Classification of Apples based Leaf Using K-Nearest Neighbors and Moment Invariant Extraction

Walidini Syaihul Huda1*, Muhammad Imron Rosadi1, Kambali2, Muhammad Syah Radian3, Syihaabul Hudaa4
1Department of Informatics, Yudharta University, Pasuruan, Indonesia
2Universitas Islam Bandung
3Department of Renewable Energy Engineering, Politeknik Negeri Jember
4Institut Teknologi dan Bisnis Ahmad Dahlan
*walidini@yudharta.ac.id

Abstract. Apples are one of several fruits that are widely favored and cultivated in various regions, in Indonesia many regions cultivate them and serve as the main source of livelihood [1], such as in East Java, especially in the area of Nongkojajar and Malang, there are many apple gardens and not a few of these apple plantations are used as apple picking tours. Of the many types of apples that flourish, not a few apple garden owners who plant more than two types of apples in one garden, and not infrequently also the owner is difficult to recognize one of the types of apples they planted, because they all look the same. The introduction of each type of apple aims to handle the treatment of apple plants that vary from one to another.

The leaf is one part of a crop where it has complete features in terms of shape, texture, and color. and can be used as a basis for classifying plant species [2]. The use of leaves as a basis for classification is easier than using cell or molecular biological methods [3] and each type of plant has a different leaf texture [2], [4].

1. Introduction

Apples are one of several fruits that are widely favored and cultivated in various regions, in Indonesia many areas are cultivated and used as the main source of livelihood [1], such as in East Java, especially in the area of Nongkojajar and Malang, there are many apple gardens and not a few of these apple plantations are used as apple picking tours. Of the many types of apples that flourish, not a few apple garden owners who plant more than two types of apples in one garden, and not infrequently also the owner is difficult to recognize one of the types of apples they planted, because they all look the same. The introduction of each type of apple aims to handle the treatment of apple plants that vary from one to another.

The leaf is one part of a crop where it has complete features in terms of shape, texture, and color. and can be used as a basis for classifying plant species [2]. The use of leaves as a basis for classification is easier than using cell or molecular biological methods [3] and each type of plant has a different leaf texture [2], [4].
In previous studies classification of plant species based on leaf features was carried out. Among these studies is the classification by combining features, colors, shapes and textures of leaves applying a fractal measure (lacunarity) with a probabilistic neural network (PNN) [5]. In other studies leaf extraction was also carried out, with leaf bone extraction method based on morphology gray-scale [6], classification based on leaves was also used on camellias [5], the study used morphology and leaf bones to distinguish camellia types. In previous studies also used color, shape and texture features on leaves for the retrieval system [7], in this study the analysis of color features using co-occurrence matrix and for shape features using local features and global features. As for the texture features using wavelet Gabor, another study used features of margins, shapes, and texture vectors for leaf extraction [8], in this study there are 3 methods used to classify the extraction results, namely Naïve Bayes, Decision Tree, and K-Nearest Neighbor. The study concluded that the method K-Nearest Neighbor had the highest accuracy rate with a value of 96.25%, while naïve Bayes produced an accuracy of 92.437% then a decision tree of 69.125%. There is also a study conducted by Arie Qur'an in 2012 [9], research conducted on apples using the method K-Nearest Neighbor (K-NN) with RGB extraction on a dataset of 50 data and in this study an accuracy value of 93.33% in terms of homogeneity features, for Red, Green Blue (RGB) features reached 100% and textures by 73.33%. Whereas in Halela's research in 2016 [10], researchers used the method K-Nearest Neighbor (K-NN) by extracting using a histogram, producing 90% accuracy with 90 data training and 10 data testing using apples as research objects. Research Riska, Laili, and M.I Rosadi The Classification of Mango Gadung and Manalagi using Bond of Leaf using Structure Element Method and SVM, result of accuracy is 78.5% [11].

In contrast to some previous studies, in this study will be carried out on the object image of an apple leaf with a dataset of 750 images. This is intended to assess the accuracy of the method K-Nearest Neighbor for determining the type of apple with invariant moments as its feature extraction. The dataset that will be used in this study uses 5 types of apple leaves with 150 types of apples taken from the nongkojajar apple plantation in East Java. In this study, the author will examine the method K-Nearest Neighbor for the classification of apples based on leaves with the extraction feature. moment invariant

2. Method

Data in this study is a picture of apple leaves totaling 750 leaves consisting of 5 types of apples, for data retrieval from this study was carried out by taking a leaf sample of 150 pieces from each apple type. The leaves taken from this sample are random from 3 trees in 2 different gardens, leaf sampling is done in nongkojajar trees in the Kayayebek village, Tutur Pasuruan, East Java. Names of the types of apples in this study are Alas Apples, Abang Apples, Ana Apples, Manalagi Apples, and Ijo Apples. Method proposed in this study is to use the KNN method by using variables obtained from the extraction of leaf image features. After preprocessing the image of an apple leaf, the image is then converted to gray love after it is processed into a binary image and feature extraction is performed. In the extraction feature of this leaf image using Moment Invariant extraction, then the variables obtained from the extraction are then made into training data and testing data, after that the data is classified and tested using the K-NN method. The following is an illustration of the method to be proposed.
For the evaluation method using a confusion matrix by examining the accuracy of the method $K$-Nearest neighbor. There is also a Confusion Matrix Is a table that contains the results of the classification calculation as a whole evaluation of data measured with accuracy, precision and recall.

| Predicted Class | Actual class | Class - 1 | Class - 2 |
|-----------------|--------------|-----------|-----------|
| Class 1         | True         | False     | positive  |
| Class 2         | False        | True      | negative  |

3. Result and Discussion
Dataset in this study in the form of leaves totaling 750 leaves of type 5 types of apples, Here are images of Apple leaves of each type.

Figure 2. Type of Apples (a) Alas Apples, (b) Abang Apples, (c) Ana Apples, (d) Ijo Apples, (e) Manalagi Apples

Removal Deletion of the background image of an apple leaf aims to eliminate noise from the leaf object, the following are the results of the image before preprocessing and after preprocessing.
Equation of image size of each image in the dataset. The dimensions of the image used is 800 x 400 pixels. Leaf image that has gone through the stages preprocessing like Figure 3.2 (b), will be converted into the image of gray (gray) as follows:

Image is an image that only has black and white, computationally binary images have two values, 0 for black and 1 for white as shown in Figure 3.4. Value 1 to show the object while 0 to indicate the background.

Performs feature extraction using the moment invariant method. The steps are as follows:

The first stage of feature extraction uses the Invariant moment method which is to calculate the moment \( m_{00}, m_{01}, m_{10} \) of the image using the equation:

\[
m_{ij} = \sum_x \sum_y x^i y^j a_{xy} \quad (3.1)
\]

Where \( xy \) is the height and width of the image, for \( i = 0,1,2 \ldots \) and \( j = 0,1,2 \ldots \) for this calculation sample is taken from an image from a dataset with a sample size of 3x4 pixels. From the above calculation, the result is:

\( m_{00} = 12204, m_{01} = 45036, m_{10} = 33588 \)

Next determine the central moment with the formula

\[
\mu_{ij} = ((x - x') (y - y') a_{xy} \quad (3.2)
\]
From the above calculation the value is generated:
\[
\begin{align*}
u_{11} & = -0.0175 \\
u_{12} & = 9.4452e-04 \\
u_{02} & = 0.0350 \\
u_{20} & = 0.1516 \\
u_{21} & = 0.0072 \\
u_{30} & = -0.0175 \\
u_{03} & = -0.0019 
\end{align*}
\]

The last of the moment invariant method is counts 7 hu moments which will be used as attributes in the classification process. Here is the formula from Hu's moment.
\[
\begin{align*}
\phi_1 & = \eta_{30} + \eta_{02} \\
\phi_2 & = (\eta_{20} + \eta_{03})^2 + 4\eta_{11}2 \\
\phi_3 & = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
\phi_4 & = (\eta_{30} + \eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \\
\phi_5 & = (\eta_{30} - 3\eta_{12}) (\eta_{30} + 3\eta_{12}) \{3 (\eta_{30} - \eta_{12})^2 - 3 (\eta_{21} - \eta_{03})^2\} + 3 (\eta_{21} - \eta_{03}) (\eta_{21} + \eta_{03}) \{3 (\eta_{30} - \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\} \\
\phi_6 & = (\eta_{20} + \eta_{02}) \{(\eta_{30} + \eta_{12})^2 - (\eta_{21} - \eta_{03})^2 + 4\eta_{111} (\eta_{30} + \eta_{12}) (\eta_{21} + \eta_{03})\} \\
\phi_7 & = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})(\eta_{30} + \eta_{12})2-3 (\eta_{12} - \eta_{03}2) + (3\eta_{21} - \eta_{03})(3\eta_{21} + \eta_{03}) \{3 (\eta_{30} - \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\}
\end{align*}
\]

\[
\phi_1 = 0.1516 + 0.0350; \\
\phi_2 = (0.1516 - 0.0350)^2 + 4 x (-0.0175); \\
\phi_3 = (-0.0175 + 3 x 9.4452e-04)^2 + (3 x 0.0072 - (-0.0019))^2; \\
\phi_4 = (-0.0175 + 9.4452e-04)^2 + (0.0072 + (-0.0019))^2; \\
\phi_5 = (-0.0175 - 3 x 9.4452e-04) x (-0.0175 + 9.4452e-04) x ((-0.0175 + 9.4452e-04)^2 - 3 x (0.0072 + (-0.0019))^2) + (3 x 0.0072 - (0.0019)) x (0.0072 + (-0.0019)) x (3 x (-0.0175 + 9.4452e-04)^2 - (0.0072 + (-0.0019))^2); \\
\phi_6 = (0.1516 - 0.0350) x ((-0.0175 + 9.4452e-04) 2 - (0.0072 + (-0.0019))^2) + 4 x -0.0175 x (-0.0175 + 9.4452e-04) x (0.0072 + (-0.0019)); \\
\phi_7 = (3 x 0.0072 + (-0.0175) x (-0.0175 + 9.4452e-04) x ((-0.0175 + 9.4452e-04)^2 - 3 x (0.0072 + (-0.0019))^2) + (-0.0175 - 3 x 9.4452e04) x (0.0072 + (-0.0019)) x (3 x (-0.0175 + 9.4452e-04)^2 - (0.0072 + (-0.0019))^2); \\
\end{align*}
\]

from the above simple resulting values monari invariant:
\[
\begin{align*}
M_1 & = 0.1866 \\
M_2 & = 0.0148 \\
M_3 & = 6.2983e-04 \\
M_4 & = 1.3659e-04 \\
M_5 & = 4.0546e-08 \\
M_6 & = 1.3139e-05 \\
M_7 & = -2.4760e-08 
\end{align*}
\]

Dataset used is is the result of apple leaf image extraction using the moment invariant method that produces 7 features. The attributes will be displayed in the following Table 3.1:

Dataset totaling 750 images of apple leaves, from 5 types of apple variants. 500 images are used as training data and 250 leaves images are used as testing data.

In processing in this study, data normalization was carried out before the data was processed. Normalization is done with the aim so that parameters do not occur in the calculation of the distance.
between data. To complete data normalization using formula 3.11. For example the normalization of the following data calculation

\[
\text{Normalization} = \frac{Data - \text{Min}}{Max - \text{Min}} = \frac{0.20607 - 0.17619}{0.33286 - 0.17619} = 0.19076
\]  

| No | Attribute | Type   |
|----|-----------|--------|
| 1  | Moment 1  | Numeric|
| 2  | Moment 2  | Numeric|
| 3  | Moment 3  | Numeric|
| 4  | Moment 4  | Numeric|
| 5  | Moment 5  | Numeric|
| 6  | Moment 6  | Numeric|
| 7  | Moment 7  | Numeric|
| 8  | Type_Apel | Polynominal |

### 3.1 K-Nearest Neighbor

Following the application of the K-NN method on the dataset obtained from apple leaf image extraction using the method moment invariant, and as explained above the dataset in this study has 7 attributes with 1 label.

| Attribute Name | value   |
|----------------|---------|
| M1             | 0.326365 |
| M2             | 0.252347 |
| M3             | 0.014531 |
| M4             | 0.012217 |
| M5             | 0.000157 |
| M6             | 0.010497 |
| M7             | 0.056669 |

There is no calculation is as follows:

1. Determine the value of k value of \( k \) let \( k = 2 \).

2. Calculate the distance \( k \) nearest neighbors using Euclidian Distance. The Euclidian Distance equation can be seen in the formula (2.10).

- Calculation of Euclidian Distance in data 1

\[
d_{1, d} = \sqrt{(0.389421 - 0.326365)^2 + (0.28783 - 0.252347)^2 + ... + (0.058285 - 0.056669)^2}
\]

\[
d_{1, d} = 0.10
\]

- Calculation of Euclidian Distance in data 2

\[
d_{1, d} = \sqrt{(0.389421 - 0.326365)^2 + (0.28783 - 0.252347)^2 + ... + (0.058285 - 0.056669)^2}
\]

\[
d_{1, d} = 0.10
\]
- Calculation of Euclidean Distance in the 3rd data
d1, dx

\[ \text{...} + (0.056553 - 0.056669) = 0.24 \]

- Calculation of Euclidean Distance in the 4th data
d1, dx

\[ = 0.18 \]

The calculation continues until the calculation in the 15th data

- Calculation of the Euclidean Distance in the third data 15
d1, dx

\[ = 0.50 \]

**Table 4.** Calculation results for distance

| No | Name   | Distance |
|----|--------|----------|
| 1  | d16, d1| 0.10     |
| 2  | d16, d2| 0.10     |
| 3  | d16, d3| 0.24     |
| 4  | d16, d4| 0.13     |
| 5  | d16, d5| 0.12     |
| 6  | d16, d6| 0.06     |
| 7  | d16, d7| 0.15     |
| 8  | d16, d8| 0.10     |
| 9  | d16, d9| 0.09     |
| 10 | d16, d10| 0.15  |
| 11 | d16, d11| 0.07  |
| 12 | d16, d12| 0.07   |
| 13 | d16, d13| 0.13   |
| 14 | d16, d14| 0.47   |
| 15 | d16, d15| 0.50   |

Sort the distance and determine the nearest neighbor based on the k-th minimum distance. After the distance has been found next sort or can be ranked by each calculation result such as pad table 4.6

**Table 5.** Ranking distance

| No | Name   | Distance | Ranking |
|----|--------|----------|---------|
| 1  | d16,d1| 0.10     | 5       |
| 2  | d16,d2| 0.10     | 5       |
| 3  | d16,d3| 0.24     | 13      |
| 4  | d16,d4| 0.13     | 9       |
After that, check whether there is data closeness or not from the data entered. For more details can be seen in table 3.6

**Table 6. Neighbouring neighbors**

| No | Name     | Distance | Ranking | Similarities | Category |
|----|----------|----------|---------|--------------|----------|
| 1  | d16, d1  | 0.10     | 4       | No           |          |
| 2  | d16, d2  | 0.10     | 4       | No           |          |
| 3  | d16, d3  | 0.24     | 13      | No           |          |
| 4  | d16, d4  | 0.13     | 8       | No           |          |
| 5  | d16, d5  | 0.12     | 7       | No           |          |
| 6  | d16, d6  | 0.06     | 1       | Yes          |          |
| 7  | d16, d7  | 0.15     | 10      | No           |          |
| 8  | d16, d8  | 0.10     | 4       | No           |          |
| 9  | d16, d9  | 0.09     | 3       | No           |          |
| 10 | d16, d10 | 0.15     | 10      | No           |          |
| 11 | d16, d11 | 0.07     | 2       | Yes          |          |
| 12 | d16, d12 | 0.18     | 12      | No           |          |
| 13 | d16, d13 | 0.13     | 8       | No           |          |
| 14 | d16, d14 | 0.47     | 14      | No           |          |
| 15 | d16, d15 | 0.50     | 15      | No           |          |

Use the simple majority of the nearest neighbor class as predictive value of new data. And the calculation results above can be seen in the table.

**Table 7. Neighboring neighbors Result**

| No | Name     | Distance | Ranking | Similarity | Category   |
|----|----------|----------|---------|------------|------------|
| 1  | d16, d1  | 0.10     | 4       | No         | Apple Abang|
| 2  | d16, d2  | 0.10     | 4       | No         | Apple Abang|
| 3  | d16, d3  | 0.24     | 13      | No         | Apple Abang|
| 4  | d16, d4  | 0.13     | 8       | No         | Apple Alas |
| 5  | d16, d5  | 0.12     | 7       | No         | Apple Alas |
| 6  | d16, d6  | 0.06     | 1       | Yes        | Apple Alas |
| 7  | d16, d7  | 0.15     | 10      | No         | Apples Ana |
| 8  | d16, d8  | 0.10     | 4       | No         | Apples Ana |
| 9  | d16, d9  | 0.09     | 3       | No         | Apples Ana |
After several experiments are carried out, the prediction results obtained from the classification method used. After that the next process is to calculate the confusion matrix to see the suitability of the predicted results.

| No | Name           | Distance | Ranking | Similarity | Category          |
|----|----------------|----------|---------|------------|------------------|
| 10 | d16, d10       | 0.15     | 10      | No         | Apples Ijo       |
| 11 | d16, d11       | 0.07     | 2       | Yes        | Apple Ijo        |
| 12 | d16, d12       | 0.18     | 12      | No         | Apple Ijo        |
| 13 | d16, d13       | 0.13     | 8       | No         | Manalagi Apple   |
| 14 | d16, d14       | 0.47     | 14      | No         | Apple Manalagi   |
| 15 | d16, d15       | 0.50     | 15      | No         | Apple Manalagi   |

Table 8. K-NN experiment results

| No | trial | Accuracy |
|----|-------|----------|
| 1  | $k = 1$ | 74.67%   |
| 2  | $k = 2$ | 47.47%   |
| 3  | $k = 3$ | 36.56%   |
| 4  | $k = 4$ | 36.80%   |
| 5  | $k = 5$ | 40.67%   |
| 6  | $k = 6$ | 40.13%   |
| 7  | $k = 7$ | 39.20%   |
| 8  | $k = 8$ | 36.67%   |
| 9  | $k = 9$ | 36.80%   |
| 10 | $k = 10$ | 35.25%  |

Table 9. Confusion matrix K-NN method

| pred. Apples Abang | true Apples Abang | true Apples Alas | true Apples Ana | true Apples Ijo | true Apples Manalagi | class precision |
|--------------------|-------------------|------------------|-----------------|----------------|----------------------|----------------|
|                    | 91                | 42               | 8               | 2              | 0                    | 63.64%         |
| pred. Alas Apples  | 40                | 79               | 3               | 11             | 3                    | 58.09%         |
| pred. Ana Apple    | 10                | 9                | 123             | 2              | 4                    | 83.11%         |
| pred. Apples       | 3                 | 11               | 9               | 128            | 2                    | Green83.66%    |
| pred. Apple Manalagi | 6                | 9                | 7               | 7              | 141                  | 82.94%         |

| class recall | 60.67% | 52.67% | 82.00% | 85.33% | 94.00% |

4. Conclusion

In this study it can be concluded that the application of the k-nn method for classifying apples based on leaves using feature extraction moment invariant can be done and works well, although the accuracy is slightly lower. This research resulted in an accuracy value of 74.93% for the accuracy value at $K=1$. In every research there must be flaws, and this research is far from perfect, therefore it is expected that the next research can be developed again with other methods. such as naïve bayes, neural networks, logistic regression, ID3 or C45. For extraction, you can use feature extraction or color extraction such as histogram, glcm and so on.

References

[1] S. Jatmika and D. Purnamasari, “Rancang Bangun Alat Pendeteksi Kematangan Buah Apel dengan Menggunakan Metode Image Processing Berdasarkan Komposisi Warna,” J. Ilm. Teknol. dan Inf. ASIA, vol. 8, no. 1, p. 8, 2014.
[2] M. Ab Jabal, S. Hamid, S. Shuib, and I. Ahmad, “Leaf features extraction and recognition
approaches to classify plant,” *J. Comput. Sci.*, vol. 9, no. 10, pp. 1295–1304, 2013.

[3] Wu, Bao, Xu, Y.-X. Wang, Y. Chang, and Q. Xiang, “A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network,” in *2007 IEEE International Symposium on Signal Processing and Information Technology*, 2007, pp. 11–16.

[4] H. Fu and Z. Chi, “Combined thresholding and neural network approach for vein pattern extraction from leaf images,” *IEE Proc. - Vision, Image, Signal Process.*, vol. 153, no. 6, p. 881, Oct. 2006.

[5] H. Lu, W. Jiang, M. Ghiassi, S. Lee, and M. Nitin, “Classification of Camellia (Theaceae) Species Using Leaf Architecture Variations and Pattern Recognition Techniques,” *PLoS One*, vol. 7, no. 1, p. e29704, Jan. 2012.

[6] Z. Xiaodong and W. Xiaojie, “Leaf Vein Extraction Based on Gray-scale Morphology,” *I.J. Image, Graph. Signal Process.*, vol. 2, no. December, pp. 25–31, 2010.

[7] H. Kebapci, B. Yanikoglu, and G. Unal, “Plant image retrieval using color, shape and texture features,” *Comput. J.*, vol. 54, no. 9, pp. 1475–1490, 2011.

[8] M. Elhadi Rahmani, A. Amine, and M. Reda Hamou, “Plant Leaves Classification,” no. c, pp. 75–80, 2015.

[9] A. Qur’ania, L. Karlitasar, and S. Maryana, “Analisis Tekstur Dan Ekstraksi Fitur Warna Untuk Klasifikasi Apel Berbasis Citra,” *Lokakarya Komputasi dalam Sains dan Teknol. Nukl.*, pp. 296–304, 2012.

[10] I. Ataka Halela, “Identifikasi Jenis Buah Apel Menggunakan Algoritma K-Nearest Neighbor (KNN) dengan Ekstraksi Fitur Histogram,” pp. 1–8, 2016.

[11] Y. Riska Suastika, L. Cahyani, and M. Imron Rosadi, “Klasifikasi Jenis Tanaman Mangga Gadung dan Mangga Madu Berdasarkan Tulang Daun,” pp. 41–50, 2014.