Identification of male and female nutmeg seeds based on the shape and texture features of leaf images using the learning vector quantization (LVQ)

E Satria¹, F Arnia² and K Muchtar³
¹,²,³ Departement of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia

Corresponding author: eri89rz@gmail.com

Abstract. One of the stages in the process of planting nutmeg seedlings on agricultural land is the identification of the sex of male and female nutmeg seedlings. The process of identifying nutmeg seedlings still uses manual methods, while the method is often inaccurate and inefficient. Therefore, this research discusses the methods and results of identification of male and female nutmeg seedlings based on the combined feature extraction and leaf texture using learning vector quantization (LVQ). The form feature extraction method used consists of slimmness, roundness, narrow factor, perimeter ratio and diameter, perimeter ratio at length while for texture is contrast, correlation, energy, and homogeneity. The results of the combined extraction of shape and texture features were applied as input vectors when classification using the LVQ method. Whereas LVQ aims at classification based on extraction results, by initializing LVQ parameters, namely learning rate and epoch. The change in the learning rate value on LVQ is very influential to get the percentage of the correctness of the data, the learning rate value must be between 0 to 1. The accuracy of LVQ predictions obtained in this study is 92.30%, with the best learning rate values from 0.001 to 0.09 and epoch 50.

Keyword: nutmeg, LVQ, shape feature extraction, texture feature extraction

1. Introduction
Pala Nutmeg is categorized as a two-house plant that has male and female flowers that are in different trees and is known as male nutmeg plants as well as female nutmeg plants. Female nutmegs produce fruit while male nutmegs only produce flowers but do not bear fruit [1]. Farmers find it very difficult to identify male nutmeg and female nutmeg seedlings so as to identify nutmeg seedlings must wait for fruiting nutmeg with 6-8 years period, overall out of 100 nutmeg seeds planted there are 55% female nutmeg, 40% male nutmeg and 5% hermaphrodite [1].

To identify nutmeg seedlings still using manual method, so the accuracy of data and parameters obtained is not yet maximum which can cause production losses in the process of planting nutmeg seedlings [1]. For this reason, a method is needed in identifying male and female nutmeg seedlings from an early age based on a combination of leaf shape and texture feature extraction using Learning Vector Quantization (LVQ) method, in previous studies LVQ method has been applied to identify fruit plants based on leaf images, to identify fruit plants, the features used are leaf shape and texture features with an accuracy of 82% [2]. Apart from leaf plants, LVQ has also been applied to identify
medicinal plants based on leaf texture, the extraction method used is 2-d gabor filter with an accuracy of 81\% [3].

Based on the 2 studies above, this study applies the LVQ method to identify male and female nutmeg seedlings based on the extraction of leaf image features, feature extraction methods used are shape and texture features. The method of extracting leaf shape features consists of slimness, roundness, narrow factor, ratio of perimeter and diameter, ratio of perimeter to width. While for texture is contrast, correlation, energy, and homogeneity. Feature extraction functions to get the characteristics of male and female nutmeg leaves based on the shape and texture features. The result of the combined feature and texture feature extraction is needed as an input vector when classifying using the LVQ method. The benefits of this study can provide information on differences in male and female nutmeg seedlings based on the results of combined extraction of shape and texture features using the LVQ method for classification. So that the identification process is more efficient.

2. Research method

The method applied in this study can be shown in Figure 1. The steps described are the technique of image acquisition of male and female nutmeg seedlings, preprocessing, region of interest, feature extraction of applied leaves consisting of shape, texture and process features. classification using the learning vector quantization (LVQ).

Based on Figure 1, the stages of the research method can be explained. The first stage is taking the image of the nutmeg leaf then the preprocessing process is carried out, namely improving the image quality aimed at facilitating the ROI process. the next step is extraction process using leaf features, while the leaf features to be extracted consist of shape and texture features, the shape and texture feature method used consists of slimness, roundness, narrow factor, perimeter and diameter ratio, ratio of perimeter to length and width leaf, contrast, correlation, energy, and homogeneity. After obtaining the desired feature extraction value, feature extraction value data will be trained and tested with the
learning vector quantization (LVQ) method for the process of recognizing male and female nutmeg leaf objects based on their shape and texture. At the LVQ process stage, database storage is performed for the final weight value of training results and class numbers representing each sample owner. Each sample owner will be introduced in certain classes consisting of class 1 for male data and class 2 for female data. Decision making is recognized or not recognized by comparing the new input with the final weight in the database.

2.1. Feature Extraction
In this study feature extraction used is shape feature consisting of slimness, roundness, narrow factor, perimeter and diameter ratio, perimeter with length and width, and texture features consisting of contrast, correlation, energy, homogeneity. The results of the extraction of shape and texture features will be used as input vector when classification using the learning vector quantization algorithm.

2.4.1 Extraction of shape features
The type of leaf feature extraction used is [2] [9]:

2.1.1.1. Slimness

\[ \text{Slimness} = \frac{L_p}{W_p} \]  

(1)

Slimness is the ratio for the ratio between the length and width of a leaf. \( L_p \) is the length of a leaf and \( W_p \) is the width of the leaf.

2.1.1.2. Roundness

\[ \text{roundness} = \frac{4\pi A}{P^2} \]  

(2)

where \( A \) is the wide area of a leaf and \( P \) is the circumference of a leaf. \( A \) adalah wilayah/ luas area dari suatu daun dan \( P \) adalah keliling daun.

2.1.1.3. Narrow factor

\[ \text{narrow factor} = \frac{D}{L_p} \]  

(3)

Narrow factor is defined as the ratio of diameter \( D \) and length of leaf \( L_p \) factor.

2.1.1.4. Perimeter ratio of diameter

\[ \text{rp}d = \frac{P}{D} \]  

(4)

Comparison between perimeter (\( P \)) and diameter (\( D \)) is used as a leaf shape feature.

2.1.1.5. Ratio of perimeter to length and width

\[ plw = \frac{P}{(L_p + W_p)} \]  

(5)

Comparison between perimeter (\( P \)) and length (\( L_p \)) and leaf width (\( W_p \)) is used as one of the leaf shape features.

2.4.2 Extraction of texture features
The type of leaf texture extraction features used are [2] [10]:

2.4.2.1. Contrast
The contrast of a pixel and its neighbors calculated for all pixels in an image can be calculated by Equation 6 where \( P_{ij} \) is a shared probability distribution of pixel pairs with gray level \( i \) and gray level \( j \).

\[ f_1 = \sum_{i=1}^{k} \sum_{j=1}^{k} (i - j)^2 P_{ij} \]  

(6)
2.4.2.2. Correlation
Correlation is a measurement of the correlation between one pixel and its neighbors which is calculated for all pixels in the image. If \( m \) is the mean (average value) of pixels at positions \( x \) and \( y \) and \( \sigma \) is the standard deviation, then the correlation can be calculated based on equation 7.

\[
 f_2 = \sum_{i=1}^{k} \sum_{j=1}^{k} (i - j)P_{ij}
\]

(7)

2.4.2.3. Energy
Energy is the sum of the Co-occurrence matrix elements calculated based on Equation 8 where \( P_{ij} \) is a joint probability distribution of pixel pairs with gray level \( i \) and gray level \( j \).

\[
 f_3 = \sum_{i=1}^{k} \sum_{j=1}^{k} (i - j)P_{ij}
\]

(8)

2.4.2.4. Homogeneity
Homogeneity is the similarity of the Co-occurrence matrix and diagonal matrix which is calculated based on Equation 9 where \( P_{ij} \) is a joint probability distribution of pixel pairs with gray level \( i \) and gray level \( j \).

\[
 f_4 = \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{P_{ij}}{1 + (i - j)}
\]

(9)

2.2. Training of LVQ Algorithms
The following LVQ training can be done in several stages. The first stage is initialisation of initial weight (\( W_j \)), the input vector of training is 146 male nutmeg leaf images and 146 female images, with \( T \) which is the target class determined by 1 for male and 2 female classes, the training process uses learning rate values of 0.001, 0.01 , 0.1 which aims to calculate the weight and epoch 50 which serves to do repetitive learning. The next stage is determining the minimum distance from the input weight vector to the reference class vector. The shortest distance is calculated using euclidean distance.

2.3. Testing of LVQ Algorithm
After training, the final weights (\( W_j \)) will be obtained. These weights are then used to carry out simulations or tests. Each weight from each class that has been trained is collected in the matrix \( W_j \). Then the weight of the test data is calculated the shortest distance to each weight of each class. The shortest distance between input weights and target class weights is the output of the test system. In this study, the tests carried out amounted to 286 images of nutmeg leaves consisting of 143 images of male leaves and 143 images of female leaves.

3. Results and Discussion

3.1. Shape Feature Extraction Results
To calculate the value of feature extraction from male nutmeg leaves and female nutmeg using equation 1-5. Form feature extraction methods consist of slimness, roundness, narrow factor, ratio of perimeter and diameter, perimeter with length and width. The shape feature extraction results were displayed using 5 samples of male nutmeg leaves and 5 different female nutmeg samples, from the total image of extracted nutmeg leaves amounted to 578 consisting of 289 images of female nutmeg seedling leaves and 289 male nutmeg images. The results of the features of female nutmeg leaf shapes are shown in Table 1 and male nutmeg leaves in Table 2.
Table 1. The result of the shape extraction feature of female nutmeg leaves

| Female Leaf | Slimness | Roundness | Narrow Factor | RPD  | RLW  |
|-------------|----------|-----------|---------------|------|------|
| citra 1     | 1.7244   | 0.4859    | 0.3365        | 4.507| 0.9599|
| citra 2     | 1.3247   | 0.4893    | 0.3337        | 4.911| 0.8541|
| citra 3     | 1.0459   | 0.4781    | 0.3377        | 4.5435| 0.7844|
| citra 4     | 1.1438   | 0.4837    | 0.3381        | 4.5173| 0.8149|
| citra 5     | 1.0597   | 0.4817    | 0.3374        | 4.5264| 0.7858|
| Average     | 1.2597   | 0.4837    | 0.3366        | 4.5176| 0.8398|

Based on Table 1, the average slimness values obtained from 5 samples of female nutmeg leaf images were 1.2597, roundness 0.4837, narrow factor 0.3366, perimeter and diameter ratio 4.5176, and perimeter ratio with length and width 0.8398. While the highest value obtained is 1.7244 slimness, roundness 0.4893, narrow factor 0.3337, perimeter and diameter ratio 4.5435, and perimeter ratio with length and width 0.9599. The lowest value of slimness is 1.0459, roundness is 0.4781, narrow factor is 0.3337, the ratio of perimeter and diameter is 4.4911, and the ratio of perimeter to length and width is 1.7844.

Table 2. The result of the shape extraction feature of male nutmeg leaves

| Male Leaf | Slimness | Roundness | Narrow Factor | RPD  | RLW  |
|-----------|----------|-----------|---------------|------|------|
| citra 1   | 1.7407   | 0.5291    | 0.4477        | 4.3192| 1.228|
| citra 2   | 2.863    | 0.5336    | 0.5411        | 4.3005| 1.7247|
| citra 3   | 1.5685   | 0.5005    | 0.4517        | 4.4409| 1.225|
| citra 4   | 1.7759   | 0.5152    | 0.4519        | 4.3768| 1.2654|
| citra 5   | 1.6944   | 0.5135    | 0.4517        | 4.3842| 1.2452|
| Average   | 1.9285   | 0.5183    | 0.4688        | 4.3643| 1.3376|

Based on Table 2 the average slimness value obtained from 5 samples of male nutmeg leaf image is 1.9285, roundness 0.5183, narrow factor 0.4688, perimeter ratio and 4.3643 diameter, and perimeter ratio with length and width of 1.3376. While the highest value obtained is 2.863 slimness, roundness 0.5336, narrow factor 0.5411, perimeter and diameter ratio 4.4409, and perimeter ratio with length and width 1.7247. The lowest value of slimness is 1.5685, roundness 0.5005, narrow factor 0.4477, perimeter and diameter ratio 4.3005, and the ratio of perimeters to length and width of 1.225.

3.2. Texture Feature Extraction Results

To calculate the value of the texture feature extraction from male nutmeg leaves and female nutmeg using equations 6-9. Texture feature extraction method consists of contrast, correlation, energy and homogeneity. The texture feature extraction results are shown using 5 samples of male nutmeg leaves and 5 samples of different female nutmegs, from the total image of extracted nutmeg leaves totaling 578 consisting of 289 images of female nutmeg seedlings and 289 images of male nutmeg. The results of the texture features of female nutmeg leaves are shown in Table 3 and male nutmeg leaves in Table 4.
Table 3. The result of the texture extraction feature of female nutmeg leaves

| Female Leaf | Contrast | Correlation | Energy | Homogeneity |
|-------------|----------|-------------|--------|-------------|
| citra 1     | 0.0016   | 0.9465      | 0.8671 | 0.9929      |
| citra 2     | 0.0015   | 0.9453      | 0.8663 | 0.9948      |
| citra 3     | 0.0032   | 0.9493      | 0.8645 | 0.9955      |
| citra 4     | 0.0033   | 0.9439      | 0.864  | 0.9949      |
| citra 5     | 0.0027   | 0.9436      | 0.8636 | 0.9951      |
| **Average** | **0.024**| **0.9457**  | **0.8651** | **0.9946** |

Based on Table 3 the average contrast value obtained from 5 samples of female nutmeg leaf image is 0.024, correlation 0.9457, energy 0.8651, and homogeneity 0.9946. While the highest value obtained contrast is 0.0033, correlation 0.9493, energy 0.8671, and homogeneity 0.9955. The lowest value of contrast is 0.0015, correlation 0.9436, energy 0.8636, and homogeneity 0.9929.

Table 4. The result of the texture extraction feature of male nutmeg leaves

| Male Leaf | Contrast | Correlation | Energy | Homogeneity |
|-----------|----------|-------------|--------|-------------|
| citra 1   | 0.0038   | 0.9897      | 0.762  | 0.9988      |
| citra 2   | 0.0039   | 0.9912      | 0.7562 | 0.9989      |
| citra 3   | 0.0042   | 0.9883      | 0.7552 | 0.9987      |
| citra 4   | 0.0041   | 0.9881      | 0.7101 | 0.9985      |
| citra 5   | 0.004    | 0.9905      | 0.7561 | 0.9986      |
| **Average** | **0.004**| **0.9895**  | **0.7479** | **0.9987** |

Based on Table 4 the average contrast value obtained from 5 samples of male nutmeg leaf image is 0.004, correlation 0.9895, energy 0.7479, and homogeneity 0.9987. While the highest value obtained contrast 0.0042, correlation 0.9912, energy 0.762, and homogeneity 0.9989. The lowest value of contrast is 0.0038, correlation 0.9881, energy 0.7101, and homogeneity 0.9987.

The results of the combined extraction of the features of the shape and texture of male and female nutmeg leaves shown in Tables 1, 2, 3 and 4 function for input vectors during the training process and testing using the LVQ algorithm in identifying male and female nutmeg seeds based on shape characteristics and leaf texture.

3.3. LVQ Test Results

Test were carried out with 286 data testing samples consisting of 143 male data samples and 143 female data samples. Each test data is carried out the process of calculating the shortest distance or smaller than one with the final weight of the results of the training that will be the winner. In the testing process, a learning rate setting was also conducted which aims to obtain the classification results with the optimal amount of accuracy. Table 5 shows the results of the testing process for LVQ testing data aimed at updating the weights in each class. The renewal of the weight value is influenced by several LVQ compiler parameters such as learning rate and number of iterations. In the LVQ training process a number of different learning rate values are arranged with a fixed number of epochs.

Table 5. LVQ Test results

| No | LR  | epoch | Data amount | Male Prediction | Female Prediction | Accuracy | Time (minute) |
|----|-----|-------|-------------|-----------------|-------------------|----------|---------------|
| 1  | 0.001 | 50    | 286         | True | False | 137 | 127 | 16 | 92.30% | 1.11 |
| 2  | 0.01  | 50    | 286         | 137  | 6     | True | False | 127 | 16 | 92.30% | 1.11 |
| 3  | 0.1   | 50    | 286         | 136  | 7     | False | True | 125 | 18 | 91.60% | 1.09 |
Pada In Table 5 explains each change in learning rate can affect the amount of truth testing data classification, time and level of accuracy. The best results from the testing process can be shown at LR 0.001 and 0.01 with the percentage level of truth is 92.30%, total data tested as many as 286 consisting of 143 male data tested, 137 predicted correctly, and 127 female data correctly predicted from 143 female data tested, with a testing time of 1.11 minutes. While the lowest accuracy is shown at LR 0.1 with a percentage of truth 91.60%, the greater the value of learning rate, the greater the area covered, but the lower the ability of observation. Tabel 5.

4. Conclusion
Based on the research that has been done, it can be concluded that the application of the Learning Vector Quantization algorithm (LVQ) method has very good accuracy in identifying the leaves of male and female nutmeg seedlings. The feature extraction method used is the form feature extraction consisting of slimmness, roundness, narrow factor, perimeter and diameter ratio, perimeter with length and width, and texture features consisting of contrast, correlation, energy and homogeneity.

The identification stages of 578 samples extracted from male and female nutmeg leaves with the application of the LVQ classification method were carried out in 2 stages, the first stage conducting data training using 292 data samples, then testing the data using 286 data samples. Changes in the learning rate in LVQ are very influential to get the percentage of the data truth level. Based on LVQ testing it can be concluded that the large learning rate value of training process time is fast, but the results of the accuracy obtained are low, on the contrary if the learning rate value is small then the accuracy of high accuracy and training time is a little slow. The highest accuracy of LVQ prediction is 92.30% with a fast testing time of 0.013 seconds, while the best LVQ parameters are at the learning rate values of 0.001 and 0.01 and epoch 50.

References

[1] N. B. S.Wahyuni, “Deteksi Tanaman Pala Jantan Dan Betina Secara Dini ,” Warta. Penelitian. dan Pengembangan. Tanaman. Industri., vol. 21, no. 2, pp. 28–31, 2015.
[2] R. F. Abdullah and R. Passarella, “Identifikasi Tanaman Buah Berdasarkan Fitur Bentuk, Warna dan Tekstur Daun Berbasis Pengolahan Citra dan Learning Vector Quantization (LVQ),” Prosiding Annual Research Seminar 2017 Computer Science and ICT, 2017, vol. 3, no. 1.
[3] R. R. Isnanto, el all., “Identifikasi Tanaman Obat Menggunakan Tapis Gabor 2-D Dengan Jaringan Saraf Tiruan Learning Vector Quantization (LVQ),” in Proceedings Seminar Nasional Teknik Elektro (FORTEI 2016), 2016
[4] H. Setiawan and E. M. Yuniarno, "Biometric Recognition Based on Palm Vein Image Using Learning Vector Quantization," 2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), Bandung, 2017, pp. 95-99
[5] A. Kadir, L.E. Nugroho, A. Susanto and P. I. Santosa, “Leaf Classification using Shape, Color, and Texture Features”, International Journal of Computer Trends and Technology, vol. 1, no. 3, 2011
[6] Y. A. Sari, R. K. Dewi, and C. Fatichah, “Seleksi Fitur Menggunakan Ekstraksi Fitur Bentuk, Warna, Dan Tekstur Dalam Sistem Temu Kembali Citra Daun,” JUTI J. Ilm. Teknol. Inf., vol. 12, no. 1, p. 1, 2014.
[7] R. K. D. Nanik Suciati, “Klasifikasi Tanaman berdasarkan Fitur Bentuk dan Tekstur pada Daun menggunakan Decision Tree,” J. Cybermatika, vol. 3, no. 2, pp. 9–15, 2015.