# Identification of Rice Variety Using Geometric Features and Neural Network

# Wahyu Srimulyani<sup>\*1</sup>, Aina Musdholifah<sup>2</sup>

<sup>1</sup>Master's Program of Computer Science, FMIPA UGM, Yogyakarta, Indonesia <sup>2</sup>Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia e-mail: \*<sup>1</sup>wahyu.sri.m@mail.ugm.ac.id, <sup>2</sup>aina\_m@ugm.ac.id

## Abstrak

Indonesia memiliki banyak varietas pangan salah satunya adalah varietas beras. Masing-masing varietas beras memiliki ciri fisik yang dapat dikenali melalui warna, tekstur, dan bentuk. Berdasarkan ciri fisik tersebut, beras dapat diidentifikasi menggunakan pendekatan Neural Network. Penelitian dengan menggunakan 12 fitur memiliki hasil yang belum optimal. Penelitian ini mengusulkan penambahan fitur geometri dengan algoritma Learning Vector Quantization dan Backpropagation yang digunakan secara terpisah.

Pecobaan menggunakan data 9 varietas beras yang diambil dari beberapa daerah di Yogyakarta. Akuisisi beras dilakukan menggunakan kamera Canon D600 dengan lensa kit dan perbesaran maksimum 55 mm. Pembagian data dilakukan untuk pelatihan dan pengujian serta dilakukan pembagian data sesuai dengan kualitas bentuk beras. Praproses data dilakukan sebelum ekstrasi fitur dengan proses segmentasi metode thresholding trial and error. Evaluasi dilakukan dengan membandingkan hasil penambahan 6 fitur geometri dan sebelum dilakukan penambahan fitur geometri.

Hasil pengujian menujukan bahwa penambahan 6 fitur geometri memberikan kenaikan nilai akurasi. Hal tersebut dibuktikan dengan algoritma Backpropagation menghasilkan kenaikan akurasi sebesar 100% dan algoritma LVQ 5,2%.

Kata kunci—Image Processing, Rice Varieties, Geometric Feature, Learning Vector Quantization, Backpropagation

## Abstract

Indonesia has many food varieties, one of which is rice varieties. Each rice variety has physical characteristics that can be recognized through color, texture, and shape. Based on these physical characteristics, rice can be identified using Neural Network. Research using 12 features has not optimal results. This study proposes the addition of geometry features with Learning Vector Quantization and Backpropagation algorithms that are used separately.

The data use in this study are 9 rice varieties that taken from several regions in Yogyakarta. The acquisition of rice was carried out using a camera Canon D600 with a kit lens and maximum magnification, 55 mm. Data sharing is carried out for training and testing, and the training data was sharing with the quality of the rice. Preprocessing of data was carried out before feature extraction with the trial and error thresholding process of segmentation. Evaluation is done by comparing the results of the addition of 6 geometry features and before adding geometry features.

The test results show that the addition of 6 geometry features gives an increase in the value of accuracy. This is evidenced by the Backpropagation algorithm resulting in increased accuracy of 100% and 5.2% the result of the LVQ algorithm.

Keywords—Image Processing, Rice Varieties, Geometric Feature, Learning Vector Quantization, Backpropagation

## 1. INTRODUCTION

Rice is the most favorable and most consuming food for human being in all over the world and researchers are working to improve the quality of rice. The quality measurement of rice is also important because it is consumed as food as well as it is used for milling process in the national and international market [1]. Based on the 2009 National Socio-Economic Survey in 2013, rice (Oryza sativa L.) is a staple food consumed by per capita households in Indonesia. The variety and quality of rice are one of the considerations for the community to consume. A large number of rice varieties that have emerged in Indonesia has made it difficult for the Indonesian market to determine the quality and variety of rice. The consumer concern on the originality of rice variety and the quality of rice leads to originality certification of rice by existing institutions. As a consequence, there should be an evaluation method for the originality of a product which is able to perform identification [2]. This is the basis for the emergence of rice certification carried out by several institutions as one of the reasons for rice grouping and quality control of rice. Control is carried out by human vision by bringing in trained inspectors.

Rice varieties can be identified by physical characteristics. These characteristics include color, shape, and texture of rice. In many researches, external features such as morphology, color and texture have been combined to modify the accuracy of grain classification [3]. These three physical characteristics can be used as parameters in processing the image of rice to determine the variety of rice. Image processing is one way to identify rice. Image processing of rice is carried out on rice images with the .jpg format which can then be carried out to take the color values of rice, namely by operating RGB and HIS color images, shapes with image morphology processes, and textures obtained from image histograms. Colour and texture features are used to develop a neural network model for the classification [4].

Rice identification research was previously conducted by Sumaryanti [2]. The results obtained were accuracy on 50% geometric features. The effect that occurs is the identification of rice varieties has not produced a high degree of accuracy. The features used in the study were 12 features. The identification method also supports the success of the accuracy of the identification of rice. Identification was carried out using the Learning Vector Quantization method.

Research on rice identification using image-based Neural Network was also conducted by Zhao (2005) by giving 14 geometric features. The accuracy of the study was 90% with identification using Backpropagation. The two studies above are the basis for researchers to conduct research on the identification of rice varieties. Research on the surface by adding geometric features to the research of [2]. Geometry features are taken from research conducted by Zhao (2005). The algorithm used in this study refers to the two studies above. The algorithms that use in this research are Learning Vector Quantization and Backpropagation.

The addition of geometric features in this study, based on research [1]. The results obtained using geometry features still have low accuracy. The number of geomaterial features in the study [1] are 4. Backpropagation and Learning Vector Quantization were used in this study on the basis of the two algorithms applied in previous studies. Backpropagation was used in Zhao (2005) 's study of Identification of rice seed varieties using neural network. The study used 14 geometry features and succeeded in providing an accuracy of 95%. Learning Vector Quantization is used in research [1] using 4 geometry features and produces an accuracy of 70%. This study took geometry features and Backpropagation algorithm in Zhao (2005) study, then applied according to the research [1] using 6 color features, 2 texture features, and 4 geometry features with Learning Vector Quantization algorithm.

#### 2. METHODS

Research begins with the data acquisition process. Image data from the acquisition results are carried out by dividing test data and training data. The test data are dividing which based on the quality of rice.

The test data that has been sharing given to the pre-process data. Pre-process data is carried out to produce more optimal feature extraction. The next step is extract feature to get the value that will be used as identification input. The results of the mixed 1: 1 data feature extract are used as algorithm architecture test data. The algorithms used for identification are Backpropagation and Learning Vector Quantization. The results of the architectural experiments were analyzed and then the best accuracy was taken to further be used as training for other test data. Whereas for the test data, pre-process data is carried out, then continued extracting features and finally testing.



#### 2.1 Data Acquisition

The data used in this study are obtained in various regions in Yogyakarta. Rice used amounts 9, there are mentik wangi, mentik susu, pandan wangi, raja lele, batang lembang, C4, cianjur, ciherang, and mekongga. The total data set used is 900 data.

The image is obtained shooting using a Canon 600D and standard lens. The camera is put on a tripod with a slope of  $45^{\circ}$  and a maximum magnification of 55mm with flash. Each rice seed is placed in a box with a lamp on one side of the box. The position of the lamp is opposite the position of the camera.



Figure 2. Image acquisition using box scheme

#### 2. 2 Data Sharing

Data is divided into training data and test data, 70% training data and 30% test data. The next stage is training data divided into 5 parts to be used in the K-Fold Cross Validation process. The sample data are 900 data divided into two parts with details 630 as training data and 270 as test data. The training data is divided into 5 parts to use the validation process using K-Fold Cross Validation so that there are 126 data in which there are 14 data on each rice variety.



Figure 3. Illustration of rice data sharing

The whole data for each variety is 100. In accordance with Figure 3 the data distribution is carried out for each variety of training data as much as 70 and for the test data 30, the details of the amount are presented in table 2.1.

Table 1. Distribution of training data and test data							
No	Variety	Training data	Testing data				
1	Mentik Wangi	70	30				
2	Mentik Susu	70	30				
3	Pandan Wangi	70	30				
4	Raja Lele	70	30				
5	Batang Lembang	70	30				
6	C4	70	30				
7	Cianjur	70	30				
8	Ciherang	70	30				
9	Mekongga	70	30				
	TOTAL	630	270				

Table 1. Distribution	n of training	data	and	test	data

The training data from each variety is divided into whole rice data and data on defective rice. Good rice, which is meant as a whole or not having a defect in shape and color, while for defective rice is the opposite.





(a) (b) Figure 4. Mentik wangi (a) Whole rice (b) Defective rice

The training data from each variety is divided into whole rice data and defective rice data. The total amount of defective rice data is 210 and whole rice is 420. Mentik wangi, mentik susu, and mekongga have a number of defective rice data each 30 and the amount of whole rice each is 40. Pandan wangi rice, raja lele, batang lembang, C4, ciajur, and ciherang have a

number of defective data each 20, while the amount of whole rice each is 50.

#### 2. 3 Pre Process Data

The image data obtained by preprocessing data is to obtain a more optimal feature extraction value. First step is converting RGB color images to gray scale image. The purpose of this step is make the image simpler, and reduce the complexity of code [5]. The segmentation process is carried out using the trial and error thresholding method. Among all the methods of image segmentation, thresholding is the simplest, yet most important and useful way of portioning an image into a front part and a background [6]. RGB images are converted to grayscale images then thresholding with a 120 threshold value. Objects attached to the border and objects with an area less than 50000 are deleted. Taking the RGB value then changes the background color to black / zero value and recombines the RGB value and looks for objects with contour tracking and finally crops on the image of rice.



## 2. 4 Feature extraction

The feature extraction process is imposed on the segmented image. In this study, the characteristics are used as a previous study conducted by [2]. There are 12 features consisting of 6 color features, 2 texture features, and 4 shape features.

Color features include red mean, gree mean, blue mean, hue mean, saturation mean, and intensity mean. Texture features are obtained by statistical calculations based on grayscale image histograms, there are mean and standard deviation. Morphological/shape features include 4 features including area, perimeter, physiological length, and physiological width.

There are few commonly used color spaces such as the RGB space, the HSV space, the HIS space and the XYZ space [7]. Morphological is agenerally known as branches in biological sciences that discuss ofshape or structure of animals and plants. In the context of digital imagery, morphology is a way to extract image components that are useful in form representations and descriptions such as boundaries, skeletons and convex hulls [8].

An additional feature that researchers propose is that geometry features include thinness ratio, aspect ratio, equivalent diameter, convect area, solidity, and extent. Thinness ratio is used to determine roundness and rice. Acpect ratio to find out the ratio of width and length of rice. Equivalent diameter is the diameter of a circle that is equivalent to that object. The Convex area is an area of outer rice polygons which is limited by an ellipse that has the same shape and size as the rice. Solidity is the proportion between the area of rice and convex area. Extent comparison of area with the area outside the area of the object or frame.

1. Perimeter

Perimeter or circumference indicates the length of the edge of an objek [9]. The perimeter of rice grains is calculated by counting the number of pixels containing the edges of the grain of rice. The perimeter can be illustrated as in Figure 6.



Figure 6. Examples of Perimeter Features

2. Area

The number of pixels inside and including the limit [9], multiplied by the calibration factor (mm2 / pixel).

3. Physiological length

The distance between the end point of the longest line of the object, or the lengthwise diameter of an object. The major axis length is also called the object diameter which can be calculated using the brute force method. The major axis length feature can be illustrated as in Figure 7.



Figure 7. Example of Physiological length

4. Physiological width

The distance between the end of the longest line that can be drawn from an object with a perpendicular axis, or a short diameter of an object [9]. The minor axis length can be illustrated as shown in figure 8.



Figure 8. Example of Physiological width

## 5. Thinnes ratio

Size of roundness of object, which can be obtained by function,

$$T = 4\pi \frac{A}{P^2} \tag{1}$$

T = Thinnes ratio

- A = Area
- P = Perimeter.

6. Aspect ratio

The aspect ratio is the ratio between the object's width and the object's height. Two points facing and are members of the perimeter set of objects with the shortest distance being the width of the object. While the height of the object is obtained from a line that is perpendicular to the line width of the object and connects the two points of the perimeter set of the farthest object.

7. Equivalent diameter

The diameter of a circle with an area equal to the object area can be illustrated as shown in Figure 9 with equation 2.



Figure 9. Example of equivalent diameter

$$d_{s} = \sqrt{\frac{A}{\pi}} \tag{2}$$

- $d_e$ = Equivalent Diameter A = Area  $\pi = 3,1415926535$
- 8. Convect area

The Convex area is a measure of how convex the object is. The convex area value is obtained from the perimeter length (edge / circumference) multiplied by the calibration factor (mm2 / pixel).



Figure 10. Ilustration of convect area

9. Solidity

Solidity is a measure of the comparison of area or area of an object compared to the area of convex that surrounds the object.

Solidity = 
$$\frac{\text{Area}}{\text{Convex Area}}$$
 (3)

10. Extent

The proportion of pixels in the bounding box that are also in the seed region, calculated as the area divided by the area of the bounding box.

$$Extent = \frac{Area}{Area of the bounding}$$
(4)

## 2. 5 Identification of Varieties Using Backpropagation

Backpropagation is done in two stages. First is Feed-forward, which is the pattern training process that will set to each unit in the input layer, then output the one generated is transmitted to the next layer, continue until the output layer. Second is backpropagation, which is the process of adjusting each weight based on the expected output, to be produced minimal

error, starting from the weight connected to the output neuron, then continue to retreat until to the input layer [10].

Input layer is a value from the feature extraction. The number of input layer neurons are 18 according to the number of features extracted. Hidden layers as many as 100 neurons with sigmoid binner activation function. The output layer has a number of neurons of rice varieties, there are 9. The activation function used in the output layer is linear. Data from feature extraction becomes input.



Figure 11. Diagram block of identification of rice varieties using backpropagation

## 2. 6 Identification of Varieties Using Learning Vector Quantization

Learning vector quantization is the method to perform classification and training on the input and target vectors given by the user. The prototype based algorithm comprises of competitive and linear layers. Competitive layer learns to classify the input vectors and further this layer's classes are altered to classify targets defined by the user known as target classes. The network is created on the basis of necessary parameters. [11]. The number of input neurons is 18 according to the number of feature extractions. Hidden size that will be used more than 10 and less than 25. The number of neurons in the output layer is 9 according to the number of varieties of rice.



Figure 12. Diagram block of identification of rice varieties using Learning Vector Quantization

## 2. 6 Testing

The stages of the testing process include the image of rice carried out pre-process data and then extracting features. The results of feature extraction were tested with several models from the previous training. To see the increase in the accuracy of the results of rice identification in this study compared with the previous study [2] used the equation as 5.

$$Percentage \ increase = \frac{end \ value - starting \ value}{starting \ value} \ x \ 100 \ \%$$
(5)

## 3. RESULTS AND DISCUSSION

Testing is done with data that has been separated from the training data. The number of test data is 270, with each class having 30 data. Testing for each algorithm is done using 8 networks, namely 4 networks for 12 features and 4 networks for 18 features. The first network uses mixed data networks randomly, the second uses mixed 50: 50 networks, the third uses the whole rice data network, and the fourth uses the disabled rice data network.

3. 1 The Result of Training

From the results of the training, an increase in accuracy was obtained as presented in table 2. Backpropagation algorithm in all data conditions results in increased accuracy. Decrease in accuracy value is obtained when training uses mixed data 50: 50 and the algorithm used is LVQ. In good rice data, the Backpropagation and LVQ algorithms provide the highest increase in accuracy. The average accuracy of training results, Backpropagation showed 28,8% and LVQ 25,5%.

	Results of training accuracy (%)							
Algorithm	Mixed	data	Good	Defective	Averege			
	Random 50:50		rice	rice	Average			
Backpropagation	23,8	2,5	68,9	19,94	28,8			
LVQ	3,6	-32,4	103,5	27,1	25,5			

Table 2. Training result

## 3. 2 The Test Results Use The Backpropagation Algorithm

Testing result of Backpropagation algorithm on randomly mixed rice network has increased accuracy by 27.9%. The network of mixed data 50:50 showed increased accuracy 138.1%, and using the whole rice data the increase obtained is 284.6%. Testing using a defective rice network has increased by 31.8%. The average increase using the Backpropagation algorithm is 100%. The increased value obtained by the initial value is the test result using 12 features and the final value using 18 features.

Table 3. Testing result of backpropagation algorithm

Network	Result of testing data (%)						
Feature	Random	andom Mixed data		Defective	Average		
	mixed	rice	rice				
12 feature	68	42	26	22	39,5		
18 feature	87	100	100	29	79		
Change in accuracy value	27,9	138,1	284,6	31,8	100		

# 3. 3 The Test Results Use The Learning Vector Quantization Algorithm

Testing uses LVQ algorithm by using a random mixed rice network showed increase accuracy 3% and when using a mixed rice 50:50 network there is a decrease in accuracy of 8.1%. In testing using utuh rice networks the increase in accuracy obtained was 8.8% and using cacat rice data gained an increase is 19.4%. The average increase using the LVQ algorithm is 5.2%. Starting value is the test result using 12 features and the final value is using 18 features.

Network	Result of testing (%)							
Feature	Random	Mixed data	Good	Defective	Average			
	mixed	50:50	rice	rice				
12 feature	33	37	34	31	33,75			
18 feature	34	34	37	37	35,5			
Change in accuracy value	3,0	-8,1	8.8	19,4	5,2			

Tabel 4. Testing result of Learning Vector Quantization algorithm

# 3. 4 Analysis of Rice Varieties in Each Class

From the results of identification using the backpropagation and LVQ algorithm, we can find out the level of difficulty in identifying each rice variety. 16 matrix confusion was obtained from the results of the identification. One example of the matrix confusion results from the test results is presented in table 5 These results are obtained from testing using the backpropagation algorithm using cacat rice data networks and 18 features.

	Mentik wangi	Mentik susu	Pandan wangi	Raja lele	lembang	C4	Cianjur	Ciherang	Mekongga	Tidak dikenali	Akurasi (%)
Mentik wangi	0	8	9	4	7	1	1	0	0	0	0
Mentik susu	0	2	11	15	2	0	0	0	0	0	7
Pandan wangi	0	1	21	5	1	0	1	1	0	0	70
Raja lele	0	0	16	11	3	0	0	0	0	0	37
Batang lembang	0	0	3	2	7	12	4	2	0	0	23
C4	0	0	0	1	7	19	3	0	0	0	63
Cianjur	0	0	0	5	16	7	2	0	0	0	7
Ciherang	0	0	1	2	1	11	6	8	1	0	27
Mekongga	0	0	0	0	0	3	10	10	7	0	23
Tidak dikenali	0	0	0	0	0	0	0	0	0	0	0
Akurasi keseluruhan									29		

 

 Table 5. Examples of confusion matrix results from testing rice using cacat rice network and Backpropagation algorithm

Mentic wangi varieties have an accuracy value of 0%. These varieties are more commonly identified as pandan wangi rice and batang lembang. Rice varieties that have the highest accuracy are pandan wangi. Data of 30 pandan rice wangi can be correctly identified by 21 and 5 data identified as raja lele varieties.

## 3. 5 Discussion

The addition of 6 geometry features carried out in this study gives an accurate accuracy. Backpropagation algorithm increase obtained 100% and the LVQ algorithm increases 5.2% on average. The geometry feature gives values according to the physical form of rice, so the weight given to the algorithm input is clearer. In the defective rice network that is also random mixed rice network used for testing has a result that is not good accuracy compared to the whole rice training network. This is because, defective rice has an imperfect shape so it cannot be identified properly.

The increase in the accuracy value of the LVQ algorithm is not as much as the Backpropagation algorithm. LVQ algorithm is very dependent or sensitive to the reference vector (reference vector) which is also used as the initial weight of the network in the LVQ training process. The reference vector is a representation of each class that represents a class in the LVQ network. The determination of the reference vector that is used is to directly select a number of input vectors as representatives of each class in the training process. The initialization of the reference vector (initial weight) in this way is very sensitive to the level of success because of the inaccuracy in the selection.

## 4. CONCLUSIONS

From the research that has been done the following conclusions are obtained. The addition of the geometric features performed can provide higher accuracy. The Backpropagation algorithm can work better with an average increase of 100% accuracy. The LVQ algorithm produces an average increase of 5.2%.

Testing using a good rice data network produced from training provides the highest results of increased accuracy. The increase obtained from the test was 284.6%. This is because the form of rice used is intact so the geometry feature provides better value compared to the data of rice with an incomplete form.

## REFERENCES

- [1] Z Parveen, M.A.Alam, and H.Shakir, "Assessment of Quality of Rice Grain Using Optical and Imgae Processing Technique", International Conference on Communication Computing and Digital Systems, 2017 [Online]. Available : https://ieeexplore.ieee.org /document/7918940. [Accesed: 25-July-2019]
- [2] L. Sumaryanti, A.Mushdolifah, and S. Hartati, "Digital Image Based Identification of Rice Variety Using Image Processing And Neural Network", Indonesian Journal of Electrical Engineering, Vol.16, No.1, P.182-190, 2015 [Online]. Available : https://www.semanticscholar.org/paper/Digital-Image-Based-Identification-of-Rice-Variety-Sumaryanti-Musdholifah/82fe813a73069e4433ccdabc692b4fef77b8376f. [Accesed: 20-July-2019]
- [3] A.R.Pazoki, F.Farokhi, and Z. Pazoki, "Classification of Rice Grain Varieties Using Two Artificial Neural Networks (MLP and Neuro-Fuzzy)", The Journal of Animal & Plant Sciences. Vol 24. No 1. p. 336-343, 2014 [Online]. Available https://www.semanticscholar.org/paper/Classification-of-rice-grain-varieties-usingtwo-Pazoki-Farokhi/9c7908f327b76b1818d4f9123d4d27d98fe2424b. [Accessed: 20-July-2019]
- [4] I.Golpur, J.A.Parian, and R.A.Chayjan, "Identification and Clasification of Bulk Paddy, Brown, and White Rice Cultivars with Colour Features Extraction using Image Analysis and Neural Network", Czech J.Food Sci, Vol. 32, No.3, P.280-287, 2014 [Online].

Available : https://www.agriculturejournals.cz/publicFiles/238\_2013. [Accessed: 20-July-2019]

- [5] N.Nagoda and L.Ranathunga, "Rice Sample Segmentation and Classification Using Image Processing and Support Vector Machine", International Conference on Industrial and Information System (ICIIS), 13<sup>th</sup>, 2018 [Online]. Available : https://ieeexplore.ieee.org /document/8721312. [Accessed: 25-July-2019]
- [6] M.A.B.Siddique, R.B. Arif, and M.M.R. Khan, "Digital Image Segmentation in Matlab: A Brief Study on OTSU's Image Thresholding", International Conference on Innovation in Engineering and Technology (ICIET), 2018 [Online]. Available : https://ieeexplore.ieee.org/document/8660942. [Accessed: 25-July-2019]
- [7] Z.Youlian, H. Cheng, Z. Kun, and P. Lingjiao, "Face Detection method using template feature and skin color feature in RGB color space", The 27<sup>th</sup> Chinese Control and Decision Conference, 2015 [Online]. Available : https://ieeexplore.ieee.org /document/7161913. [Accessed: 25-July-2019]
- [8] C.S. Nandi, B. Tudu, and C. Koley ,"A Machine Vision-Based Maturity Prediction System for Sorting of Harvested Mangoes", IEEE Transactions on Instrumentation and Measurement., Vol.63, No.7, 2014 [Online]. Available:https://ieeexplore.ieee.org/document/6730653. [Accessed: 25-July-2019]
- [9] K. Syaban and A. Harjoko, "Klasifikasi Varietas Cabai Berdasarkan Morfologi Daun Menggunakan Backpropagation Neural Network", Indonesian Journal of Computing and Cybernetics Systems (IJCCS), Vol.10, No.2, 2016 [Online]. Available : https://jurnal.ugm.ac.id/ijccs/article /16628/11686. [Accessed: 27-July-2019]
- [10] Ardiansyah and E. Rainarli, "Implementasi Q-Learning dan Backpropagation pada Agen yang Memainkan Permainan Flappy Bird", Jurnal Nasional Teknik Elekto dan Teknologi Informasi (JNTETI), Vol.6, No.1, 2017 [Online]. Available: http://ejnteti.jteti.ugm.ac.id/index.php/JNTETI/article/287/216. [Accessed: 27-July-2019]
- [11] S.K. Khatri, S. Dutta, P.Johri, "Recognizing images of handwritten digits using learning vector quantization artificial neural network", International Conference on Reliability Infocom Technologies and Optimization (ICRITO), Vol.10, 2015 [Online]. Available: https://ieeexplore.ieee.org/document/7359307. [Accessed: 27-July-2019]