Application of colour, shape, and texture parameters for classifying the defect of Gayo Arabica green coffee bean using computer vision

A Maghfirah¹, I S Nasution*²

¹Master of Agricultural Industry Technology, Universitas Syiah Kuala, Jalan Tgk. Hasan Krueng Kalee 3, Darussalam-Banda Aceh 23111, Indonesia
²Agricultural Engineering Department, Universitas Syiah Kuala, Jalan Tgk. Hasan Krueng Kalee 3, Darussalam-Banda Aceh 23111, Indonesia

*Email: i.nasution@unsyiah.ac.id

Abstract. Coffee is the most important commodity in the trading industry. Determination of the quality of coffee is still done manually so that it cannot separate good quality coffee beans with bad quality coffee beans. This research conducted the development of a visual-based intelligent system using computer vision to be able to classify the quality of rice coffee based on the Indonesian National Standard (SNI). The models used in the study are the K-Nearest Neighbour (K-NN) method and the Support Vector Machine (SVM) method with 13 parameters used such as: area, contrast, energy, correlation, homogeneity, circularity, perimeter, and colour index R(red), G(green), B(blue), L*, a* and b*. A total of 1200 Arabica green coffee bean captured using Kinect V2 camera with training data of 1000 samples and testing data of 200 samples.

1. Introduction
Coffee is one of the most important commodities in the trading industry. Based on statistics, Indonesia is the fourth producers and exporters in the world [1]. Coffee is not only used for consumption but also has become a lifestyle. The system for handling and improving the quality of coffee must be in accordance with the quality of the coffee beans that have been determined. Coffee quality determination in Indonesia is generally still done manually, which has drawbacks such as a low level of quality consistency and requires a relatively time-consuming. These constraints make it difficult to separate quality coffee beans for large production.

In order to overcome these problems, it is necessary to develop a coffee processing process, one of which is a sorting system. Separation of green coffee beans can be accomplished using several characteristics such as shape, colour, defect, and size [2–5]. Several studies have been developed for coffee sorting using computer vision, assessing the quality of Arabica and Robusta green coffee using the L*a*b* colour feature with several criteria such as white, sugarcane green, green, bluish green [3]. Classification of green coffee using digital image processing and neural network based on texture feature, obtained high accuracy values [4]. However, research on computer vision on Gayo Arabica bean that uses combinations features such as colour, shape and texture need to be more attention in
order to increase more accuracy. The study aims to classify the defect on Gayo Arabica green coffee based on colour, shape, and texture by using computer vision.

2. Materials and methods

2.1. Green coffee sample

A total of 700 Gayo Arabica green coffee beans was obtained from Bener Meriah regency with a semi-wash process which has been separated into 5 criteria such as: normal, sour bean, partly black beans, broken beans, and black beans (figure 1).

![Figure 1. Images of Gayo Arabica green coffee obtained based on five criteria.](image)

2.2. Image acquisition

The images were taken using a Kinect v2 camera with a resolution of 1920 x 1080 pixels. The distance between the object and the image is 30 cm. The obtained image recorded on a laptop for further processing. The image acquisition process is equipped with lighting from LED lights.

2.3. Histogram smoothing

Green coffee bean and background was processed using Otsu method and histogram smoothing. Firstly object was extracted based on L*a*b* colour space, and b* channel was chosen for separating between the objects and the background using Otsu method [8,9]. Secondly, the grey value of relative histogram was determined, and the significant minima were isolated from the histogram, which then used as boundaries for thresholding. The histogram was smoothed using a Gaussian filter in order to reduce the number of minima [10].

2.4. Colour, shape, and texture

Colour parameters can be obtained from each colour pixel which is the intensity value for each colour such as red, green, and blue. Some texture features such as contrast, energy, correlation, and homogeneity are used to define the characteristics of the green coffee beans. Shape features such as perimeter, circularity are used to describe objects, while areas calculated based on the number of pixels in the object [11].

\[
\text{Area} = \sum_{m=1}^{X} \sum_{n=1}^{Y} bw_{\text{area}} (x, y) \quad (1)
\]

Perimeter is the number of pixels at the boundary of an object. If \(X_1 \ldots X_2\) is a list of limits, the perimeter equations based on equation (2).

\[
\text{Perimeter} = \sum_{i=1}^{N-1} d_i = \sum_{i=1}^{N-1} |X_i - X_{i+1}| \quad (2)
\]
The distance \( d \) is 1 for 4 connected borders, and 1 or \( \sqrt{2} \) to 8 for connected borders. The green coffee beans circularity calculates the similarity of the input region with a circle (equation 3).

\[
\text{Circularity} = \frac{4 \pi \cdot \text{Area}}{\text{Convex Perimeter}^2}
\]  

(3)

There are four texture parameters were used such as: homogeneity, energy, correlation, and contrast. Homogeneity measure how similar certain pixels of image, energy used to measure image homogeneity, correlation measure grey value dependencies, and contrast measure the size of the intensity differences. The texture of green coffee beans calculated based on the equation 4-7. Where width is width of co-occurrence matrix and \( C_{ij} \) is entry of co-occurrence matrix.

\[
\text{Contrast} = \sum_{i,j=0}^{\text{width}} (i - j)^2 C_{ij}
\]  

(4)

\[
\text{Correlation} = \frac{\sum_{i,j=0}^{\text{width}} (i - u_x)(j - u_y) C_{ij}}{S_x S_y}
\]  

(5)

\[
\text{Energy} = \sum_{i,j=0}^{\text{width}} C_{ij}^2
\]  

(6)

\[
\text{Homogeneity} = \sum_{i,j=0}^{\text{width}} \frac{1}{1 + (i - j)^2} C_{ij}
\]  

(7)

2.5. Classification

2.5.1. Method K-Nearest Neighbour (K-NN). Classification using the K-Nearest Neighbour (K-NN) method is calculated based on learning data obtained by the value of \( k \) from the nearest object. Previous studies using K-Nearest Neighbours (K-NN) based on colour and texture [12], and also [13] classify based on texture and shape. Figure 2 shows the illustration of the used of \( k \) value using K-NN.

![Figure 2. Illustration of K-NN classification.](image)

Euclidean distance is the calculation of the distance between test data and training data based on the equation 8 [14].
\[ d_1 = \sqrt{\sum_{i=1}^{p} (X_{2i} - X_{1i})^2} \]  

(8)

2.5.2 Support Vector Machine (SVM) method. Support Vector Machine (SVM) is a method that separate classes by determining the separator field based on the training data used for unknown class data [15]. The best dividing field is obtained by finding the maximum point from the data closest to each class [16], which can be seen in Figure 3. In the SVM method, there are three main kernel functions that will be used in the test such as linear, radial basis function, and polynomial. Related to coffee product there is few study using SVM for coffee classification, Christian et al. [16] have been conducted research using the SVM method based on colour parameters, the accuracy is 90.8%.

![SVM classification illustration](image)

**Figure 3.** SVM classification illustration.

3. Results and discussion
The colour features including red, green and blue are shown in Figure 4, where the red colour pixel value is higher than all green coffee defectives criteria. Since the value of red colour is higher than other colour indexes (green and blue), red colour is preferable to be used as a predictive parameter. Figure 5 shows the shape parameters consisting of area, perimeter, and circularity. The area feature has a significant effect in distinguishing the five defectives of Arabica green coffee defects. Meanwhile, circularity tends to be low value which is not recommended for predicting the Arabica green coffee. Figure 6 shows texture parameters which include contrast, correlation, energy and homogeneity. The contrast value has a good average value to distinguish the five criteria; however the energy has an insignificant value for distinguishing the five criteria [17].
Figure 4. Colour features of Arabica green coffee beans based on defective criteria.

Figure 5. Shape features of Arabica green coffee beans based on defective criteria.
3.1. Classification using the K-Nearest Neighbour (K-NN) method

The K-Nearest Neighbour (K-NN) method is influenced by the value of k to determine the most optimal class in the classification process. All features used in this method are calculated to produce significant accuracy values. In this study, the best k value was obtained k = 7 with an accuracy rate of 91.2%. Previous studied by Yhurinda et al [18] identified the disease of arabica coffee plants using the KNN using a k value of 5 with an accuracy of 80%. The magnitude of the effect of the value of k is equivalent to the specified criteria.

Table 1. Accuracy of training and testing data of Arabica green coffee using K-NN method.

| Sample | Number of samples | Normal | Sour Beans | Partly Black Beans | Broken Beans | Black Beans | Average |
|--------|-------------------|--------|------------|-------------------|-------------|-------------|---------|
| Training | 500               | 100    | 90         | 97                | 95          | 95          | 95.4    |
| Testing  | 200               | 93     | 94         | 97                | 80          | 92          | 91.2    |

According to Table 1, 95.4% of the training data were classified correctly, while 91.2% of the testing data were classified correctly. This result has been compared by Ikhsan et al [19] which uses six features with the classification of Arabica green coffee beans is 63.5%.

3.2. Classification using the Support Vector Machine (SVM) Method

The support vector machine (SVM) measure the size of the divisor plane and finding its maximum point for separating data from its class. The accuracy obtained is 91.4% for training data and 90.8% for testing data, which the radial basis function kernel was chosen as the best kernel with a C value of 10, a gamma value of 0.05, an epsilon value of 0.0001 and a regularization value of 0.005, respectively.

Figure 6. Texture features of Arabica green coffee beans based on defective criteria.
Table 2. Accuracy of training and testing data of Arabica green coffee using SVM method.

| Sample  | Number of Samples | Normal | Sour Beans | Partly Black Beans | Broken Beans | Black Beans | Average  |
|---------|-------------------|--------|------------|--------------------|--------------|-------------|----------|
| Training | 500               | 94     | 93         | 89                 | 90           | 91          | 91,4     |
| Testing  | 200               | 93     | 90         | 92                 | 89           | 90          | 90,8     |

According to Table 2, testing data of 200 samples from 5 criteria for green coffee beans obtained very good classification results, especially on the criteria for normal coffee beans. Overall, the classification using the SVM method for all the respective criteria produces high accuracy.

4. Conclusion

Based on this study, the two methods (SVM and KNN) were able to classify the defects of Gayo Arabica green coffee very well. The K-NN method gives higher accuracy results than the SVM method. The adoption of the K-NN method is expected to be very useful for further research, particularly for real time classification.

References

[1] Kemenperin 2019 Kemenperin: Kopi Kita Untuk Dunia 2019
[2] Arboleda E R 2019 Comparing Performances Of Data Mining Algorithms For Classification Of Green Coffee Beans Int. J. Eng. Adv. Technol. 8 1563–7
[3] Bicho N C, Leitão A Nio E, Ramalho J C And Lidon F C 2014 Application Of Colour Parameters For Assessing The Quality Of Arabica And Robusta Green Coffee Emirates J. Food Agric. 26 9–17
[4] Ramalakshmi K, Kubra I R And Rao L J M 2007 Physicochemical Characteristics Of Green Coffee: Comparison Of Graded And Defective Beans J. Food Sci. 72
[5] Srisang N, Chanpaka W And Chungcharoen T 2019 The Performance Of Size Grading Machine Of Robusta Green Coffee Bean Using Oscillating Sieve With Swing Along Width Direction Jop Conf. Ser. Earth Environ. Sci. 301
[6] Silvia E, Sidebang B, Chanona-P J, Juan V M, Perea-Flores J, Arzate-V I, Polit I, Rico-Fernández M P, Rios-Cabrera R, Castelán M, Guerrero-Reyes H, Kang H, Chen C, Paper T, Cheng H, Damerow L, Sun Y, Blanke M, Ali H, Lali M I, Nawaz M Z, Sharif M, Saleem B A, Behera S K, Rath A K, Mahapatra A, Sethy P K, Arakeri M P, Choi D, Lee W S, Ehsani R, Schueller J K, Roka F, Chen Y, Chao K, Kim M S, Choi D, Lee W S, Schueller J K, Ehsani R, Roka F, Diamond J, Bargoti S, Oliveira E M De, Leme D S, Henrique B, Barbosa G, Rodarte M P, Gualberto R, Alvarenga F, Dorj U, Lee M, Yun S, Fawcett T, Mochida K, Koda S, Inoue K, Habaragamuwa H, Ogawa Y, Suzuki T, Shiigi T, Ono M, Hamuda E, Glavin M, Jones E, Kamilaris A, Prenafeta-Boldú F X, Keightley K E, Bawden G W, Leme D S, Alves S, Henrique B, Barbosa G, Meira F, Gualberto R, Alvarenga F, Mavridou E, Vrochidou E, Papakostas G A, Pachidis T, Vassilis G, Momin M A, Rahman M T, Sultana M S, Igathinathane C, Ziauddin A T M, Grift T E, Nasution I S, Gope H L, Fukai H, Wang Z, Walsh K B, Arboleda E R, Jos C, Ang C L, Chalfoun I S, Mass O, Estimation S, Et Al 2018 A Computer Vision System For Coffee Beans Classification Based On Computational Intelligence Techniques Comput. Electron. Agric. 156 103–12
[7] Nasution T H And Andayani U 2017 Recognition Of Roasted Coffee Bean Levels Using Image Processing And Neural Network Jop Conference Series: Materials Science And Engineering Vol 180 (Institute Of Physics Publishing) P 012059
[8] Nasution I S, Satriyo P, Yolanda S And Alma A 2021 Non-Destructive Measurement Of Leaf Area And Leaf Number Of Hydroponic Pak-Choy Plants (Brassica Rapa) Iop Conf. Ser. Earth Environ. Sci.

[9] Otsu N 1979 A Threshold Selection Method From Gray-Level Histograms Ieee Trans. Syst. Man. Cybern. 9 62–6

[10] Mvtec 2020 Solution Guide I Basics Halcon 20.05 Progress

[11] Nasution I S And Gusriyan K 2019 Nutmeg Grading System Using Computer Vision Techniques Iop Conf. Ser. Earth Environ. Sci. 365 012003

[12] Rizal M A Klasifikasi Mutu Biji Kopi Menggunakan Metode K – Nearest Neighbor Berdasarkan Warna Dan Tekstur

[13] Pawening R E And Ja W 2020 Klasifikasi Kualitas Jeruk Lokal Berdasarkan Tekstur Dan Bentuk Menggunakan Metode K-Nearest 8 10–7

[14] Pariyandani A 2019 Klasifikasi Citra Ikan Berformalin Menggunakan Metode K-Nn Dan Glcm Pros. Semin. Nas. Teknol. Inform. 2 42–7

[15] Windrawati A N 2020 Klasifikasi Varietas Kopi Arabika Menggunakan Metode Support Vector Machine ( Svm ) Skripsi Classification Of Arabic Coffee Variety Using Support Vector Machine ( Svm ) Method

[16] Christian B F, Lestari U, Andayati D, Informatika P S And Industri F T 2019 Support Vector Machine 7 248–56

[17] Arboleda E R 2018 Classification Of Coffee Bean Species Using Image Processing , Artificial Neural Network And K Nearest Neighbors 2018 Ieee Int. Conf. Innov. Res. Dev. 1–5

[18] Yhurinda A, Putri P And Sodik A 2019 Identifikasi Penyakit Tanaman Kopi Arabika Dengan Metode K-Nearest Neighbor (K-Nn) Semin. Nas. Sains Dan Teknol. Terap. Vii 759–64

[19] Ikhsan D, Utami E And Wibowo F W 2020 Metode Klasifikasi Mutu Greenbean Kopi Arabika Lanang Dan Biasa Menggunakan K-Nearest Neighbor Berdasarkan Bentuk J. Ilm. Sinus 18 1