Matrix (GLCM): Texture features are calculated from the grayscale image to obtain characteristics such as dissimilarity, correlation, homogeneity, contrast, ASM, and energy.
- Shape Analysis: Using regionprops to calculate shape features such as area, perimeter, and eccentricity of the detected objects.

**KNN Classification**

The K-Nearest Neighbors (KNN) method is an algorithmic method in nonparametric machine learning that is used for classification. This algorithm works by identifying the "k" closest data points (neighbors) of the data being tested based on distance metrics, such as Euclidean distance. To create an evaluation method using KNN, several steps are taken

including data preparation, data separation into training and testing sets, KNN implementation, and model performance evaluation (Fajri, Septian, & Sanjaya, 2020).

**System Implementation**

The final stage in this research is the development of a web-based system interface using HTML and CSS for the display, and Flask as the backend framework. This web interface allows users to upload leaf images and receive identification results along with information on the efficacy of the leaves. Identification results are presented in a format that is easy for users to understand, with additional relevant information related to the benefits of herbs in medicine.

## 4. RESULTS AND DISCUSSION

In this section, the results of the development and implementation of a web-based traditional herbal leaf identification system are presented and analyzed. The system was tested with various leaf image samples to ensure accuracy, speed, and ability to identify leaf types based on their visual characteristics, including leaf shape and color. The following are the results and discussions of each stage carried out.

**Perancangan Sistem**

A flowchart is a visual representation that depicts a series or steps in a process sequentially using standard symbols, such as ovals for start and finish, rectangles for processes, parallelograms for input/output, and diamonds for decisions. Flowcharts are used to simplify and visualize complex processes, facilitate communication, and assist in the analysis and improvement of workflows in a variety of fields, such as system design.

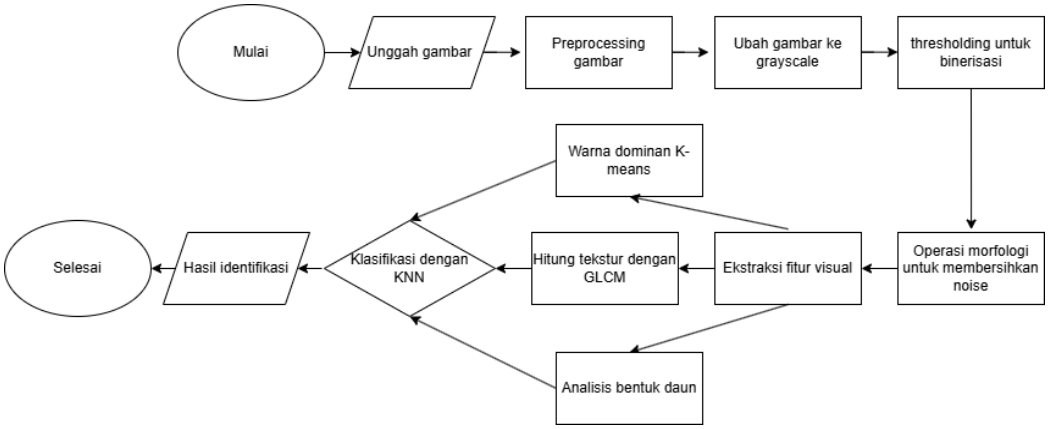
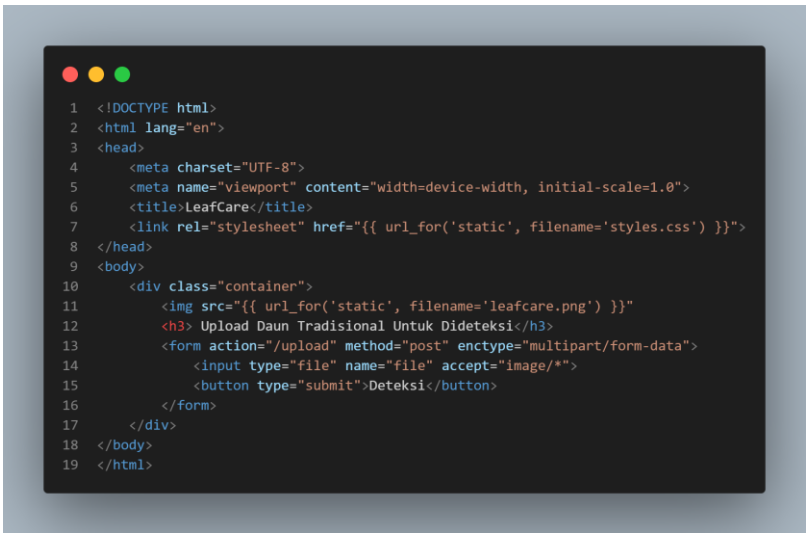


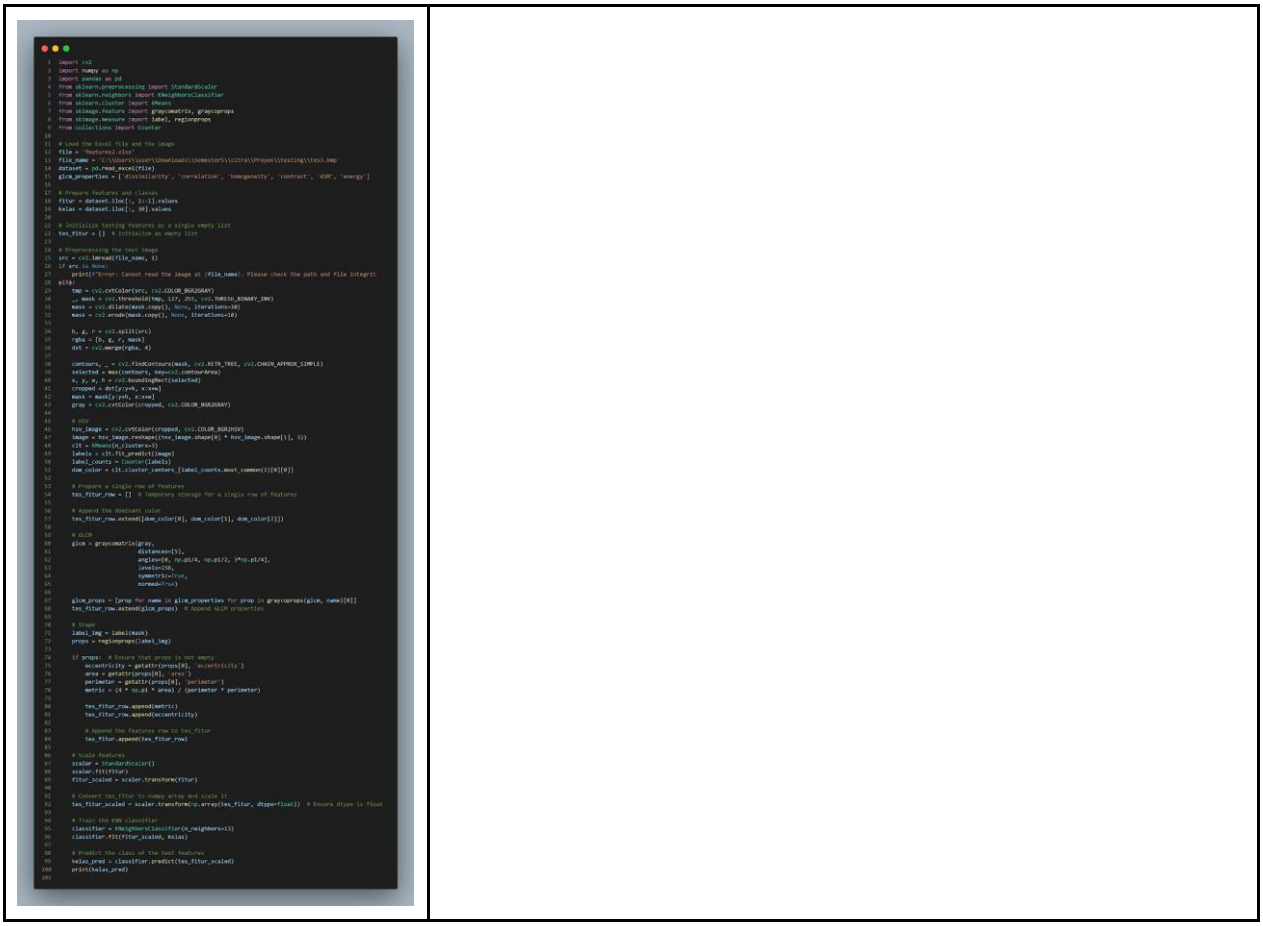

**Figure 12. System Design**

**System Coding and Implementation**

This system is developed using Python programming language with image processing libraries like Flask for web backend. Here is an example of code used for extracting shape and color features from leaf images:

**Table 1. System Code**

| result.html | index.html |
|---|---|
|  |  |
| main.py | app.py |
|  |  |
| knn.py | |



The system is implemented in the form of a web-based interface that allows users to upload leaf images. Once the image is uploaded, the system will process the image to produce identification output, including information about the type of leaf and its benefits. An example of the identification results displayed on the interface is as follows:

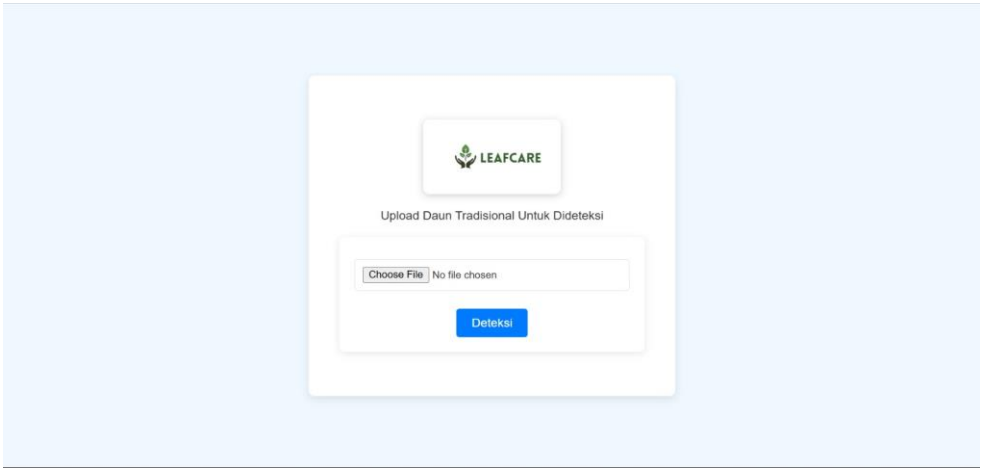



**Figure 13. System View**

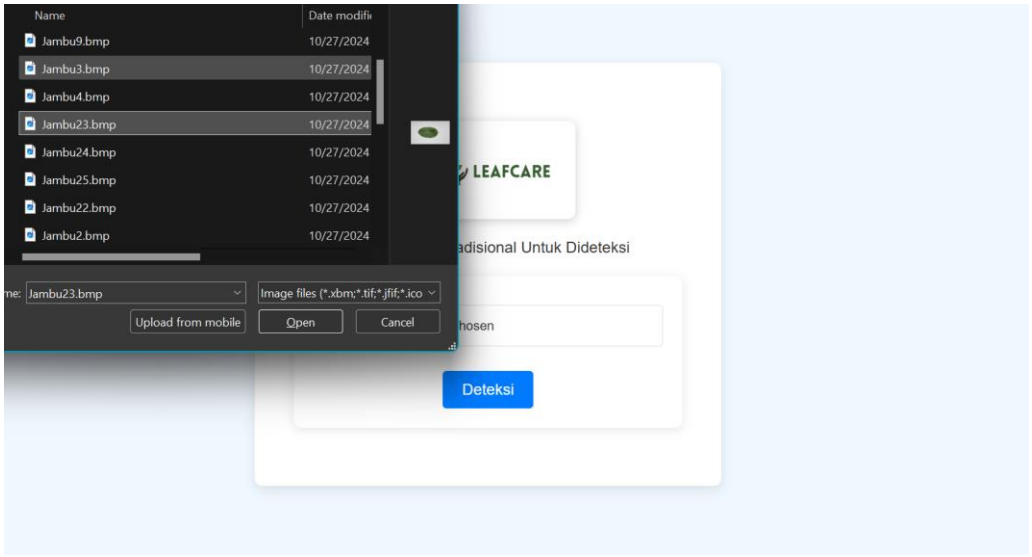



**Figure 14. Leaf Image Input**

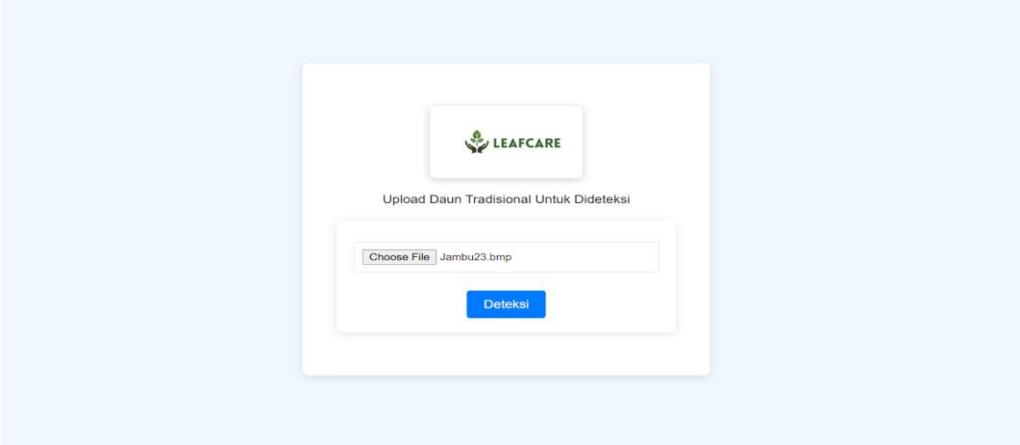



**Figure 15. Leaf Detection**

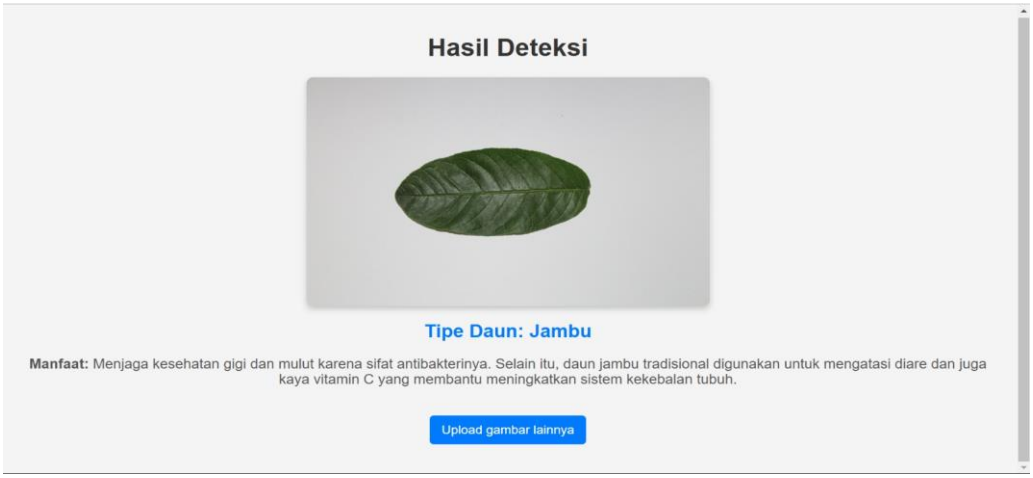



**Figure 16. Detection Results**

The test results show that the system is able to recognize leaf types with an average accuracy of 87%. The highest accuracy is achieved on leaves with very distinctive shapes and colors, such as Guava leaves which have a wide shape and bright green color, and Mango leaves which are elongated with many clear leaf veins.

However, there are some obstacles in leaves that have similarities in shape and color, such as soursop leaves and bay leaves which are difficult to distinguish by the algorithm. This indicates that certain visual features, such as leaf texture, may need to be considered as additional parameters to improve identification accuracy.

**Speed and Accuracy Analysis**

The developed system shows a fairly good response time, with an average identification time of about 1.2 seconds per image on the test device. The color segmentation and contour detection processes are the stages that most affect the response time, especially on high-resolution images. However, decreasing the resolution can reduce the accuracy of leaf feature extraction.

## 5. CONCLUSION

This study successfully developed a web-based system for identifying traditional herbal leaves using the K-Nearest Neighbors (KNN) method and image processing techniques. This system can classify types of herbal leaves based on visual characteristics, such as leaf shape and color, which are extracted through several stages of preprocessing and visual features. This system is also equipped with information on the medical benefits of each identified leaf, so that it can provide useful information for users in recognizing and understanding the efficacy of herbal leaves.

Testing shows that the system is able to identify leaf types with sufficient accuracy and can be accessed through a user-friendly interface, making it easy for users to utilize the system for everyday purposes. With this system, it is hoped that the public can more easily recognize herbal leaves that are useful as natural medicine, and can increase awareness of the use of herbal plants as an alternative solution to traditional medicine.

## REFERENSI

Heryadi, Y., & Wahyono, T. (2020). MACHINE LEARNING (Konsep dan Implementasi). Gaya Media.

Jamaliah, I., Whidhiasih, R. N., & Maimunah, M. (2017). Identifikasi jenis daun tanaman obat hipertensi berdasarkan citra RGB menggunakan jaringan syaraf tiruan. PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic, 5(1), 1-11

Jananto, A. (2022). Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika. *Progresif: Jurnal Ilmiah Komputer*, *18*(1), 69-76.http://dx.doi.org/10.35889/progresif.v18i1.769

Kim, H., Kim, M., Park, Y., Yang, S. Y., Chung, H., Kwon, O., & Yeo, H. (2019). Visual classification of wood knots using k-nearest neighbor and convolutional neural network. Journal of the Korean Wood Science and Technology, 47(2), 229–238. https://doi.org/10.5658/WOOD.2019.47.2.229

M. S. Fajri, N. Septian, and E. Sanjaya, "Evaluasi Implementasi Algoritma Machine Learning KNearest Neighbors (kNN) pada Data Spektroskopi Gamma Resolusi Rendah," Al-Fiziya: Journal of Materials Science, Geophysics, Instrumentation and Theoretical Physics, vol. 3, no. 1, pp. 9–14, Aug. 2020, doi: 10.15408/fiziya.v3i1.16180

Nugraha, I.N.B.S., Noviana, L.P.R., 2023, Perbandingan Klasifikasi Citra Daun Herbal Menggunakan Metode Logistic Regression dan Decision Tree Classifier Berdasarkan Fitur (Warna, GLCM, Bentuk), Journal Informatic Technology And Communication, 7(2), 126-133.

Pratama, E. F. A., Khairil, K., & Jumadi, J. (2022). Implementasi Metode K-Means Clustering Pada Segmentasi Citra Digital. *Jurnal Media Infotama*, *18*(2), 291-301.doi:https://doi.org/10.37676/jmi.v18i2.2899

Raharja, B. D., & Harsadi, P. (2018). Implementasi Kompresi Citra Digital Dengan Mengatur Kualitas Citra Digital. Jurnal Ilmiah SINUS, 16(2), 71–77. https://doi.org/10.30646/sinus.v16i2.363

Shofrotun, F., Sutojo, T., Ignatius, D. R., & Setiadi, M. (2018). Identifikasi Tumbuhan Obat Herbal Berdasarkan Citra Daun Menggunakan Algoritma Gray Level Co-occurence Matrix dan K-Nearest Neighbor. 6(November 2017), 51–56. https://doi.org/10.14710/jtsiskom.6.2.2018.51-56.

Sibero, A. F. K., & Saleh, A. (2020). Identifikasi Tanaman Herbal Berdasarkan Citra Daun Menggunakan Cosine Similarity dan Features Extraction. 5(1).

Siregar, R. R. A., Siregar, Z. U., & Arianto, R. (2019). Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor. Kilat, 8(1), 81–92. https://doi.org/10.33322/kilat.v8i1.421

Sunandar, H. (2017). Perbaikan kualitas Citra Menggunakan Metode Gaussian Filter. MEANS (Media Informasi Analisa Dan Sistem), 2(1), 19–22. https://doi.org/10.54367/means.v2i1.18

Suyanto, Rachmawati, E., Sulistiyo, M. D., Wulandari, G. S., & Fachrie, M. (2022). Explainable Artificial Intelligence Menggunakan Metode-Metode Berbasis Nearest Neighbors. informatika.