Identification of Medicine Leaf Images Using Invariant Moment and K-Nearest Neighbor

Fredianto[1,a], Enny Itje Sela*[,2,b], Suhirman[3,c], Ardhi Wicaksono Santoso[4,d]

[a,b,c]Universitas Teknologi Yogyakarta, Indonesia
[d]National Central University, Taoyuan, Taiwan
[1]fredianto248@gmail.com, [2]ennysela@uty.ac.id*, [3]suhirman@uty.ac.id, [4]ardhi.wicaksono@gmail.com

Abstract

Medicinal plants have benefits for preventing or curing various diseases. The number of medicinal plants and the lack of knowledge about the types of medicinal plants make it difficult for people to distinguish the types of medicinal plants. This difficulty causes people to prefer to use chemical drugs rather than medicinal plants. This study develops a system of identification of medicinal plants. There are four steps to build the system: input leaf images, pre-processing, invariant moment feature extraction, and K-Nearest Neighbours (K-NN) pattern recognition. A 100 images samples images from 5 types of medicinal plants were involved in this study. The identification process of leaf image begins with the cropping, resizing process, and several morphological operations. Then feature extraction stage uses invariant moments. The final stage of pattern recognition uses K-NN. The result of this research is that the system can identify the types of medicinal plants. Using the Manhattan distance, the study archives the highest average accuracy.

Keywords: Medical Leaf Image, Invariant Moment, K-NN

INTRODUCTION

Traditional medicine and medicinal plants have played an essential role in maintaining health, maintaining stamina, and even treating disease. Therefore, traditional medicine and medicinal plants are firmly rooted in people's lives to this day [1]. Knowledge of medicinal plants and their properties has been passed down from generation to generation, based on experience and skills [2].

Medicinal plants are plants that have benefits for preventing or curing various diseases. In the use of medicinal plants, knowledge is needed to recognize the characteristics and forms of these plants. The recognition stage is critical to ascertain whether the plants around us are medicinal plants or not. This stage is done because there are many similarities between one plant and another [3]. The number of medicinal plants and the lack of knowledge about the types of medicinal plants makes it difficult for people to distinguish the types of medicinal plants, so many prefer to use chemical drugs rather than medicinal plants. Usually, the introduction of medicinal plants can be seen from the shape of the leaves. Almost all the leaves are green, so recognition based on leaf color becomes irrelevant.

Several studies have been conducted studies on the introduction of plant species, especially medicinal plants.

The research performed RGB value for feature extraction, and an artificial neural network was implemented on six types of medicinal plants [4].

Feature extraction using the Gray Level Coordination Matrix (GLCM) algorithm and K-Nearest Neighbor classification was performed to identify ten types of medicinal plants [5]. This study resulted in the ability to identify medicinal plants with an average accuracy of 83.33%.

Research using morphological feature extraction was performed to classify five types of medical plants [2]. The artificial neural network classification was used to identify them. This study resulted in the ability to from the previous studies, the use of moment invariant has not been implemented for feature extraction, even though this method is suitable for identifying leaf patterns with unique characteristics, such as size and position. This study is carried out to identify the leaf pattern of medical plants, known locally as *alpukat*, *jambu*, *pepaya*, *sirih*, and *salam*. The five leaves are helpful to relieve the symptoms of various diseases, easy to plant, and these plants are commonly found around the house. From the previous studies, the use of moment invariant has not been implemented for feature extraction, even though this method is suitable for identifying leaf patterns with unique characteristics, such as size and position. This study is carried out to identify the leaf pattern of medical plants, known locally as *alpukat*, *jambu*, *pepaya*, *sirih*, and *salam*. The five leaves are helpful to relieve the symptoms of various diseases, easy to plant, and these plants are commonly found around the house [1], [6]. Moments Invariant or Geometric Moments Invariant was first introduced by Hu [7]. The moment can describe an object in terms of area, position, orientation, and other defined parameters. By getting several moment information, both zero (m00) and first (m10 and m01) or central moments, and moments at level 2 or moment invariant of an object, the object can be identified even though it has undergone a translation, rotation, or changes in scale [8]. K-Nearest Neighbors is a relatively simple pattern recognition method and is also widely used by researchers [9].

Table 1. The Benefit of Leaf Types

Locally name of leaf	Traditionally treatment	Image
Jambu	• Reducing the frequency of abdominal pain and diarrhea • Overcoming stomach cramps during menstruation • Reducing the risk of gum inflammation • Lowering high blood • Reduing joint pain in the knee • Increasing platelets and prevent bleeding when exposed to DHF • Treating acne and dark spots • Preventing premature aging	
Alpukat	• Improving eye health • Rich in nutrients • Providing a longer full effect • Delicious and easy to combine into a diet menu • Maintaining heart health • Preventing osteoporosis • Containing more potassium than bananas	
Pepaya	• Full of nutrition • Anticancer properties • Improving heart health • Fighting inflammation • Protecting against skin damage • Lowering the risk of Alzheimer's disease • Helping control blood sugar levels • Improve gastrointestinal health	
Sirih	• Stopping bleeding • Cough medicine • Eliminating bad breath • High blood pressure medication	
Salam	• Lowering cholesterol levels • Lowering the risk of heart attack • Lowering blood sugar levels • Helping cancer treatment • Treating kidney stones • Boosting immunity	

METHOD

Material

Medicinal plants used as research objects are *alpukat*, *jambu*, *pepaya*, *sirih*, and *salam*. The number of samples from each type of leaf is 20 images data of medicinal plants so that the total data used is 100 images leaf of medicinal plants. The benefits of leaf types and image examples can be seen in Table 1 [10].

Proposed model

The analysis is carried out to minimize expenses and consider data needs so that the tools and materials are not excessive. The stages and steps that the author does in this study can be seen in Figure 1. The explanation of each stage can be seen in the following description.

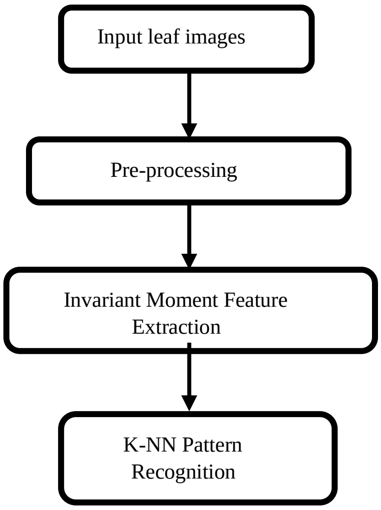

Figure 1. Proposed model

1. Input Leaf Image

This stage is a process where the system will receive input of leaf images in RGB format. The following is an example of a leaf image that will be entered into the system (Figure 2).

Figure 2. Example of the medicine leaf image

2. Pre-processing

The preprocessing process is a leaf image processing process before feature extraction is carried out. The steps for image preprocessing can be seen in the following description.

a. Convert leaf image from RGB to grayscale using COLOR_BGR2GRAY in python. The grayscale image shows in Figure 4.

Figure 4. Grayscale image

b. Edge detection using the *Canny method.* The edge detection image shows in Figure 5.

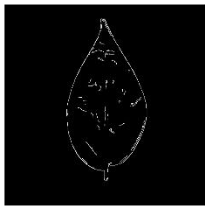

Figure 5. Dilation image

c. Dilation. The dilation image is shown in Figure 6.

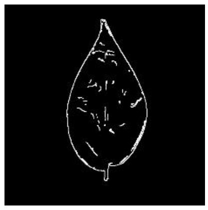

Figure 6. Dilation image

d. Filling. The filling image is shown in Figure 7.

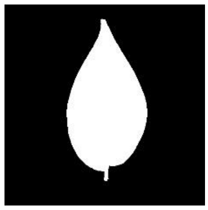

Figure 7. Filling image

e. Erosion. The erosion image is shown in Figure 8.

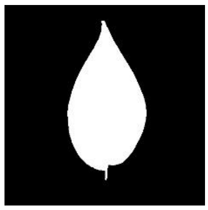

Figure 8. Erosion image

f. Cropping and resizing. This stage produces a 100 x 300 size image, as shown in Figure 9.

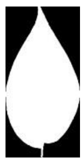

Figure 9. Cropping image

3. Invariant Moment Feature Extraction

The feature extraction process obtains features in the form of vector values from the results of the leaf image pre-processing process using the moment invariants algorithm.

Moment invariant is a non-linear function that is invariant to rotation, translation, and scale in the image geometry moment. If there is an image with an intensity value of f(x,y) the value of x as a row and y as a column, then the moment invariant transforms the function f(x,y) on a discrete system [7], [11] Moment invariant is a feature extraction method that can handle translation, scaling and rotation of images. This method will produce seven features that can later be used in character recognition.

Moment invariant feature extraction is used to extract shape features so that when the image is transformed, it can still be recognized. This process is carried out to produce features in the form of vectors from binary images. The feature used is the seven-moment invariant, which will produce seven values in the feature vector. The writer uses the moment invariant is because this method can handle the shape of the image when the same image transforms.

This method will produce seven traits that can later be used in character recognition. The seven values in the feature vector are obtained using equation (1).

Image moments invariant to image translation are of the order m00, m01, m02, m03, m10, m11, m12, m20, m21, and m31. Furthermore, the coordinates of the centre of the image are determined based on the image moment, which is calculated using equation (2) and equation (3). m00 is the total number of pixels that make up the object, while m10 and m01 are the object's center of mass. Furthermore, to obtain the moment invariant to the rotation, the central moment is obtained by using equation (4) to obtain the moment invariant to the rotation. The formed center moment is sensitive to rotation and scaling transformations. Therefore, normalization of the central moment is carried out using equation (5) to equation (7).

Based on the normalization of the central moment, seven-moment invariant values can be calculated using equation (8) to equation (14). The results of feature extraction can be seen in Table 2.

$$m_{pq} = \sum_{x=0}^{W-1} x \sum_{y=0}^{H-1} y \; x^p y^q \; f(x,y) \tag{1}$$

$$\bar{x} = \frac{m10}{m00} \tag{2}$$

$$\bar{y} = \frac{m01}{m00} \tag{3}$$

$$\mu_{pq} = \sum_{x=0}^{W-1} x \sum_{y=0}^{H-1} y \; (x-\bar{x})^p (y-\bar{y})^q \; f(x,y) \tag{4}$$

$$\eta_{pq} = \frac{\mu_{pq}}{{\mu_{00}}^{\gamma}} \tag{5}$$

$$\gamma = \left(\frac{p+q}{2}\right) + 1 \tag{6}$$

$$\mu_{00} = m_{00} \tag{7}$$

$$\varphi 1 = \eta_{20} + \eta_{02} \tag{8}$$

$$\varphi 2 = (\eta_{20} - \eta_{02})^2 + 4{\eta_{11}}^2 \tag{9}$$

$$\varphi 3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{10}$$

$$\varphi 4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{11}$$

$$\varphi 5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + (3\eta_{21} - \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{12}$$

$$\varphi 6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})] \tag{13}$$

$$\varphi 7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{14}$$

Where:
m = Image moment
p, q = Orde
f = Intensity image value
x, y = Pixel coordinate
H, W = Height, Weight image
μ : Central moment
$\bar{x}, \bar{y}$: Central image

Table 2. Example of Feature Extraction

No	Label	$\varphi 1$	$\varphi 2$	$\varphi 3$	$\varphi 4$	$\varphi 5$	$\varphi 6$	$\varphi 7$
1	*Alpukat*	3,0781	6,5295	10,9814	11,7894	17,8183	15,2795	23,2535
2	*Alpukat*	3,1189	6,7320	11,6275	12,5122	-19,4988	15,9140	24,7409
3	*Alpukat*	3,1045	6,6578	10,8866	11,8299	17,6989	15,3175	23,1949
4	*Alpukat*	3,0943	6,6120	12,0654	12,9384	-18,7734	17,2094	25,9669
5	*Jambu*	3,1019	6,6315	11,7508	12,4499	19,4389	15,7665	25,3521
6	*Jambu*	3,1009	6,6258	11,7532	12,4465	19,4066	15,7605	25,3141
7	*Jambu*	3,0782	6,5137	12,9631	13,9647	20,8307	17,3708	27,4302
8	*Jambu*	3,0693	6,4774	11,6565	12,4137	19,1434	15,6564	-24,9960
9	*Pepaya*	3,0946	8,5851	10,9350	11,8601	-19,0938	-16,4449	23,3145
10	*Pepaya*	2,0386	4,1552	6,2371	6,0846	12,0076	8,1623	13,9180
11	*Pepaya*	3,1359	8,8250	10,5871	12,8113	-17,9929	-18,3568	24,7639
12	*Pepaya*	3,1368	9,2952	11,0360	11,7582	-18,1621	-16,8200	-23,2178
13	*Sirih*	3,0738	6,5169	10,8811	11,6777	17,5556	15,3736	23,8522
14	*Sirih*	3,0918	6,5923	11,3336	12,2711	17,7310	16,4490	24,6353
15	*Sirih*	3,0650	6,4693	11,7712	12,1207	-18,7776	15,3553	25,0227
16	*Sirih*	3,0872	6,5767	12,2226	13,2346	-19,6627	16,7248	25,9757
17	*Salam*	3,1238	6,8154	10,5815	11,2511	-17,2137	14,6675	22,4762
18	*Salam*	3,1215	6,7954	10,5441	11,2001	17,3629	14,6004	-22,6997
19	*Salam*	3,1438	7,0071	10,4606	11,2716	16,9445	14,7939	-22,3451
20	*Salam*	3,1263	6,8643	10,3437	11,0709	16,7373	14,5090	-22,1444

4. Pattern Recognition with K-NN

K-Nearest Neighbors (K-NN) is a supervised learning algorithm. The results of the new query instance are classified based on the majority of the categories in the k-nearest neighbors. The class that appears the most will be the classified class [12]. Near or far neighbors are calculated using the algorithm Euclidean Distance using equation (15), Manhattan distance using equation (16), dan Chebyshev distance using equation (17).

$$d_{(x,y)} = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \tag{15}$$

$$d_{(x,y)} = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \tag{16}$$

$$d_{(x,y)} = \sum_{i=1}^{m} |x_i - y_i| \tag{17}$$

$$d_{(x,y)} = \max_{i=1}^{n} |x_i - y_i| \tag{18}$$

Where:

x : trained features
y : tested features
m : size of features

RESULT AND DISCUSSION

Result

At this stage, the authors test the system using 100 data in the form of images of medicinal plant leaves. An example of system output can be seen in Figures 10 and 11. The test accuracy table can be seen in Table 3 to Table 5.

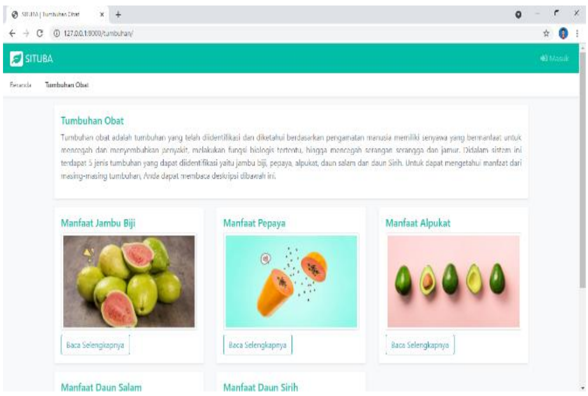

Figure 10. Initial page

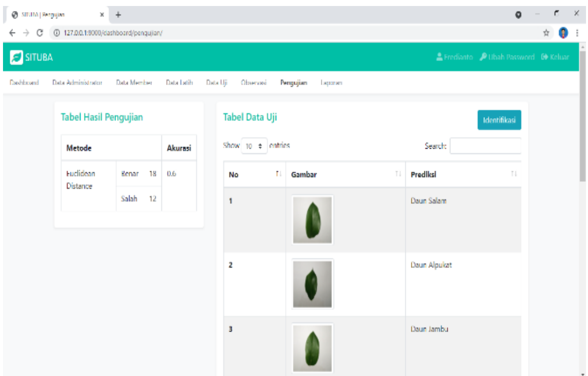

Figure 11. Leaf recognition page

Discussion

In this study, the system produces nine models. The model is grouped based on the distance calculation algorithm used in the K-Nearest Neighbours algorithm's classification process. Based on the accuracy tables obtained in Tables 1 to 3, it can be concluded that the highest average accuracy for all types of leaves uses the Manhattan distance algorithm and the k value of 3 is 70%.

Table 3. Accuracy of Testing Using Euclidean Distance

Leaf	**K**		
	3	**5**	**7**
Alpukat	50%	50%	50%
Jambu	25%	50%	75%
Pepaya	75%	50%	50%
Sirih	75%	25%	25%
Salam	100%	100%	100%

Table 4. Accuracy of testing using Manhattan Distance

Leaf	**K**		
	3	**5**	**7**
Alpukat	50%	50%	50%
Jambu	50%	50%	75%
Pepaya	75%	50%	25%
Sirih	75%	25%	25%
Salam	100%	100%	100%

Table 5. Accuracy of testing using Chebyshev Distance

Leaf	**K**		
	3	**5**	**7**
Alpukat	25%	50%	25%
Jambu	25%	50%	50%
Pepaya	50%	50%	50%
Sirih	75%	25%	25%
Salam	100%	75%	75%

While the lowest accuracy on all types of leaves using a distance calculation algorithm with Chebyshev distance and k value of 7 is 45%.

The performance of image identification of medicinal plant leaves using feature extraction of hu moment invariants and classification with K-Nearest Neighbours shows the identification results are quite good. It is indicated by the average accuracy of all types of leaves, which is 70%.

CONCLUSION

Based on the results of the tests that have been carried out on the identification system of medicinal plants based on leaf patterns, it can be concluded that, first, the identification system of medicinal plants based on leaf patterns using hu moment invariants and K-Nearest Neighbours can identify medicinal plant species well at a value of k = 3 using the Manhattan distance calculation algorithm. Second, the highest average accuracy produced by the system is 70%.

The research that has been done still needs to be improved; to improve the system in further investigation, the following methods can be used, namely:
1. Developing a leaf identification system of medicinal plants with other feature extraction methods.
2. Adding a dataset for each type of leaf.
3. Developing a real-time identification system for medicinal plant leaves.

REFERENCES

[1] A. Hikmat, E. A. M. Zuhud, Siswoyo, E. Sandra, and R. K. Sari, "Revitalisasi Konservasi Tumbuhan Obat Keluarga (Toga) Guna Meningkatkan Kesehatan Dan Ekonomi Keluarga Mandiri Di Desa Contoh Lingkar Kampus Ipb Darmaga Bogor," *Jurnal Ilmu Pertanian Indonesia*, vol. 16, no. 2, pp. 71–80, 2011.

[2] K. Saputra S and M. I. Perangin-Angin, "Klasifikasi Tanaman Obat Berdasarkan Ekstraksi Fitur Morfologi Daun Menggunakan Jaringan Syaraf Tiruan," *Jurnal Informatika*, vol. 5, no. 2, pp. 169–174, 2018, doi: 10.31311/ji.v5i2.3770.

[3] F. Y. Manik, Y. Herdiyeni, and E. N. Herliyana, "Leaf Morphological Feature Extraction of Digital Image Anthocephalus Cadamba," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 14, no. 2, p. 630, 2016, doi: 10.12928/telkomnika.v14i2.2675.

[4] I. Jamaliah, "Identifikasi Jenis Daun Tanaman Obat Hipertensi Berdasarkan Citra Rgb Menggunakan Jaringan Syaraf Tiruan," *Penelitian Ilmu Komputer Sistem Embedded dan Logic*, vol. 5, no. 1, pp. 1–11, 2017.

[5] M. A. Agmalaro, A. Kustiyo, and A. R. Akbar, "Identifikasi Tanaman Buah Tropika Berdasarkan Tekstur Permukaan Daun Menggunakan Jaringan Syaraf Tiruan," *Jurnal Ilmu Komputer dan Agri-Informatika*, vol. 2, no. 2, pp. 73–82, 2013.

[6] Ahmad Izzuddin and M. Rizal Wahyudi, "Pengenalan Pola Daun untuk Membedakan Tanaman Padi dan Gulma Menggunakan Metode Principal Components Analysis (PCA) dan Extreme Learning Machine (ELM)," *ALINIER: Journal of Artificial Intelligence & Applications*, vol. 1, no. 1, pp. 44–51, 2020, doi: 10.36040/alinier.v1i1.2521.

[7] G. Ananggadipa, A. Hidayatno, and A. A. Zahra, "Pengenalan Huruf Alfabet

Menggunakan Tujuh Invarian Momen Hu Dan Jaringan Saraf Tiruan Lvq (Learning Vector Quantization)," *Transient*, vol. 3, no. 4, pp. 429–432, 2014.

[8] A. Sunyoto, "Analisa Metode Moment Invariant Untuk Mendeteksi Obyek Yang Telah Mengalami Transformasi," *Data Manajemen dan Teknologi Informasi (DASI)*, vol. 14, no. 1, p. 1, 2013.

[9] N. N. Dzikrulloh and B. D. Setiawan, "Penerapan Metode K – Nearest Neighbor (KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus: Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri)," *Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 5, pp. 378–385, 2017.

[10] H. Dewoto, "Pengembangan Obat Tradisional Indonesia menjadi Fitofarmaka," *Majalah Kedokteran Indonesia*, pp. 205–11, 2007.

[11] Eskanesiari, A. Hidayatno, and R. R. Isnanto, "Jaringan Saraf Tiruan Perambatan Balik," *Transient*, vol. 3, p. 98, 2014.

[12] F. S. Ni'mah, T. Sutojo, and D. R. I. M. Setiadi, "Identification of Herbal Medicinal Plants Based on Leaf Image Using Gray Level Co-occurence Matrix and K-Nearest Neighbor Algorithms," *Jurnal Teknologi dan Sistem Komputer*, vol. 6, no. 2, pp. 51–56, 2018, doi: 10.14710/jtsiskom.6.2.2018.51-56.