Classification of Dried Chilli Quality Using Image Processing

Y A Azis, N Khuriyati* and A Suyantohadi
Department of Agroindustrial Technology, Universitas Gadjah Mada, 1st Flora Street, Bulaksumur, Yogyakarta, Indonesia

*E-mail : nafis.khuriyati@ugm.ac.id

Abstract. Chillies are a high-value export commodity. One of the chilli products that has a high export value, such as to the European market, is dried red chillies. Dried chillies' quality is the main parameter that must be maintained if this commodity is to be continuously exported. This study aimed to precisely and accurately classify the quality of dried chillies using physical parameters of the length and color of chillies based on digital image processing. In this research, the quality of dried red chillies was classified using a combination of digital image processing and artificial neural networks (ANN). The quality parameters of dried chillies used as inputs were chilli length, mean energy, mean a*, mean blue, and mean contrast. This study used 150 dried chillies for training data and 36 samples for test data. The classification of the quality of dried red chilli was divided into three classes, which were extra class, class I, and class II. The result of this study was an artificial neural network structure consisting of 5 input layer cells, 16 hidden layer cells, and two output layer cells. The testing of the system using 36 testing data that determined the values of dried chillies reached 94.4%.

1. Introduction
Chillies (Capsicum annum L.) are one of Indonesia's leading commodities that can influence the country's inflation rate. Outside of harvest seasons, their price is very high but very cheap when in season. One of the alternatives to overcome this huge disparity is processing fresh chillies into dried chillies[1]. Dried chillies have many advantages, such as having high resistance to rot, having a good shelf life, and having high levels of capsaicin[2]. In addition, the capsaicin can prevent cancer[3].

Europe is the biggest importer of dried chillies in the world. Up to the present, India has become the largest dried chilli exporting country in the world. It opens up opportunities for Indonesia to compete in the European market. To compete in the international market, Indonesia must improve the quality standards of dried chillies by following European market commodity trading standards, namely the United Nations Economic Commission for Europe (UNECE). The quality classification of dried chillies according to the UNECE standards uses a combination between the color and length of chillies. Color classification for whole dried chillies that is mandatory for the commercial types is divided into three classes. Extra class chillies have dark red color (red content >98%) and are >9cm in length. Class I chillies have a bright red color (red content >95%) and are >9cm in length, and class II chillies have a yellowish red or orange color (red content >90%) and are either >9cm or <9 cm in length. Chillies that cannot be classified into the third grade are disqualified. In addition, chillies must meet the absolute requirements of dry chillies that they must have a maximum moisture content of 9.0%.

To obtain dried chillies with the high quality and meet the standards, the sorting must be accurate and precise. The sorting process of dried chillies in Indonesia is still done manually. As a result, the
bias caused by human errors is high. It happens due to inconsistencies in the classifying and sorting. A technology designed to classify the quality of dried chillies accurately and precisely is needed to minimize the errors during the sorting. One example is the use of the image processing and artificial neural network (ANN) approach.

Based on previous research, there have been several studies related to the quality classification based on image processing using ANN on chilli plants[4]. There are already quality classifications of dried chillies, with which they are grouped based on their color features of various varieties[5]. However, in the previous studies no one has examined and combined individual quality classification (per piece) by combining color values and length values. This is interesting as according to the existing standard, the length is one of the determinants of chilli quality.

2. Methods

2.1. Samples preparation
Fresh chilli (Capsicum annum L.) samples, OR 42 variety, were obtained from farmers in Purwobinangun Village, Sleman Regency, Indonesia. This region is the largest producer of fresh red chillies in the Special Province of Yogyakarta. Sampling was done from 08.00 to 10.00 Western Indonesian Time (WIT) during a harvest time. The total number of samples was 186 chillies, and these chillies had different levels of colors that were dark red, bright red, tinge (orange), green, and black.

The fresh chilli samples were brought to the laboratory and dried by following a procedure. Firstly, the chillies were blanched at 80° to 90° C for 4 to 6 minutes. The purpose of the blanching was aimed to deactivate the oxidation enzymes and to increase B-carotene[6]. Furthermore, the blanching also increased the drying rate and maintained the chillies’ quality during drying. The drying was conducted for 16 hours at 60° C to reach a maximum water content standard of 9.0%.

2.2. Image acquisition
The primary data required were extraction data of color and texture features, actual lengths, and estimated lengths (pixels). The photo was captured using Logitech C270 camera manually and connected to the computer. The image extracted color and texture features using the MATLAB application. The scheme of image capturing is shown in Figure 1.

![Figure 1. Image capturing scheme](image_url)

The image capturing tool used in this study was Logitech C270 HD Resolution (720p) camera or computer vision with a resolution of 1024x768 pixels, and each image was taken from the image capture box. The box was made from dark acrylic (black color), and it had 4 LED lamps above and a black sample background, as shown in Figure 2.
2.3. Image processing
The image extraction step consisted of converting an RGB image to an HSV image, converting an
HSV image to a black and white (BW) image, thresholding and segmentation, removing chillies
stalks in green, and returning the image to an RGB image. The image extraction was done by means
of a GUI to make the process easier as shown in Figure 3.

2.3.1. Color feature extraction. The dried chilli images generally consisted of RGB components such
as red, green, and blue. RGB color space was converted into another color space such as HSV color
using “rgb2hsv” code and Lab color using rgb2lab in MATLAB application.

2.3.2. Texture feature extraction. The texture of the image measured the roughness, coarseness,
and smoothness by calculating the outer parts of the images [7]. Gray level co-occurrence matrix was
used to calculate different texture features, namely GLCM. The GLCM matrix gave several derivative
values such as contrast, energy, entropy, correlation, and homogeneity value.

2.4. Classification using ANNs
The development of artificial neural network structures (ANNs) consisted of two steps, which were
the training step to determine the ANNs structure and the validating step to measure the level of ANNs
accuracy of the quality classification system. The classification system was compiled and developed
by means the toolbox and MATLAB application. The data used for the training and classification system validation were called a dataset.

The dataset used to compile this quality classification system were the dried chillies. The ratio of training data and testing data was 80:20, and this stood for 150 training samples and 36 testing data samples. The dataset consisted of three quality classes and defect category (the black and green dried chillies). The standard used for the quality classification was the UNECE Standard.

3. Result and Discussion

3.1. Determination of ANN parameters

Extracted sample data were stored in a spreadsheet file (.xlsx) created with Microsoft Excel software, but they were still scattered. Subsequently, the data were filtered by using control data to remove samples that were out of bounds (upper and lower limits). The 150 training samples were classified into three quality classes (i.e. Extra, Class I, and Class II) and disqualified class according to the UNECE standard as shown in Table 1.

| No | Class         | Dried chillies color                                      | Data Training |
|----|---------------|-----------------------------------------------------------|---------------|
| 1  | Extra class   | Dark red (red > 98%) and Length > 9 cm                    | 31            |
| 2  | Class I       | Bright red (red > 95%) and length > 9 cm                  | 20            |
| 3  | Class II      | Red—orange (red > 90%) with all length                    | 26            |
|    |               | Extra class and Class I with length < 9 cm               |               |
| 4  | Disqualified  | Black and green chillies                                 | 38            |

Included in the disqualification class were black and green dried chillies; the details of the number of samples belonging to the fourth class as shown in the Table 2.

| No | Class       | Dried chillies color | Data Training |
|----|-------------|----------------------|---------------|
| 1  | Disqualification | Black                | 20            |
| 2  |              | Green                | 18            |

Subsequently, the correlation test was done to determine the relationship between the color of the dried chillies and the quality classes as shown in Table 2. This correlation analysis was done with SPSS software, and the results are shown in Table 3. The extraction data parameters consisted of Red, Green, Blue, Mean L*, Mean a*, Mean b*, Hue, Saturation, Value, Entropy, Contrast, Homogeneity, and Energy.

The parameters chosen as inputs for the artificial neural network (ANN) were those that had correlation values of ≥0.4. Correlation values between 0.4 and 0.6 indicate moderate correlation values while correlation values of ≥ 0.6 indicate strong correlations, so that they were able to significantly influence the results[8]. Based on the correlation test, the selected parameters for further analysis were length of dried chillies, Mean Energy, Mean a*, Mean b*, Hue, Saturation, Value, Entropy, Contrast, Homogeneity, and Energy.

The parameters represented the levels of characteristics for classifying the color quality class of dried chillies. Mean a* and blue value represented the color features. Mean a* was categorized as a chroma color that had a range of colors between green and red. For the value of mean a*, the higher the value of mean a*, the more red the color was, and the lower the color, the closer the color to green.

For the blue value, the higher the blue value, the greener the color was, and the lower the color, the closer the color to yellow. The mean energy and mean contrast represented the texture features. The
texture features not only indicated the difference between pixels but also provided correlation values whether they were wither homogeneous or heterogeneous[9].

Table 3 The result of the correlation analysis on color textures

| Parameter               | Correlation Value (p-value) | Significance Value |
|-------------------------|-----------------------------|--------------------|
| Length of chillies      | 0.595                       | 0.000              |
| Mean Energy             | 0.593                       | 0.000              |
| Mean a*                 | 0.453                       | 0.000              |
| Mean Blue               | 0.437                       | 0.000              |
| Mean Contrast           | 0.416                       | 0.000              |
| Mean Correlation        | 0.392                       | 0.000              |
| Mean Green              | 0.370                       | 0.000              |
| Mean b*                 | 0.341                       | 0.000              |
| Mean Entropy            | 0.225                       | 0.006              |
| Mean Red                | 0.220                       | 0.006              |
| Mean Value (V)          | 0.154                       | 0.600              |
| Mean Saturation (S)     | 0.136                       | 0.095              |
| Mean Homogeneity        | 0.080                       | 0.326              |
| Mean L*                 | 0.070                       | 0.396              |

3.2. Structure preparation and training of ANNs

The structure development was carried out by using the neural network toolbox through MATLAB software. The structure consisted of three layers that were the input layer, hidden layer, and output layer. The structure development was aimed to create a good system for the classification of the dried chillies’ quality.

The initial stage of the development began with determining the training algorithm that was used, number of hidden layers, value of learning rate, value of constant momentum, activation function, and amount of iteration. The determination of the structure development was done by using the trial and error method to obtain a model with high performance. The design performance was able to measure from a MSE value. A good model is a model that has the smallest possible MSE value[10]. It will better if the model has a MSE value that is very close to the value of 0.001[9]. The output layer consisted of 2 cells that consisted of numbers 0 and 1 representing four quality classes according to the UNECE standard. After the compilation and training process were conducted, the smallest MSE value was selected with the following details:

1) Performance target value (goal): 0.001
2) Maximum iteration value: 3000
3) Learning rate value (LR): 0.35
4) Momentum constant value (MC): 0.55
5) Amount of input and activation: 5 cells (tansig)
6) Amount of hidden layer and activation: 16 (logsig)
7) Amount of output and activation: 2 cells (logsig)

3.3. Validation of ANNs

The process of network validation was aimed to predict or classify a sample using weight and bias values, and it was then stored in a file and called when classification was conducted. Firstly, the
classification system was validated by using 150 samples of training data. This classification combined the features of colors and the estimated lengths of dried chillies. The estimation of dried chillies' lengths was done by using the formula \( y(x) = 0.014x + 0.7716 \) cm where \( y \) was the estimated length in centimeter or cm, and \( x \) was the maximum number of x-axis in pixels. The length estimation equation of dried chillies formula followed a linear regression approach and used dried chillies training data. The equation formula can be seen in Figure 4.

![Figure 4. Equation formula for length estimation](image)

The stages of validation using training data and measuring accuracy with testing data obtained the following results with a combination of color features and length quality components as shown in Table 4 and Figure 4. From several combinations of the trial and error experiments, the highest and most constant level of accuracy during validation and testing was MSE 0.00333.

| Number | MSE     | The level of accuracy of validation using data training | The level of accuracy of testing using data testing |
|--------|---------|-------------------------------------------------------|---------------------------------------------------|
| 1      | 0.00667 | 98.6%                                                 | 83.3%                                             |
| 2      | 0.00668 | 98.6%                                                 | 86.1%                                             |
| 3      | 0.00668 | 98.6%                                                 | 88.9%                                             |
| 4      | 0.01000 | 98.6%                                                 | 88.9%                                             |
| 5      | 0.01100 | 98.6%                                                 | 91.7%                                             |
| 6      | 0.00333 | 99.3%                                                 | 94.4%                                             |
| 7      | 0.00334 | 99.3%                                                 | 83.3%                                             |
| 8      | 0.00380 | 99.3%                                                 | 77.7%                                             |
| 9      | 0.00677 | 99.3%                                                 | 80.5%                                             |
| 10     | 0.01000 | 99.3%                                                 | 86.1%                                             |

Table 5 provides information about the results of the testing using testing data of 36 samples, and it achieved an accuracy of 94.4%. The quality classification system for dried chillies was said to be good if the system can recognize previously unknown samples. Therefore, the qualities of the dried chillies were classified by using the chosen ANN chosen, which was when the MSE level was 0.00333.
was because at this level 99.3% validation accuracy was achieved when validated by using the training data set (150 samples), and 94.4% testing accuracy was achieved when tested with the testing data set (36 samples). The details of testing result are shown in Table 5.

### Table 5 The results of testing

| Class       | Actual data testing | Correct prediction | False prediction | Level of accuracy |
|-------------|---------------------|--------------------|------------------|-------------------|
| Extra       | 8                   | 8                  | 0                | 100%              |
| Class I     | 5                   | 5                  | 0                | 100%              |
| Class II    | 13                  | 11                 | 2                | 84.6%             |
| Disqualification | 10               | 10                 | 0                | 100%              |

**Average accuracy** 94.4%

The false prediction for class II in the above table was caused by the mean a* value of 2 samples of the class II that was close to class I and was near to the middle coordinate of O (0, 0) or a value close to 0 (zero), so that the chillies’ colors were darker. Therefore, class II dry chillies had a mean value of a*, which was almost the same as class I [11]. In addition, it was also influenced by the small mean value of blue and within the mean blue range for class I, so that the colors of the chillies were darker, the chillies were predicted to belong to class I. The UNECE classification standard was used to determine the quality classification of dried chillies as presented in Table 6 below.

### Table 6 The UNECE standard for the classification of dried chillies’ quality

| Number | Class       | Color                  | Length (cm) | Moisture Content |
|--------|-------------|------------------------|-------------|-----------------|
| 1.     | Extra class | Dark red (red > 98%)   | > 9         |                 |
| 2.     | Class I     | Bright red (red > 95%) | > 9         |                 |
| 3.     | Class II    | Extra & Class I        | < 9         | Max. 9.0%       |
|        |             | Red—orange (red > 90%) | All length  |                 |
| 4.     | Disqualification | Black   | All length   |                 |
|        |             | Green                  | All         |                 |

This high accuracy value was resulted from the linear regression equation that had the best prediction for dried chillies’ lengths. The results of the linear equation validation for the length prediction had R² of 0.9901, RMSE of 0.207, and a MAPE value of 1.91% as presented in Figure. 5. The R² value was close to 1.0, so it indicated that the model was very accurate and precise for the prediction because it had a small RMSE value and a small MAPE that was less than 10% (MAPE 3.60%)[12]. The dried chillies quality classification system could be relied upon and accepted because it had a high accuracy value and exceeded the accuracy of the quality classification system achieved in previous research (only 90%)[5].
Previous studies made mango classification using color, size, and skin conditions with an accuracy of up to 80%, strawberry classification using color and size features obtaining a 88% accuracy, and lemon classification using color and size feature reaching a 94% accuracy [13]. Based on this, overall, the classification system of dried chillies’ quality by means of the ANN method based on the UNECE standard can be used and trusted.

4. Conclusions
Input parameters for the classification of dried chillies’ quality were lengths of the chillies, Mean Energy, Mean a*, Mean Blue, and Mean Contrast. They had a correlation value of 0.595, 0.593, 0.453, 0.437, and 0.416. The input parameters were used to compile the ANN architecture. The ANN architecture consisted of 3 layers. The number of input layer cells was 5 cells; the hidden layer cells was 16 cells, and the output layer cells was 2 cells. The output layer used was extra class (0 and 0), Class I (0 and 1), class II (1 and 0), and disqualified class (1 and 1). Validation and testing results provided an accuracy rate of 99.3% and 94.4%, respectively.

References
[1] Alsebaeai M, Chauhan A, Arvind, Yadav P, Al-sayadi M, Al-Salali S, an Al Eryani H 2018 Microscopic Characterizations of Green Chilli Powder Int. J. Curr. Microbiol. Appl. Sci. 7 2488–2497
[2] Muamaroh S and Wahyono 2018 A TINGKAT KETAHANAN BEBERAPA VARIELAS CABAI MERAH (Capsicum annuum L.) HIBRIDA PADA KEMASAKAN BUAH TERHADAP PENYAKIT ANTRAKNOSA Colletichum acutatum RESISTANCE LEVEL OF HOT PEPPER VARIETIES (Capsicum annuum L.) HYBRIDS IN FRUIT MATURITY OF ANTRACHNOSA Jurnal Produksi Tanaman 66 19–628
[3] Montoya-Ballesteros L C, González-León A, García-Alvarado M A, and Rodríguez-Jimenes G C 2014 Bioactive Compounds During Drying of Chilli Peppers Dry. Technol 32 1486–1499
[4] Patil A and Lad K 2019 Review of Diseases Detection and Classification for Chilli Leaf using Various Algorithms Proc. 2019 3rd IEEE Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2019 (Tamil Nadu : IEEE) 1–4, 2019.
[5] Sajjan M M, Kulkarni L and Sajjan M 2019 Combined Color , Edge and Texture Features for Chilly Recognition Combined Color, Edge and Texture Features for Chilly Recognition International Conference on Sustainable Advanced Computing and Statistical Techniques ICSACST-19 (Bangalore).

[6] Anoraga S B, Sabarisman I and Ainuri 2018 M Effect of different pretreatments on dried chilli (Capsicum annum L.) quality IOP Conference Series: Earth and Environmental Science 131 0–5.

[7] Bairwa D and Sharma G 2017 Classification of Fruits Based on Shape, Color and Texture using Image Processing Techniques Int. J. Eng. Res. V6 110–114

[8] Schober P and Schwarte L A 2018 Correlation coefficients: Appropriate use and interpretation Anesth. Analg. 126 1763–1768

[9] Khuriyati N, Pamungkas A P and Pambudi AA 2019 The Sorting and Grading of Red Chilli Peppers (Capsicum annuum L.) Using Digital Image Processing Int. J. Agric. Environ. Sci. 6 17–23

[10] Rougier J 2016 Ensemble averaging and mean squared error J. Clim. 29 8865–8870

[11] Bora D J, Gupta A K and Khan F A 2015 Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation International Journal of Emerging Technology and Advanced Engineering 5 192-203

[12] Khair U, Fahmi H, Al Hakim S and Rahim R 2017 Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error J. Phys. Conf. Ser. 930

[13] A. Bhargava and A. Bansal 2018 Fruits and vegetables quality evaluation using computer vision: A review J. King Saud Univ. - Comput. Inf. Sci.

Acknowledgements
This research was financially supported by Universitas Gadjah Mada via Students’ Final Project Recognition Program Year 2020.