Ilomata International Journal of Social Science (IJSS)
P-ISSN: 2714-898X; E-ISSN: 2714-8998
Volume 2, Issue 1 January 2021
Page No. 81-91 https://www.ilomata.org/index.php/ijss

# Algorithm Configuration K-Nearest To Clarification Medicine Tree Based On Extraction, Variation Of Color, Texture And Shape Of Leaf

**Ardhi Dinullah Baihaqie[1], Rayung Wulan[2]**
[12]Universitas Indraprasta PGRI, Fakultas Teknik & Ilmu Komputer , Jakarta – Indonesia
Correspondent: nufus.ardhi@gmail.com

Submitted : December 7, 2020 Revised : December 26, 2020 Published : January 31, 2021

## ABSTRACT
At this time to overcome difficulties in identifying medicinal plants that have an impact on the frequent errors in the use of medicinal plants. The formulation of the problem to be discussed in this study is how to identify medicinal plants based on feature extraction of color, texture, and leaf shape. Steps to resolve this problem by collecting image data of medicinal plants, then the image data extracted leaf color features using Red Green Blue (RGB) and Hue Saturation Value (HSV), based on leaf texture using the Gray Level Co-occurrence Matrix (GLCM), based on the shape leaves use eccentricity and metrics. and then classified by the K-Nearest Neighbor (KNN) method. The results in this study the accuracy of Chinese Petai leaves is superior to other types of leaves, which is 98%, which occurs at each K value. Other types of leaves have various values. Saga leaves range between 94% - 97%, Green Betel leaves between 92.8% - 97%, and Red Betel leaves between 91.7% - 95%, Optimal K values indicated by K = 3 have an average accuracy rate of 96.7% also have sensitivity value of 93.3%. The addition of K = 5, K = 7, K = 9, and K = 11 tends to decrease the average value of accuracy and sensitivity.
Keywords: Clarification, KNN Algorithm, Digital Image Processing, Feature Extraction, Medicinal Plants

## INTRODUCTION
Medicinal plants are plants that are approved and known based on research on humans that contain beneficial plants, cure diseases, perform certain biological functions, and use in the prevention of insect and fungal attacks. At least 12 thousand compounds have been isolated from various medicinal plants in the world, but this amount is only ten percent of the total number of compounds that can be extracted from all medicinal plants (Tapsell LC1, Hemphill I, Cobiac L, Patch CS, Sullivan DR, Fenech M, Roodenrys S, Keogh JB, Clifton PM, Williams PG, Fazio VA, 2014).

Research Data conducted (Aditama., 2015) National Health Research was conducted by the Department of Health's research and development at the Ministry of Health, showing that 30.4% of households in Indonesia are utilizing traditional health services, including 77.8% of traditional types of Health utilizing skills without tools, and 49.0% of households make use of herbs. Meanwhile, Riskesdas 2010 shows that 60% of the population is above the age of 15 years, Indonesia said that they drink potions once, and 90% of them stated the benefits of drinking herbs.

A problem is the large number of medicinal plants and the lack of knowledge about the types of medicinal plants in terms of distinguishing types of medicinal plants which have an impact

on the frequent errors in identifying types of medicinal plants. Image data used as the object of research are leaves on medicinal plants.

Research conducted (Gustina, Fadlil, and Umar, 2016) using neural network methods can be implemented, one of which is the introduction of Cambodian leaf patterns. Texture patterns that have frangipani leaves are needed by farmers to recognize the types and characteristics of leaves. This study is an application of artificial neural networks to determine the types of frangipani flowers based on their leaf patterns. The application was carried out using the Back Propagation algorithm method for 2 types of Cambodian petal Leaves, namely Cambodia Japan, and Cambodia Bali, each of which has the same leaf pattern. The neural network method is designed by determining the type of frangipani leaves, Research conducted.

## METHODS

In this study including collecting image data of medicinal plants, the data features extracted image images based on leaf color using *Red Green Blue* (RGB) and *Hue Saturation Value* (HSV), based on leaf texture using the *Gray Level Co-occurrence Matrix* (GLCM) and leaf shape uses eccentricity and metric. Then it is classified by *K-Nearest Neighbor* (KNN) to get identification of medicinal plant variants because the results of classification are good (Li, Jiang, and Yin, 2014) (Eliyen, Tolle and Muslim, 2017), and Fast (Hwang and Wen, 1998), (Zhao, Qian, and Li, 2018). The design of the identification model is shown in Figure 1.

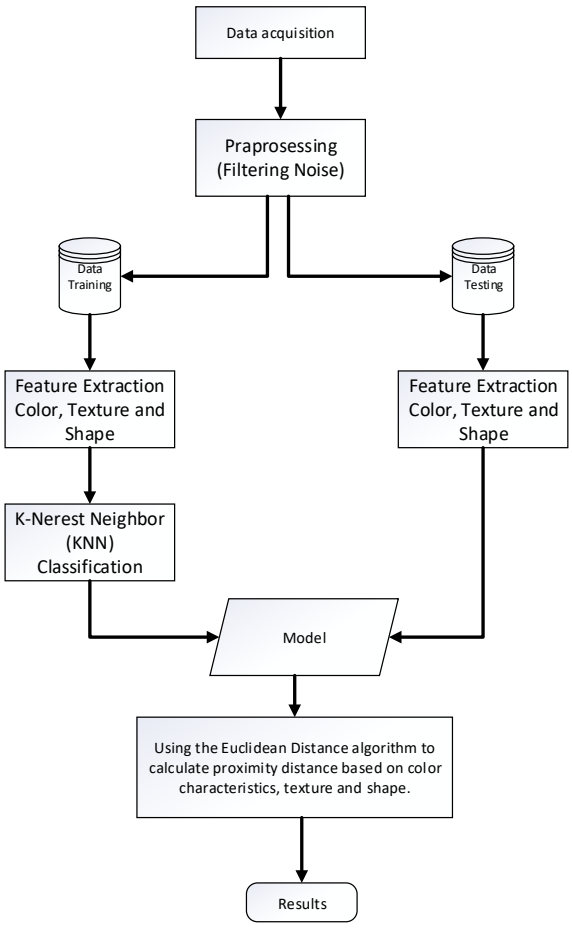



**Figure 1. Design of Identification Model**

1. Data acquisition. Collecting plant leaf images Medicines from several variants of medicinal plants such as saga leaves, Chinese petai, green betel leaf, and red betel leaf are taken using a camera.
2. Processing. Pre-Before extraction removal noise removal of feature segmentation is done to produce training data images that are better than noise in the image not only imperfections in the form of images or during transmission. But also because of poor shooting. Some noise can be attached to the image, one of which is salt and paper in the form of the spread of black or white dots in the image according to (Pardosi, 2016).
3. Training data and data testing. In this phase, the training algorithm only performs storage features and vector classification. Data samples in the training phase, the same classification features are calculated to test data (some of which are unknown). The distance from this new vector trains all vector samples calculated from the nearest k. Several new experiences are mostly euclidean point distance classification algorithms for calculating proximity based on color, texture, and shape (Sikki, 2009). In this study, 280 training data were divided into 4 classes of leaf types of medicinal plants, each with 70 training data for each class. While the testing data amounted to 120 data, which was divided into 4 classes of plant types each of 30 testing data. The Medicinal planting image shown in Figure 4.2 is obtained by photographing directly using data stored in the .jpg file format.
4. Feature
   a. Extraction Color feature extraction in the process of dredging color features begins by changing the direction for gray RGB (grayscale) color. Each group of five sums then by dividing the length and breadth of the image (many) pixel color compiler image to distinguish between objects of a certain color can use that color is a representation of visible light (red-orange, yellow, green, blue, purple). Color values can be combined in saturation and color brightness. Getting this third class, is needed to change the color of the image, initially RGB red (green, blue) for HSV (hue, saturation, value). (Alviansyah, Ruslianto and Diponegoro, 2017) The colors compiled by mixed color models are the primary colors of *Red, Green*, and *Blue* based on certain compositions. It is expected that in the form of RGB HSV can be formulated in the following equations 3.1, 3.2, and 3.3;

$$r = \frac{R}{R+G+B} \text{.....(3.1)}$$

$$g = \frac{G}{R+G+B} \text{.....(3.2)}$$

$$b = \frac{B}{R+G+B} \text{.....(3.3)}$$

By utilizing the normalized values of r, g, and b, the RGB to HSV conversion formula can be formulated in equations 3.4, 3.5, and 3.6 the extract results are shown in the following figure 2:

$$V = max(r, g, b) \quad \text{.....(3.4)}$$

$$S = \begin{cases} 0 & jika\ V = 0 \\ V - \frac{min(r,g,b)}{V} & jika\ V > 0 \end{cases} \quad \text{.....(3.5)}$$

$$H = \begin{cases} 0 & jika\ S = 0 \\ \frac{60 \times (g-b)}{S \times V} & jika\ V = r \\ 60 \times \left[2 + \frac{(b-r)}{S \times V}\right] & jika\ V = g \\ 60 \times \left[4 + \frac{(r-g)}{S \times V}\right] & jika\ V = b \end{cases} \quad \text{.....(3.6)}$$

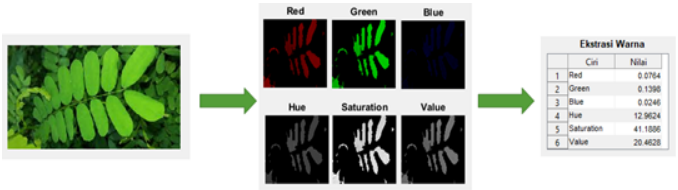



**Figure 2. Results of RGB to HSV Color**

b. Extraction texture characteristics are the characters possessed by the region (in the picture) with the values of *contrast, correlation, energy*, and *homogeneity* of course these properties can be done at the level of the Gray Level Co-occurrence Matrix (GLCM) this parameter is extracted in an image. (Kasim and Harjoko, 2014) The method used for feature extraction is the Gray Level Co-occurrence Matrices (GLCM). GLCM is a method for obtaining second-order statistical characteristics by calculating probability relationships between two neighboring pixels at a distance and invariant orientation. Co-occurrence (co-occurrence) means a shared event, it means the number of occurrences of one neighboring pixel level with pixel values to another level in (d) and a certain angular orientation (θ). Distance is expressed in pixels and orientation is expressed in degrees. Orientation is formed in four angles with an interval of 45 °, namely 0 °, 45 °, 90 °, and 135 °. While the distance between pixels is usually set at 1 pixel. The direction and distance of the GLCM can be seen in Figure 2.1. Whereas the process of forming GLCM is an image with 4 levels of gray (gray level) at a distance d = 1 and direction 0o. The direction and distance of pixels in the GLCM shown in Figure 3.3 illustrates the process of the cursurence matrix formed to extract the image characteristics based on GLCM, the extraction results are shown in Figure 3.5.

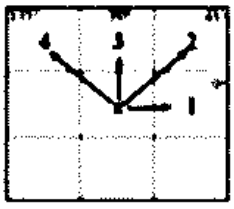


Figure 3 x pixels in the middle, pixels 1, pixels 2, pixels 1 all distances d = 1

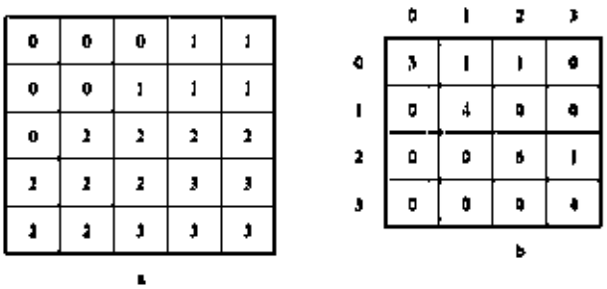


Figure 4 (a). Image with 4 levels of gray (b). GLCM with direction = 0o and distance = 1

$$contrast = \sum_{a,b} P_{a,b}(a-b)^2 \quad \text{.....(3.4)}$$

$$Energi = \sum_{a,b} P^2_{\theta,d}(a,b) \quad \text{.....(3.5)}$$

$$homogenitas = \sum_a \quad \sum_b \frac{1}{1+(a-b)^2} P_{\theta,d}(a,b) \quad \text{.....(3.6)}$$

$$korelasi = \frac{\sum_{a,b}[(ab)P_{\theta,d}(ab)]-\mu_x\mu_y}{\sigma_x\sigma_y} \quad \text{.....(3.7)}$$

With

$$\mu_x = \sum_a a \sum_b P_{\theta,d}(a,b),\ \mu_y = \sum_b b \sum_a P_{\theta,d}(a,b)$$
$$\sigma_x = \sum_a (a-\mu_x)^2 \sum_b P_{\theta,d}(a,b),$$
$$\sigma_y = \sum_b (b-\mu_y)^2 \sum_a P_{\theta,d}(a,b)$$

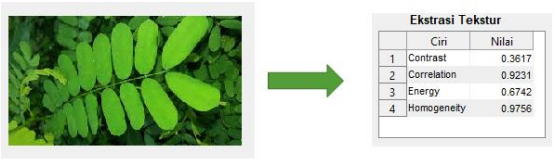



**Figure 5. Image Texture Feature Extraction Using GLCM**

c. Extraction features of the image shape are very important for image segmentation because it can detect objects or boundaries. To distinguish objects from one object to another, you can use a parameter called eccentricity. Eccentricity is the value of the comparison between the distance between the minor ellipse and the major ellipse of the object. Eccentricity has various values between 0 and 1 (Nixon and Aguado, 2008). The calculation of eccentricity is shown in Figure 3.6.

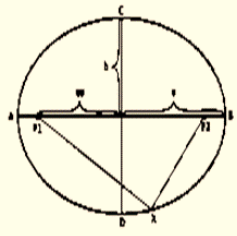


Figure 6. Calculation of eccentricity

The calculation of eccentricity is presented in Equation 3.8

$$eccentricity = \sqrt{1 - = \frac{b^2}{a^2}} \quad \text{.....(3.8)}$$

Objects that are shaped like elongated or close form a straight line, the metric value is close to 0, while the object is round or circular, the metric value is close to number 1. The calculation of the metric is shown in Figure 3.7.

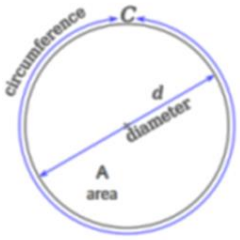
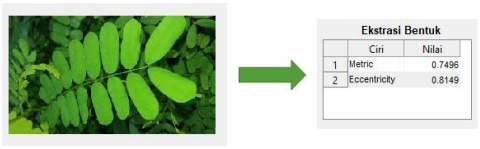



Figure 7. Calculation of metrics

$$metric = (4\pi * A)/c\text{^}2 \quad \text{.....(3.9)}$$

With C is the circumference of an image with pixel size, and A is the area of the image area with pixel units. To distinguish the size of an object from another object, you can use the area and perimeter parameters. An area called the area is the number of pixels that make up an object. While the circumference which is also called the perimeter is the number of pixels that surround an object. Based on the size of the metric, area, and perimeter values can be derived a value called a shape factor such as the roundness, aspect ratio, and triangle values, the shape feature extraction results are shown in figure 8.




5. Image Classification uses K Nearest Neighbor (KNN). Method K-Nearest Neighbor (KNN) to classify objects based on training data. The purpose of this algorithm is to classify new objects based on training data attribute values. The best k value for this algorithm depends on the data, generally, a high k-value will reduce the sound effects on classification, but make the boundaries between each classification more blurred. This cross-validation technique is used to find the optimal k value in finding the best parameters in the model. Euclidean distance (McAndrew, 2004), is used to calculate the distance between two vectors which serves to test the size that can be used as an interpretation of the proximity of the two objects represented in equation 3.10. with d (x, y): Euclidean distance between vector x and vector y; xi: a feature I of vector x; Yi: a feature I of vector y; n: number of features in the x and y vectors.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad \text{.....(3.10)}$$

The concept uses the Euclidean distance function, to avoid differences in the K-NN range of values for each attribute (recommended) it is necessary to do a transformation. Transformation is used to equate scale features to certain types, such as-as-1 for 1 or 0-1. The transformation method used is Min-Max normalization which results in a linear transformation in the original data to produce the same range of values (Han, Kamber, and Pei, 2012) as in equation (2.11).

$$V^1 = \frac{V - min_A}{max_A - min_A} \quad \text{.....(2.11)}$$

with:

V1 : new value Min-Max Normalization

V : value of the feature to be transformed

minA : min value of the field in the same feature
maxA : max value of the field in the same feature
new_minA : minimum value desired feature
new_maxA : maximum value of feature desired

6. Implementation of system implementation. This is implemented by making a model of digital image processing computing approach,
7. Testing Confusion Matrix Multi-Class. A confusion matrix is one method used to measure the performance of classification methods. The confusion matrix contains information that compares the results of the classification carried out by the system with the results of the classification must. Based on the number of output classes, the classification system can be divided into four (4) binary types of classification, that is, labels of multi, multi hierarchical, and class (Sokolova and Lapalme, 2009).

In the *confusion matrix*, there are four (4) provisions as a representation of the results of the classification process. The fourth term issue Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). MR value is the number of negative data detected correctly, while FP is negative, but data is detected as positive data. TP is positive data detected correctly, while FN is positive but data is detected as negative data. Based on the value of TN, FP, FN, and TP can be obtained values of accuracy, specificity, and sensitivity. Accuracy values describe how accurately the system can classify data correctly. In other words, the accuracy value is a comparison between data that is correctly classified with the whole data (Equation 3.12). Value Sensitivity shows how many percent of the positive category data is correctly classified by the classifier (Equation 3.13) (Han, Kamber, and Pei, 2012).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad \text{.....(3.12)}$$

$$Sensitivitas = \frac{TP}{FN + TP} \times 100\% \quad \text{.....(3.13)}$$

with,

a. TP is True Positive, which is the amount of positive data that is correctly classified by the classifier.
b. TN is True Negative, which is the amount of negative data that are correctly classified by the classifier.
c. FN is False Negative, that is the number of negative data but incorrectly classified by classifier
d. FP is False Positive, which is the number of positive data that are classified incorrectly by the classifier.

In the multi-class classification, how to calculate accuracy and sensitivity, here are the equations to calculate the average accuracy value (Equation 3.13), and micro average Sensitivity (Equation 3.14) in the multi-class classification (Sokolova and Lapalme, 2009).

$$Akurasi\ rerata = \frac{\sum_{i=1}^{l} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} \times 100\% \quad \text{.....(3.14)}$$

$$Sensitivitas_{\mu} = \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l} (TP_i + FN_i)} \times 100\% \quad \text{.....(2.15)}$$

where,

〖TP〗_i is True Positive in class I, i.e. the amount of positive data that is correctly classified by the classifier.

〖TN〗_i is True Negative in class I, which is the amount of negative data that are correctly classified by the classifier.

〖FN〗_i is False Negative in the I class, i.e. the number of negative data but is incorrectly classified by the classifier.

〖FP〗_i is False Positive in class I, which is the amount of positive data that are classified incorrectly by the classifier.

l is the number of classes

μ is micro mean

## RESULTS AND DISCUSSION

Model-identification of medicinal plants based on color, texture, and leaf shape feature extraction has been made using 280 leaf image data for Training data and 120 leaf images for Data Testing. The following are the results of the Classification of KNN using K = 3, K = 5, K = 7, K = 9 and K = 11 shown in Tables 4.1, 4.2, 4.3, 4.4 and 4.5 below. Testing classification uses Confusion Matrix Multi-Class, image class is calculated using equation (3.12), Sensitivity is calculated using equation (3.13), Average accuracy is calculated using equation (3.14) Input data are classified into several classes in multi-class classification, to calculate accuracy, and sensitivity is calculated using equation (3.15) the calculation results are shown in tables 4.1 to 4.5.

**Table 1.**
**Classification Using KNN K = 3**

| Citra Daun | Kelas Prediksi K= 3 | | | | Akurasi % |
|---|---|---|---|---|---|
| | Saga | Petai Cina | Sirih Hijau | Sirih Merah | |
| Saga | 28 | 0 | 1 | 1 | 97 |
| Petai Cina | 1 | 28 | 0 | 1 | 98 |
| Sirih Hijau | 0 | 0 | 27 | 3 | 97 |
| Sirih Merah | 1 | 0 | 0 | 29 | 95 |
| Total Akurasi % | 96,7 | | | | |
| Sensitifitas % | 93,3 | | | | |

**Table 2.**
**Classification Using KNN K = 5**

| Citra Daun | Kelas Prediksi K= 5 | | | | Akurasi % |
|---|---|---|---|---|---|
| | Saga | Petai Cina | Sirih Hijau | Sirih Merah | |
| Saga | 26 | 0 | 3 | 1 | 94 |
| Petai Cina | 1 | 28 | 0 | 1 | 98 |
| Sirih Hijau | 0 | 0 | 28 | 2 | 96 |
| Sirih Merah | 2 | 0 | 0 | 28 | 95 |
| Total Akurasi % | 95,8 | | | | |
| Sensitifitas % | 91,7 | | | | |

**Table 3.**
**Classification Using KNN K = 7**

| Citra Daun | Kelas Prediksi K= 7 | | | | Akurasi % |
|---|---|---|---|---|---|
| | Saga | Petai Cina | Sirih Hijau | Sirih Merah | |
| Saga | 26 | 0 | 3 | 1 | 95,0 |
| Petai Cina | 1 | 28 | 0 | 1 | 95,0 |
| Sirih Hijau | 0 | 0 | 26 | 4 | 95,0 |
| Sirih Merah | 1 | 0 | 2 | 27 | 95,0 |
| Total Akurasi % | 94,6 | | | | |
| Sensitifitas % | 89,2 | | | | |

**Table 4.**
**Classification Using KNN K = 9**

| Citra Daun | Kelas Prediksi K= 9 | | | | Akurasi % |
|---|---|---|---|---|---|
| | Saga | Petai Cina | Sirih Hijau | Sirih Merah | |
| Saga | 27 | 0 | 2 | 1 | 95,0 |
| Petai Cina | 1 | 28 | 0 | 1 | 98,3 |
| Sirih Hijau | 0 | 0 | 27 | 3 | 93,3 |
| Sirih Merah | 2 | 0 | 3 | 25 | 91,7 |
| Total Akurasi % | 94,58 | | | | |
| Sensitifitas % | 89,2 | | | | |

**Table 5.**
**Classification Using KNN K = 11**

| Citra Daun | Kelas Prediksi K= 11 | | | | Akurasi % |
|---|---|---|---|---|---|
| | Saga | Petai Cina | Sirih Hijau | Sirih Merah | |
| Saga | 27 | 0 | 2 | 1 | 94,2 |
| Petai Cina | 1 | 28 | 0 | 1 | 98,3 |
| Sirih Hijau | 0 | 0 | 26 | 4 | 94,2 |
| Sirih Merah | 3 | 0 | 1 | 26 | 91,7 |
| Total Akurasi % | 94,58 | | | | |
| Sensitifitas % | 89,17 | | | | |


Figure 8. Classification KN Test Graph K = 3, K = 5, K = 7, K = 9 and K = 11

Figure 8. The picture above shows that the accuracy of the leaves of Chinese Petai is superior to other leaf types, which is 98%, which occurs at each K value. Other leaf types have a value diversity. Saga leaves range from 94% - 97%, Green Betel Leaves between 92.8% - 97% and Red Betel leaves between 91.7% - 95%. It can be seen that in the leaf class Saga the addition of K3 to K5 values decreased by 95.7% to 94.2% but the addition of K values from K5 to K7 was able to increase the accuracy from 94.2% to 95%. Furthermore, the addition of K values in green betel has decreased the value of accuracy from K5 to K7 95.8% to 92.8% but has increased after the addition of the values of K9 and K11 93.3% to 94.2% there is an increase in the value of accuracy in the class of green betel leaf. Whereas in the Red Betel leaf class the opposite occurs. The addition of the K value causes the accuracy in the Red Betel leaf class to decrease. Whereas the accuracy of Saga and Green Betel leaves fluctuated, where the addition of K values caused a decrease in the value of accuracy, but then the addition of the K value was able to increase the value of accuracy again. Added-value of K increased the number of FN and FP especially in the class of Green Betel leaves and Red Betel leaves. When the value of K is too large it will make the classification results more blurred (Karomi, 2015).

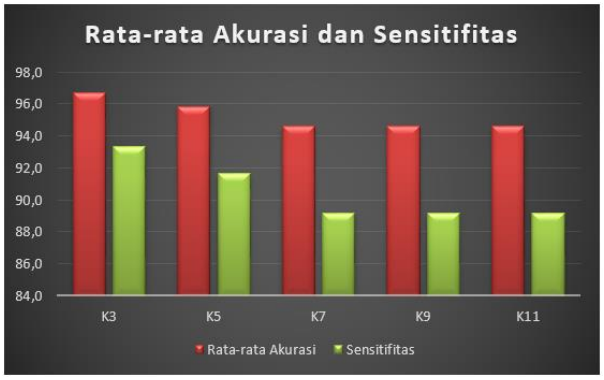



**Figure 9. Graphs of Accuracy Testing and Sensitivity of Classification of KN = 3 KN, K = 5, K = 7, K = 9 and K = 11**

Figure 9. above Sensitivity shows the percentage of images of leaves that can be classified correctly. It is seen that K = 3 in addition to having a high average accuracy also has a high sensitivity value. The addition of K values tends to decrease the average value of accuracy and sensitivity. From the discussion above, it can be said that K = 3 is the optimal K for the identification model of medicinal plants based on color, texture, and leaf shape feature extraction and uses the KNN classification.

## CONCLUSION

Based on the analysis and description above, it can answer questions from the problem formulation namely a conclusion that identification of medicinal plants based on color feature extraction, texture, and leaf shape using the KNN classification can produce an accuracy of 96.7% and sensitivity of 93.3% can be good in identifying medicinal plants. The optimal K value indicated by K = 3 has an average accuracy rate of 96.7% also has a sensitivity value of 93.3%. The addition of K = 5, K = 7, K = 9, and K = 11 tends to decrease the average value of accuracy and sensitivity shown in Figure 4.2

## REFERENCES

Aditama., T. Y. (2015) Jamu & Kesehatan. Available at: terbitan.litbang.depkes.go.id/penerbitan/index.php/catalog/book/160/160-99Z_Book Manuscript-372-1-10-20150521.pdf.

Alviansyah, F., Ruslianto, I. and Diponegoro, M. (2017) 'Identifikasi Penyakit Pada Tanaman Tomat Berdasarkan Warna Dan Bentuk Daun Dengan Metode Naive Bayes Classifier Berbasis Web', Jurnal Coding Sistem Komputer Untan, 05(1), pp. 23–32.

Eliyen, K., Tolle, H. and Muslim, M. A. (2017) 'K-NEAREST NEIGHBOR UNTUK KLASIFIKASI PENILAIAN PADA VIRTUAL PATIENT CASE Kunti', Scholarpedia, 4(2), p. 1883. DOI: 10.4249/Scholarpedia.1883.

Gustina, S., Fadlil, A. and Umar, R. (2016) 'Identifikasi Tanaman Kamboja menggunakan Ekstraksi Ciri Citra Daun dan Jaringan Syaraf Tiruan', 2(1), pp. 128–132.

Han, J., Kamber, M., and Pei, J. (2012) Data Mining. Concepts and Techniques, 3rd Edition.

Hwang, W., and Wen, K.-W. (1998) 'Fast kNN Classification Algoritma Based On partial Distance Search', Electronics Letters, pp. 3–4.

Karomi, M. A. A. (2015) 'Optimasi Parameter K Pada Algoritma KNN Untuk Klasifikasi Heregistrasi Mahasiswa Program Studi Teknik Informatika STMIK Widya Pratama', Information Processing and Management, p. IC-TECH X (285): 5.

Kasim, A. A. and Harjoko, A. (2014) 'Klasifikasi Citra Batik Menggunakan Jaringan Syaraf Tiruan Berdasarkan Gray Level Co- Occurrence Matrices ( GLCM )', Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta, 21 Juni 2014, pp. 7–13. Available at: http://jurnal.uii.ac.id/index.php/Snati/article/viewFile/3256/2936.

Li, X., Jiang, H. and Yin, G. (2014) 'Detection of Surface Crack Defect on Ferrite Magnetic'.

McAndrew, A. (2004) An Introduction to Digital Image Processing with Matlab. Australia: Thomson.

Nixon, M. S., and Aguado, A. S. (2008) 'Feature Extraction & Image Processing (Second)', Elsevier B.V.

Pardosi, I. (2016) 'Salt and Pepper Noise Removal with Spatial Median Filter dan Adaptive Noise Reduction', 17(2), pp. 127–136.

Sikki, M. H. (2009) 'Pengenalan Wajah Menggunakan KNearest Neighbor dengan Proses Transformasi Wavelet'.

Sokolova, M. and Lapalme, G. (2009) 'A systematic analysis of performance measures for classification tasks', Information Processing and Management. Elsevier Ltd, 45(4), pp. 427–437. DOI: 10.1016/j.ipm.2009.03.002.

Tapsell LC1, Hemphill I, Cobiac L, Patch CS, Sullivan DR, Fenech M, Roodenrys S, Keogh JB, Clifton PM, Williams PG, Fazio VA, I. K. (2014) 'Health benefits of herbs and spices: the past, the present, the future', Journal Of the Australian Medical Association, 185(21 Agustus 2006).

Zhao, Y., Qian, Y. and Li, C. (2018) 'Improved KNN text classification algorithm with MapReduce implementation', 2017 4th International Conference on Systems and Informatics, ICSAI 2017, 2018–Janua(Icsai), pp. 1417–1422. DOI: 10.1109/ICSAI.2017.8248509.