Eggs Classification based on Egg Shell Image using K-Nearest Neighbors Classifier

Eko Hari Rachmawanto
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
eko.hari@dsn.dinus.ac.id

Christy Atika Sari
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
atika.sari@dsn.dinus.ac.id

Rivalda Villadelfiya
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
111201609580@mhs.dinus.ac.id

De Rosal Ignatius Moses Setiadi
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
moses@dsn.dinus.ac.id

Nova Rijati
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
nova.rijati@dsn.dinus.ac.id

Etika Kartikadarma
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
etika.kartikadarma@dsn.dinus.ac.id

Mohamed Doheir
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka
Melaka, Malaysia
Jeruslem.20088@gmail.com

Setia Astuti
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
setia.astuti@dsn.dinus.ac.id

***Abstract*— Chicken eggs are one of the foods that are widely consumed by humans. The quality of eggs will affect the nutritional quality of eggs. One method that can be used to determine the quality of the outer shell is the quality of the eggshell. This research proposes egg classification techniques based on eggshell images using the K-Nearest Neighbors (KNN) classifier based on two feature extractions, namely the extraction of Hue Saturation Value (HSV) color features, and the Gray Level Co-Occurrence Matrix (GLCM). The experiment was carried out using 100 egg images consisting of three classes, namely eggs of good quality, rotten, and defective. Of the 100 images used 21 images as testing images and the rest as training images. The test was conducted with parameter values k = 1.3, and 9 while the distance used for each k was 1.2, and 4. Based on the test results obtained the highest accuracy of 85.71%, where the parameter value k = 1; d = 2 and k = 1; d = 4.**

Keywords*— *egg classification, KNN, HSV, GLCM, chicken egg

I. INTRODUCTION

Chicken eggs are one of the animal foods that are consumed by many people and have high nutrition [1]–[3]. A simple way that can be used to determine the quality of a chicken egg is to look at the eggshell visually in the eggshell. According to Chukwuka et al in [3], five things affect eggshell damage, namely integrity, shape, texture, color, and cleanliness. Another thing that can be seen visually is the large number of areas containing black, green, or red dots on the egg that are aimed at microbial contamination. Poultry diseases can also affect the quality of the shell and can even make the shell become deformed such as having a rough, thin texture, not even having a shell[3], [4].

Classification of egg quality can be done with egg image processing technology. This process can be carried out in several stages, namely preprocessing, segmentation, feature extraction, classification, and measurement or evaluation[5]–[7]. Preprocessing and segmentation are stages that aim to get Region of Interest (ROI), usually done with the process of cropping, resizing, contrast adjustment, etc. While the segmentation aims to separate an object from other objects, feature extraction is a method of structural and statistical analysis to obtain patterns, textures, features, and colors of objects. Classification is the process of finding information from a data set that aims to predict the label or class of objects based on learning outcomes. Where each of these stages is very influential on the results of classification. Whereas evaluation or measurement is a process to find out the quality of the method.

Egg quality is visually visible from its texture and color so that the combination of texture feature extraction and its logical color can be used to produce a good classification. Gray Level Co-Occurrence Matrix (GLCM) is an extraction of texture features that uses gray degree distribution statistics by measuring the degree of contrast, granularity, and roughness of an object from the relationship between neighbors in pixels in the image. This method has been widely used in various studies on classifications based on digital images as in research [8]–[10], whereby extracting this feature can produce high classification accuracy. GLCM can certainly be used to detect rough, unclean egg texture, having areas that have black, red or green dot spots, even damage to the eggshell like cracks. Hue Saturation Value (HSV) color feature extraction is also combined so that the features used are more complete. HSV feature extraction has also been widely applied to classification research conducted by [1], [6], [9] which with this feature is also proven to improve classification accuracy.

In the classification process, several classifiers have been widely applied, such as Naïve Bayes (NB)[11], Support Vector Machine (SVM)[1], [5], and K-Nearest Neighbor (KNN)[9], [12], [13]. The KNN method has the advantage of a simple calculation process and a relatively high level of accuracy. So in this research the proposed KNN classification method and GLCM and HSV feature extraction to visually identify the egg quality of the eggshell based on its image.

II. LITERATURE REVIEW

A. Hue Saturation Value (HSV)

The HSV color space of three integral parts namely Hue, Value, and Saturation (chroma). Hue is used to distinguishing between colors so that the classification of each color can be identified. In general, the meaning is value. In color, value means the brightness of color because each color emits a different brightness. Some colors appear bright and colors that appear dark. The highest brightness is owned by white, and the lowest is black. While other colors are between these two colors. The whiter elements a color, the brighter it will appear. Vice versa, the blacker elements mixed in color, the darker it will appear. While color chroma is color intensity. What is meant by intensity is the weakness of an element of the color[9], [14]. The strength is measured by how close and far a color is to the original pigment. To obtain an HSV value, you must convert an RGB image to HSV with Eq 1 to Eq. 7.

$$r = \frac{R}{(R+G+B)} \tag{1}$$

$$g = \frac{G}{(R+G+B)} \tag{2}$$

$$b = \frac{B}{(R+G+B)} \tag{3}$$

$$V = max(r,g,b) \tag{4}$$

$$S = \begin{cases} 0, & V = 0 \\ 1, \quad -\dfrac{min(r,g,b)}{v}, & V > 0 \end{cases} \tag{5}$$

$$H = \begin{cases} 0, & if\ S = 0 \\ \dfrac{60 * (g-b)}{S * V}, & if\ V = r \\ 60 * \left[2 + \dfrac{b-r}{S * V}\right], & if\ V = g \\ 60 * \left[4 + \dfrac{r-g}{S * V}\right], & if\ V = b \end{cases} \tag{6}$$

$$H = \mathrm{H} + 360\ if\ \mathrm{H} < 0 \tag{7}$$

Where R = Red value has not been normalized; r = normalized red value; G = Green value has not been normalized; g = normalized Green value; B = Blue value has not been normalized; b = normalized Blue value; V = declare value of Value; S = declare the saturation value; H = declare the value of Hue

B. *Grey Level Co-Occurrence Matrix (GLCM)*

GLCM is defined as a tabulation of image pixel data which illustrates how often different combinations of gray values appear in the image. There are four directions commonly used to make the GLCM matrix, namely the angle of 0°, 45°, 90°, and 135°, see Fig. 1. In one direction, there is one GLCM matrix for each value chosen from distance and degrees[13], [15].

From the center of the matrix (x, y) seen from Fig. 1, four directions indicate the angle according to the value of distance d = 1. The following are the steps to extract the GLCM feature[16]:

1. Quantization: Represents the conversion value of the grayscale (0-255) image into a certain range of values.

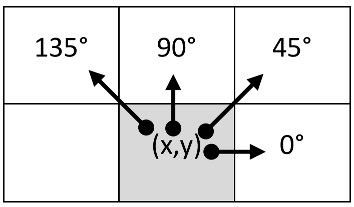

Fig. 2. Example of a figure caption. (*figure caption*)

The purpose of this quantization is to reduce the number of calculations and lighten the computational process. For example, eight value ranges (0 - 7) are specified where each range represents 32 gray values.

1. Co-occurrence: Number of occurrences of one level of neighboring pixel intensity value with one level of intensity of another pixel in a certain distance and orientation angle. Distances are expressed in pixels and angles are expressed in degrees. The orientation is formed in four angular directions with an angle interval of 45 °, from 0 ° to 135 °, while the distance between pixels is set at one pixel.
2. Symmetric: referred to as the appearance of the same pixel position. Symmetric is the sum of the cohesion matrix with its transpose matrix.
3. Normalization: done by dividing each number of the matrix in the symmetric matrix (D) by the sum of all the numbers in the matrix.
4. Feature Extraction: The extracted features are energy, contrast, correlation, and homogeneity. Energy value indicates the size of the image homogeneity. A high energy value occurs when the image texture tends to be uniform, the energy formula can be seen in Eq.8. Contrast is a calculation of the difference in intensity between one pixel and adjacent pixels throughout the image. Contrast is zero for a constant image, contrast calculation can be seen in Eq. 9. Homogeneity shows the homogeneity of an image with the same degree of gray. The homogeneous image will have a large homogeneity, homogeneity calculations can be seen in Eq. 10. While correlation measures the dissimilarity of an image where the value will be large if random and small if uniform, the correlation can be calculated with Eq.11.

$$energy = \sum_i \sum_j p(i,j)^2 \tag{8}$$

$$contras = \sum_i \sum_j (i-j)^2\, p_{(i,j)} \tag{9}$$

$$homogeneity = \sum_i \sum_j \frac{p(i,j)}{1 + |i-j|} \tag{10}$$

$$correlation = \sum_i \sum_j \frac{(i-\mu i)(j-\mu j)p_{(i,j)}}{\sigma i\ \sigma j} \tag{11}$$

Where p (i, j) = values of the matrix elements in a row (i) and column (j); μi, μj = the average value of elements in the matrix row and column; σi, σj = standard deviation values in the matrix row and column.

C. *K-Nearest Neighbor (K-NN)*

K-NN is a supervised learning method that classifies objects based on training data that is the closest distance to the object. In the test data is usually taken more than one of the closest neighbors to the training data then this algorithm is used to determine the class. Following is the algorithm of the K-Nearest Neighbor (KNN) classification :

1. Determine the value of the neighborhood (k).

2. Calculate the distance between new data to each labeled data.
3. Determine k the labeled data that has the minimum distance.
4. Classification of new data into labeled data in which the majority of KNN is selected based on distance metrics.

There are various ways in which KNN can be used to determine class, namely some distance rules that are used, one of which is Euclidean distance [8]. Euclidean distance is the usual distance between two points or coordinates derived from the Pythagoras formula. Euclidean distance is the hypotenuse of the line formed on the x-axis and the y-axis between the coordinates of point a and point b.

III. Proposed Method

The study proposes an egg quality classification method based on the visual appearance of the eggshell using the KNN classifier based on the extraction of HSV color features and GLCM features. Fig. 3 shows the steps taken in the proposed method, in detail can be explained as follows:

1. The egg image dataset is read and used as input to be copied or stored in a variable, for example, D1 and D2 for two different extraction processes. These D1 and D2 variables contain the entire image in the dataset, where D1 for HSV feature and D2 for the GLCM feature.
2. Convert the color space on D1, from the RGB color space to HSV.
3. Extract the color features on D1.

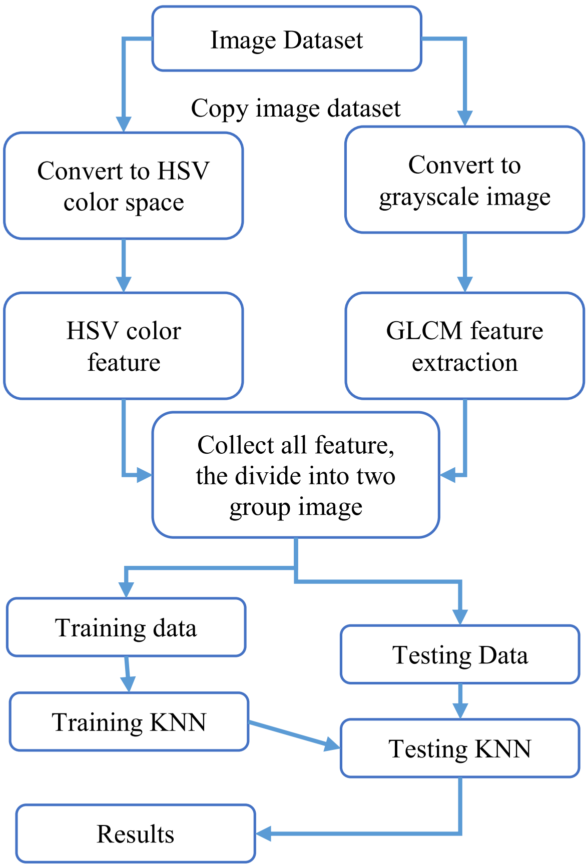

Fig. 3. Proposed Method

4. Perform color space conversion on D2, from RGB to Grayscale color space.
5. Extract GLCM features (Contrast, Correlation, Energy, and Homogeneity) on D2.
6. Collection of all feature extraction results, then divide into two groups, namely training data, and testing data, where the training has a class and the testing data does not have a class.
7. Perform KNN training, on training data.
8. Perform a classification test using KNN and Euclidean distance calculation in the testing data.
9. Get class or egg classification results.

IV. Results And Analysis

At this stage, 100 egg images are used as a dataset. The image dataset is divided into three classes, i.e. good, rotten and defective, see Fig.4.

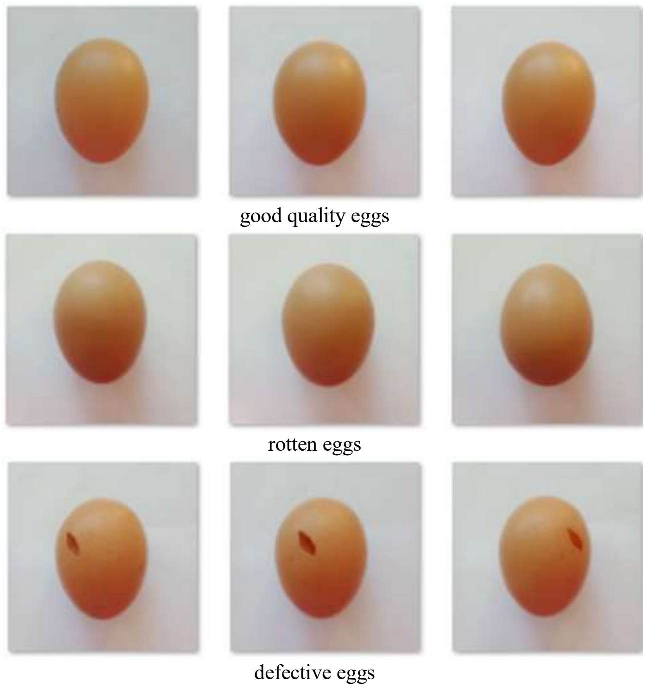

Fig. 4. Sample image dataset

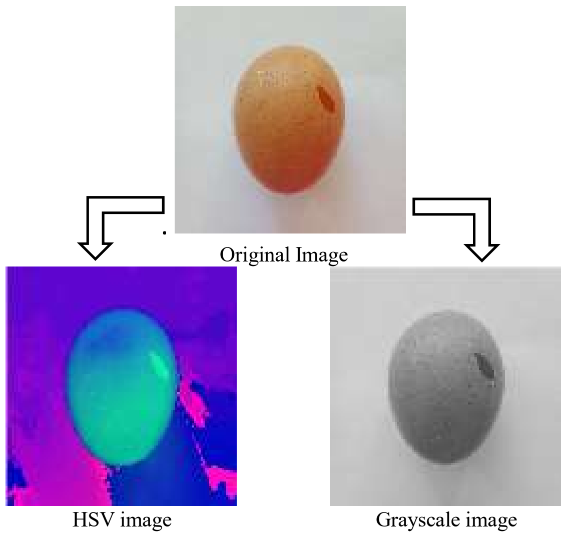

Fig. 5. Sample image after converting color space

In the class with good quality has several images of 34, and the quality of rot and defects of 33 images. The egg image was stabilized using an Oppo A83 camera with a 13MP camera quality with white paper as background carried out in a room with sufficient lighting, at 11.00 - 14.00 WIB GMT + 7. Sample dataset images are presented in Fig. 3. The color space conversion process is then performed on the egg image. An example is shown in Fig.5

After the image warrants conversion process is carried out, the HSV color feature extraction on the HSV image and the GLCM feature extraction on the grayscale image. Table 1, shows an example of the value of the feature extraction in one of the egg images, with the value k = 1 and pixel distance (d) = 1.

At the classification stage using KNN, the values k = 1, 5, and 9 are used where at a distance of 1, 2, and 4 pixels, respectively. The testing data used were 21 pieces in which each class consisted of seven images. The classification results are taken from the average value of GLCM feature extraction from all angles, and the results are presented using the confusion matrix in Tables 2 to 10.

TABLE I. SAMPLE FEATURES EXTRACTION FROM ONE IMAGE

Features	**Degrees**			
	0°	***45°***	***90°***	***135°***
Contrast	0.042569	0.054064	0.032972	0.054932
Correlation	0.98586	0.98215	0.98907	0.98187
Energy	0.38962	0.38678	0.39591	0.38688
Homogeneity	0.9806	0.97669	0.98384	0.97601
Hue	0.5732			
Saturation	0.57271			
Value	0.57031			

TABLE II. KNN CLASSIFICATION RESULTS (K=1 AND D=1)

		Prediction class		
		good	***rotten***	***defective***
Actual class	***good***	6	2	0
	rotten	1	5	1
	defective	0	0	6

TABLE III. KNN CLASSIFICATION RESULTS (K=1 AND D=2)

		Prediction class		
		good	***rotten***	***defective***
Actual class	***good***	6	2	0
	rotten	1	5	0
	defective	0	0	7

TABLE IV. KNN CLASSIFICATION RESULTS (K=1 AND D=4)

		Prediction class		
		good	***rotten***	***defective***
Actual class	***good***	6	2	0
	rotten	1	5	0
	defective	0	0	7

TABLE V. KNN CLASSIFICATION RESULTS (K=5 AND D=1)

		Prediction class		
		good	***rotten***	***defective***
Actual class	***good***	5	1	0
	rotten	2	6	0
	defective	0	0	7

TABLE VI. KNN CLASSIFICATION RESULTS (K=5 AND D=2)

		Prediction class		
		good	***rotten***	***defective***
Actual class	***good***	6	2	1
	rotten	1	5	0
	defective	0	0	6

TABLE VII. KNN CLASSIFICATION RESULTS (K=5 AND D=4)

		Prediction class		
		good	***rotten***	***defective***
Actual class	***good***	5	2	0
	rotten	2	4	0
	defective	0	1	7

TABLE VIII. KNN CLASSIFICATION RESULTS (K=9 AND D=1)

		Prediction class		
		good	***rotten***	***defective***
Actual class	***good***	4	2	0
	rotten	3	5	0
	defective	0	0	7

TABLE IX. KNN CLASSIFICATION RESULTS (K=9 AND D=2)

		Prediction class		
		good	***rotten***	***defective***
Actual class	***good***	5	2	0
	rotten	2	4	0
	defective	0	1	7

TABLE X. KNN CLASSIFICATION RESULTS (K=9 AND D=4)

		Prediction class		
		good	***rotten***	***defective***
Actual class	***good***	5	2	0
	rotten	1	5	0
	defective	1	0	7

Based on the matrix confusion presented in Tables 2 through 10, it appears that the highest accuracy is at k = 1; d = 2 and k = 1; d = 4, with the number of correct classifications of 18 images from a total of 21 testing images. Of all tests, the highest accuracy is in the classification of defective classes, where almost all tests can classify these classes perfectly. This is because eggs with deformed shells have a different texture where the value of features certainly has quite a dominant difference compared to the other two classes. In the class of eggs with good quality and visually rotten eggs do not have a

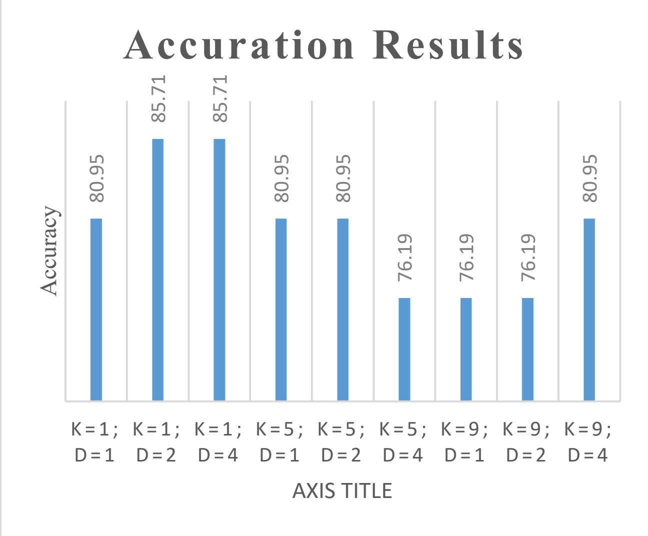

Fig. 6. Graph of the accuracy of all experiments

striking difference, it's just that if observed closely the characteristics of eggs with poor quality, relatively more spots, and more prone to rupture. From Table 2 to Table 10 the accuracy results can be summarized and presented using the graph in Fig. 6.

V. Conclusions

This research proposes the classification technique of chicken egg quality using the KNN method as a classifier and HSV and GLCM as extraction features. Classification is done based on the image of the eggshell because the eggshell can also be used as a parameter that determines egg quality. Based on testing with values k = 1, 5, and 9, where each value of d (distance) in each k is 1, 2, and 4, the highest value is 85.71. These results indicate that this method is quite effective for classifying eggs based on the image of the eggshell. In future research, the accuracy value still needs to be optimized again, for example by adding some other feature extraction and segmentation processes, as well as tests with several other classifiers to get better accuracy.

References

[1] R. Waranusast, P. Intayod, and D. Makhod, "Egg size classification on Android mobile devices using image processing and machine learning," in *Proceedings of the 2016 5th ICT International Student Project Conference, ICT-ISPC 2016*, 2016, pp. 170–173.

[2] A. F. A. Nasir, S. S. Sabarudin, A. P. P. A. Majeed, and A. S. A. Ghani, "Automated egg grading system using computer vision: Investigation on weight measure versus shape parameters," *IOP Conf. Ser. Mater. Sci. Eng*, vol. 342, 2018.

[3] J. R. Chambers, K. Zaheer, H. Akhtar, and E. S. M. Abdel-Aal, "Chicken Eggs," in *Egg Innovations and Strategies for Improvements*, Elsevier Inc., 2017, pp. 3–11.

[4] Jeremy, K. Thomas, and D. W. Engels, "Identification and Classification of Poultry Eggs: A Case Study Utilizing Computer Vision and Machine Learning," 2019.

[5] J. Thipakorn, R. Waranusast, and P. Riyamongkol, "Egg weight prediction and egg size classification using image processing and machine learning," in *ECTI-CON 2017 - 2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications, and Information Technology*, 2017, pp. 477–480.

[6] M. A. Rahman, M. T. Haque, C. Shahnaz, S. A. Fattah, W. P. Zhu, and M. O. Ahmed, "Skin lesions classification based on color plane-histogram-image quality analysis features extracted from digital images," in *Midwest Symposium on Circuits and Systems*, 2017, vol. 2017-August, pp. 1356–1359.

[7] L. Huang, A. He, M. Zhai, Y. Wang, R. Bai, and X. Nie, "A multi-feature fusion based on transfer learning for chicken embryo eggs classification," *Symmetry (Basel).*, vol. 11, no. 5, pp. 1–15, 2019.

[8] T. Sutojo, P. S. Tirajani, D. R. I. M. Setiadi, C. A. Sari, and E. H. Rachmawanto, "CBIR for classification of cow types using GLCM and color features extraction," in *International conferences on Information Technology, Information Systems, and Electrical Engineering*, 2017, pp. 182–187.

[9] O. R. Indriani, E. J. Kusuma, C. A. Sari, E. H. Rachmawanto, and D. R. I. M. Setiadi, "Tomatoes classification using K-NN based on GLCM and HSV color space," in *Proceedings - 2017 International Conference on Innovative and Creative Information Technology: Computational Intelligence and IoT, ICITech 2017*, 2018, vol. 2018-Janua, pp. 1–6.

[10] T. T. Htay and S. S. Maung, "Early Stage Breast Cancer Detection System using GLCM feature extraction and K-Nearest Neighbor (k-NN) on Mammography image," in *International Symposium on Communications and Information Technologies*, 2018, pp. 171–175.

[11] A. Kusuma, D. R. I. M. Setiadi, and M. D. M. Putra, "Tomato Maturity Classification using Naive Bayes Algorithm and Histogram Feature Extraction," *J. Appl. Intell. Syst.*, vol. 3, no. 1, pp. 39–48, Aug. 2018.

[12] Riandini and M. K. Delimayanti, "Feature extraction and classification of thorax x-ray image in the assessment of osteoporosis," in *International Conference on Electrical Engineering, Computer Science and Informatics*, 2017, pp. 1–5.

[13] C. Irawan, W. Listyaningsih, D. R. I. M. Setiadi, C. Atika Sari, and E. Hari Rachmawanto, "CBIR for Herbs Root Using Color Histogram and GLCM Based on K-Nearest Neighbor," in *Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, iSemantic 2018*, 2018, pp. 509–514.

[14] E. Hamuda, B. Mc Ginley, M. Glavin, and E. Jones, "Automatic crop detection under field conditions using the HSV colour space and morphological operations," *Comput. Electron. Agric.*, vol. 133, pp. 97–107, Feb. 2017.

[15] C. Irawan, E. N. Ardyastiti, D. R. I. M. Setiadi, E. H. Rachmawanto, and C. A. Sari, "A Survey: Effect of the Number of GLCM Features on Classification Accuracy of Lasem Batik Images using K-Nearest Neighbor," in *International Seminar on Research of Information Technology and Intelligent Systems*, 2018, pp. 33–38.

[16] P. N. Andono, T. Sutojo, and M. Muljono, *Pengolahan Citra Digital*. 2017.