Corn Seeds Identification Based on Shape and Colour Features

Haddad Alwi Yafie[1,2], Ema Rachmawati[1], Esa Prakasa[2*], Amin Nur[3]

[1]School of Computing
Telkom University
Bandung
[2]Computer Vision Research Group
Research Center for Informatics
Indonesian Institute of Sciences (LIPI)
Bandung
[3]Assessment Institutes for Agricultural Technology of Gorontalo
Ministry of Agriculture of the Republic of Indonesia
Gorontalo
*Correspondence: esa.prakasa@lipi.go.id

Abstract-Corn is one of the agricultural products that are essential as daily food sources or energy sources. Corn selection or sorting is important to produce high-quality seeds before its distribution to areas with varying conditions and agricultural characteristics. Hence, it is necessary to build corn seeds identification. In this paper, we propose a corn seed identification technique that incorporates the advantage of combining shape and color features. The identification process consists of three main stages, namely, ROI selection, feature extraction, and classification using the Artificial Neural Network (ANN) algorithm. The shape feature originates from the eccentricity value or comparison value between a distance of minor ellipse foci and major ellipse foci of an object. Meanwhile, the color features are extracted based on the HSV (Hue-Saturation-Value) channel. The experimental result shows that the proposed system achieves excellent performance for the identification of poor and good corn quality for BIMA-20 and NASA-29 species. The classification result for BIMA-20 Good vs. BIMA-20 Bad gives an accuracy of 89%, while the classification accuracy of BIMA-20 Good vs. NASA-29 Good is 97%.

Keywords: artificial neural network, eccentricity, feature extraction, region of interest

1. Introduction

In the field of food and agriculture technology, to meet production requirements is very difficult. the difficulty is related to the level of product diversity and its non-homogeneous nature. To assess the quality of products or goods, it is necessary to consider the feature specifications including corn [1].

Corn is one of the agricultural products that can be used either as essential food sources in daily life or energy sources. The selection or sorting process must be carried out to produce quality seeds which will then be distributed to areas with varying conditions and agricultural characteristics. Prakasa et al. have successfully developed a system to automatically detect a region of interest (ROI) using the K-means algorithm. ROI will produce images containing only one corn seed by determining the location

and boundary box of each corn seed in the input image. Based on the test results, the model is proven to be able to detect ROI with an accuracy exceeding 90% [2].

It is very important to check the quality of corn seeds so that they are safe and of a high standard. Traditionally inspecting corn seeds way is very time-consuming and also depends on the skills of employees/humans. it would be very interesting to be able to control human error and also expect the same quality from many staff with various levels of skill [3].

Feature extraction has a fundamental role in classification techniques[4]. The colour value is one of the object features that widely used in classification algorithms. The colour feature is significant in determining the quality of any fruit[5]. Colour characteristics can be represented in a variety of colour representations or commonly referred to as colour space. There are several colour spaces used to

represent colour values. One of the colour space is called HSV (hue, saturation, value). The HSV is a result of the geometrical transformation from The RGB (red, green, blue) colour map[6]. Another feature that can be used to characterize the physical properties is object shape. The shape features can be represented in roundness, convexity, compactness, elongation, eccentricity, and sphericity [7]. Use the shape feature to classify pollen grains [8].

Moallem et al. [4], in their research, have also succeeded in identifying healthy apples and defective apples by applying several segmentation algorithms such as background removal, stem tip detection, calyx detection, primary segmentation defects, and repair of defective areas. By combining statistical, texture, and geometric features, they achieve excellent classification result for healthy apples of 91.75%, and defective apples is an average of 91.5%.

Xiao Ling et al. classifying waxy corn kernels based on a combination of spectral, morphological, and texture features. This feature is extracted from hyperspectral images that look and are almost infrared. Morphological features are used as many as five features, including area, circularity, aspect ratio, roundness, and solidity. Then the texture features used amounted to eight features including energy, contrast, correlation, entropy, and standard deviation. Both features were extracted as the appearance features of each peeled corn. Furthermore, the model was built for the classification of corn seed varieties based on various feature groups, namely Support vector machine (SVM) and partial least squares-discriminant analysis (PLS-DA). The accuracy results achieved using the SVM model are higher than using the PLS-DA model [9].

According to the grain-grading standards, the corn quality is classified based on several conditions, such as moisture, weight, colour, shape, odour, and damage. Among these criteria, the moisture, weight, and odor of corn can still be evaluated using electronic analyzers or other special instruments. Moreover, the damage can still be seen manually. Therefore, the assessment can introduce some errors. A plausible way to classify corn is by using computer vision to classify corn automatically [10].

In this study, we propose the identification of corn seeds using the combination of shape and color features. The data used in the study consisted of three categories, BIMA-20 Good, BIMA-20 Bad, and NASA-29 Good.

2. Method

Figure 1 shows our proposed corn seed identification. According to Figure 1, the corn seed image data will be divided into training and testing data. The corn image data contains many corn seeds in one image. The first step is carried out by preprocessing corn image data and detecting the region of interest (ROI) to produce images containing only one corn seed. The next step will be feature extraction based on shape and colour. Then at the last stage, the classification will be carried out; the corn seed data extracted from the training data will be used to conduct

training with ANN algorithm that produces output in the form of a classification model. The classification model will then be used to test the test data based on the shape and colour features that have been extracted.
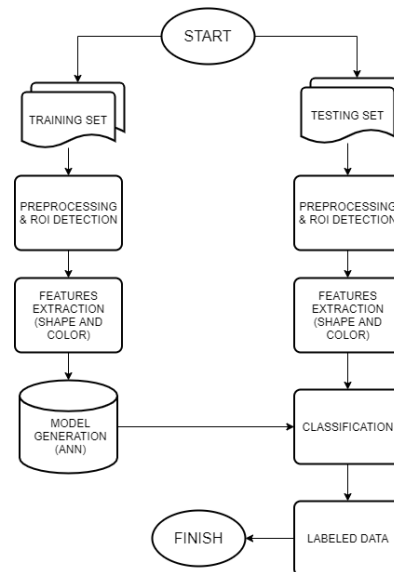

Figure 1: Flowchart System

a. Capturing the corn image data

The capturing process of corn image data is conducted at the Laboratory of Computer Vision Research Group, Research Center for Informatics - Indonesian Institute of Sciences (LIPI) Bandung, with a modified apparatus from the previous research[2]. The con seed samples are provided by the Assessment Institutes for Agricultural Technology of Gorontalo, Ministry of Agriculture Republic of Indonesia. The method of capturing corn image data using a Canon EOS 70D DSLR camera, with the lighting arranged so that the corn can be seen clearly. The apparatus arrangement is displayed in Figure 2.
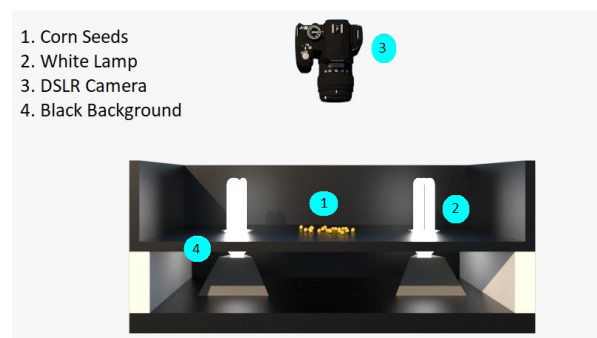

Figure 2: Environment setting corn seed imaging

b. Detection of the region of interest (ROI) and Preprocessing

The purpose of ROI detection is to obtain a single corn seed in each image. Prakasa et al.[2] had successfully developed an algorithm to detect the ROI automatically using the k-means algorithm. The ROI will produce images containing only one corn seed by determining the location

and boundary box of each corn seed in the input image. Figure 3 is an input image in the RGB channel, which is converted into a red channel, as shown in Figure 4. Figure 5 shows the preprocessing median filtering done to eliminate noise in the image. Figure 6 shows the result of segmenting the corn seed from background objects by calculating the threshold value using the Otsu method. The result of the erosion process is depicted in Figure 7.
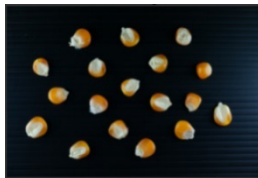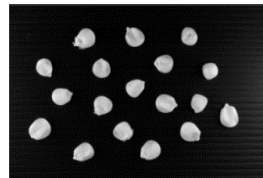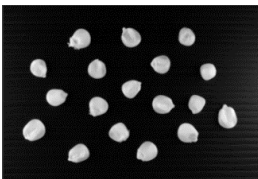

Figure 3: RGB Channel

Figure 4: Red Channel
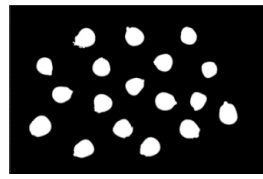

Figure 5: Median Filter
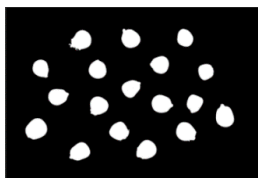

Figure 6: Otsu Thresholding

Figure 7: Erode

c. Feature Extraction

The feature used in this study is 2 (two) types, namely, shape and colour feature. We use the eccentricity value to represent the shape feature and HSV value to represent the colour feature.

1) Shape feature: eccentricity

Shape features are physical dimensional measures that characterize the appearance of kernels [11]. Eccentricity is the comparison value between the distance of the minor ellipse foci with the major ellipse foci of an object with the eccentricity value of an elongated object or straight-line shape is close to one. In contrast, a circular object will have an eccentricity value near to zero [12][13].

$$Ecc = \sqrt{1 - (minor\ axis\ /\ major\ axis)^2} \qquad (1)$$

Figure 8 shows an example of the minor and major axis. The minor axis is the shortest line (line p3 to p4), while the major axis is the longest line (line p1 to p2). The length of the two lines is calculated using Euclidean distance [14].

2) Colour feature: HSV

The colour value is one of the essential indicators to measure the quality of fruit appearance [15][16].

To distinguish an object with a particular colour, we can use the Hue value, which is a representation of visible light (red, orange, yellow, green, blue, purple). Hue can be combined with Saturation and Value, which is the brightness level of colour. The original RGB value is converted to HSV to obtain the values of these three parameters [17][18]using linear support vector machine (SVM.
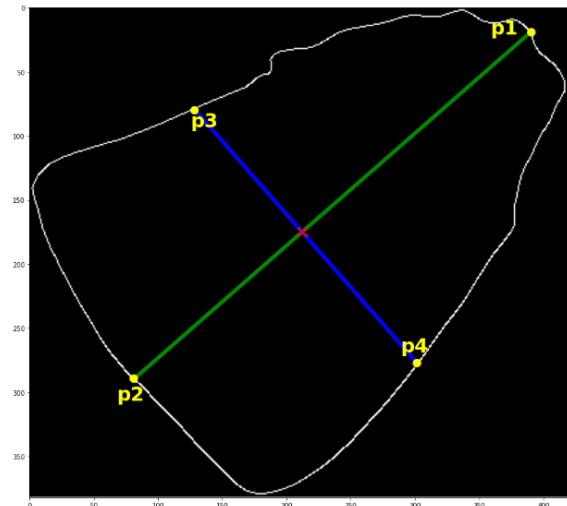

Figure 8: Minor axis & Major axis

d. Artificial Neural Network

Artificial Neural Network (ANN) is a method that models the nervous system of the human brain or commonly referred to as neurons, while the task is to introduce patterns, especially assemblages. This model is based on the ability of the human brain to organize neurons so that they can improve patterns effectively [6]. The ANN characteristics can be seen from the pattern of relationships between neurons, the method of determining the weights of each connection, and its activation function. In general, ANN consists of :

1) An input layer, which contains several input neurons that functions to send data
2) Hidden layer, which is where the primary process of ANN occurs
3) Output layer, which contains several neuron outputs
4) Activation function.

The ANN process starts from the input received along with the weight value, after entering the input values that will be added by a propagation function. The results of the addition will be processed by the activation function of each neuron. Then the results will be compared with a certain threshold. If the value exceeds the threshold, the activation of neurons will be canceled. Conversely, if it is still below the threshold value, the neurons will be activated. Once active, the neuron will send the output value to the output layer [7][19][20].

Table 1 The Result of Features Extraction

No	Shape			Colour			Label
	Major axis	Minor axis	Eccentricity	H	S	V	
1	10309	9351	0.421	26167	163251	144307	BIMA20_GOOD
2	9735	9138	0.345	28347	157900	141916	BIMA20_GOOD
3	9785	9726	0.109	29348	161522	149893	BIMA20_GOOD
4	10406	8885	0.521	23477	199624	141244	BIMA20_GOOD
5	10122	9675	0.294	24821	207008	137560	BIMA20_GOOD
6	9748	8719	0.447	26137	167724	132771	BIMA20_GOOD
7	10542	9543	0.425	29248	157730	147806	BIMA20_GOOD
8	10458	8658	0.561	25941	168078	119700	BIMA20_GOOD
9	10301	10283	0.06	25993	186081	140040	BIMA20_GOOD
10	10309	9351	0.421	26167	163251	144307	BIMA20_GOOD

3. Result and Discussion

The following is the amount of corn seed images that we used in the experiment:
a. BIMA-20 Good: 632 images
b. BIMA-20 Bad: 487 images
c. NASA-29 Good: 488 images

Figure 9 shows some examples of corn images used in the experiment. From each image, we extract shape and colour feature, with the result examples can be seen in Table 1. We split the data into training and testing data with 70% and 30% for training and testing data, respectively.
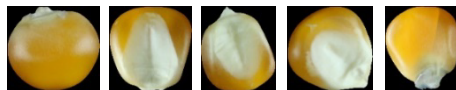

Figure 9: Corn seeds image that have been extracted

The setting of the ANN parameter used in this paper consists of three hidden layers and one output layer. The number of neurons in each layer is as follows: layer 1= 128 neurons, layer 2= 128 neurons, and layer 3= 64 neurons. The setting is selected because it was the most optimal and produced the highest accuracy. We have tried to implement the neurons setting as follows; layer 1 = 32 neurons, layer 2 = 32 neurons, and layer 3 = 16 neurons. However, the accuracy obtained by using this setting was 57% for BIMA-20 Good vs. BIMA-20 Poor and 71% in the classification of BIMA-20 Good vs. NASA-29 Good. We also have used another setting as follows; layer 1 = 256 neurons, layer 2 = 256 neurons, layer 3 = 128. The accuracy of this setting is only 70%.

We investigate two scenarios in our corn seed identification: (1) classification of BIMA-20 Good vs. BIMA-20 Bad, and (2) classification of BIMA-20 Good vs. NASA-20 Good. The results of the classification by the ANN method are summarized in a format called a confusion matrix. Confusion matrices are commonly used to describe the performance of a classification method whose actual class is known[12]. The data in the confusion matrix shows the number of class predictions that correspond to the actual class. F1 score is used to measure classification performance by applying Equation 2 [21], where TP, TN, FP, and FN are the number of true -positives, true-negatives, false-positives, and false-negatives, respectively[1].

$$F1\ Score = \frac{TP+TN}{TP+TN+FP+FN} \qquad (2)$$

a. Classification result of BIMA-20 Good vs. BIMA-20 Bad

In this section, we classify the corn seed of BIMA-20 Good and BIMA-20 Bad classes. The results of classification using the ANN method and summarized in the form of a confusion matrix is shown in Figure 10, Figure 11, and Figure 12. True Positive is the amount of data from the BIMA-20 Good class that was successfully classified correctly into the Good class. True Negative is the amount of data from the BIMA-20 Bad class that was successfully classified correctly into the Bad class. False-positive is the amount of data from the BIMA-20 Bad class that was not successfully classified into the Bad class. False Negative is the amount of data from BIMA-20 Good, which was not successfully classified into the Good class.
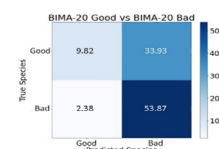

Figure 10: Confusion matrix BIMA-20 Good vs. BIMA-20 Bad using shape feature

The classification result which is shown in Figure 10 only uses the eccentricity properties of the corn seed with the accuracy is 64%. In contrast, the classification result

using colour features is depicted in Figure 11, in which by using H, S, and V we achieve a classification accuracy of 88%. Meanwhile, accuracy results in Figure 12 with feature extraction using shape and colour obtained from BIMA-20 Good and BIMA-20 Bad data that is equal to 89%.
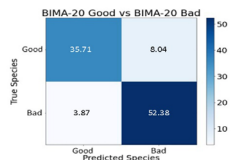

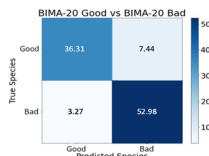**Figure 11: Confusion matrix BIMA-20 Good vs. BIMA-20 Bad using colour feature**

Figure 12: Confusion matrix BIMA-20 Good vs. BIMA-20 Bad using shape and colour feature

b. BIMA-20 Good vs. NASA-29 Good

In this section, we show the classification result of BIMA-20 Good vs. NASA-29 Good. The results of classification using the ANN method and summarized in the form of a confusion matrix, as shown in Figure 13, Figure 14 and Figure 15. True positive is the amount of data from the BIMA-20 Good class that was successfully classified correctly into the Good class. True Negative is the amount of data from the NASA-29 Good class that was successfully classified correctly into the Good class. False-positive is the amount of data from the NASA-29 Good class that was not successfully classified into the Good class. False Negative is the amount of data from BIMA-20 Good that was not successfully classified into the Good class.

In Figure 13, we only use the shape feature extraction, namely eccentricity, with an accuracy of 71%, whereas, in Figure 14 we use colour feature extraction, namely H, S, and V values, with an accuracy of up to 96%. Meanwhile, accuracy results in Figure 15 with feature extraction using shape and colour obtained from the BIMA-20 Good vs. NASA-29 Good data that is equal to 97%.
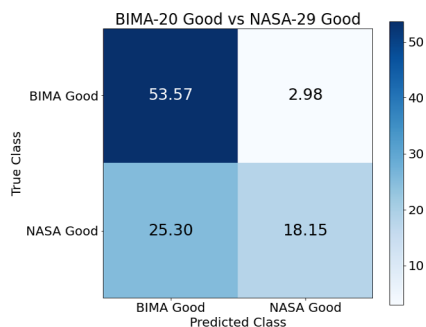

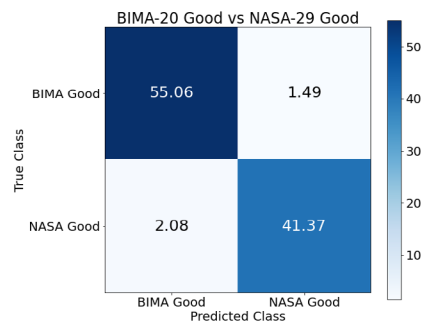**Figure 13: Confusion matrix BIMA-20 Good vs. NASA-29 Good using shape feature**

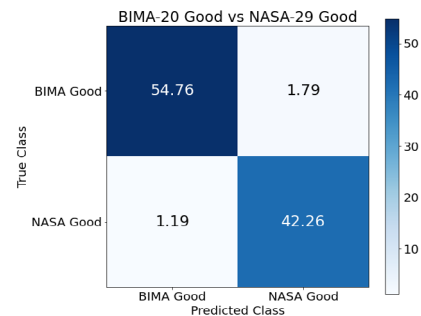**Figure 14: Confusion matrix BIMA-20 Good vs. NASA-29 Good using colour feature**

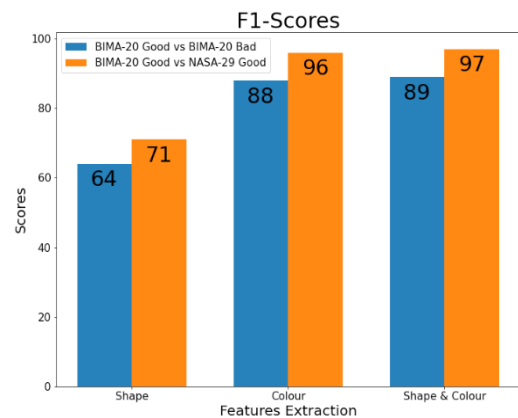**Figure 15: Confusion matrix BIMA-20 Good vs. NASA-29 Good using shape and colour feature**

Figure 16: F1-Scores Diagram

The F1 score is shown in Figure 16. Based on Figure 16, the classification results between BIMA-20 Good vs. NASA-29 Good are better, both in terms of shape features, colour features, or shape, and colour features. However, the colour feature classification results are better than the shape feature classification results. Colour features have a strong influence in terms of this classification.
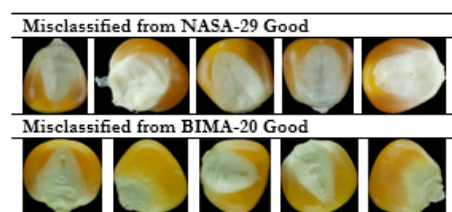

Figure 17: The example of misclassification cases.

Figure 17 is an example of a picture that was misclassified from NASA-29 Good vs. BIMA-20 Good. The misclassification is because the colour of corn seeds from NASA-29 Good has a brighter colour than corn seeds from BIMA-20 Good.

4. Conclusion

In this study, we proposed corn seed identification using color and shape features. We used eccentricity as the shape feature and HSV channel as the color feature. Based on the results obtained in the experiment, we can conclude that the proposed algorithm is performed better in classifying the corn varieties at the same quality than quality grading. In this case, the using of colour feature produce a higher degree of accuracy than the shape feature. This condition is caused due to the shape of corn seed is relatively the same.

Data Availability

The corn images are available in the RIN (Repositori Ilmiah Nasional/National Scientific Repository) of the Indonesian Institute of Sciences. The images are stored as three datasets, BIMA-20 Good, BIMA-20 Bad, and NASA-29 Good. The dataset can be found in the following link: https://data.lipi.go.id/dataverse/seed-grading

Author Contributions

All authors are equally contributed in the paper preparation.

References

[1] P. Daskalov, E. Kirilova, and T. Georgieva, 'Performance of an automatic inspection system for classification of Fusarium Moniliforme damaged corn seeds by image analysis', MATEC Web Conf., vol. 210, pp. 1–10, 2018, doi: 10.1051/matecconf/201821002014.

[2] E. Prakasa et al., 'Automatic region-of-interest selection for corn seed grading', Proc. - 2017 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Comput. Sci. Eng. IC3INA 2017, vol. 2018-Janua, pp. 23–28, 2017, doi: 10.1109/IC3INA.2017.8251734.

[3] K. Kiratiratanapruk and W. Sinthupinyo, 'Color and texture for corn seed classification by machine vision', 2011 Int. Symp. Intell. Signal Process. Commun. Syst. "The Decad. Intell. Green Signal Process. Commun. ISPACS 2011, pp. 7–11, 2011, doi: 10.1109/ISPACS.2011.6146100.

[4] P. Moallem, A. Serajoddin, and H. Pourghassem, 'Computer vision-based apple grading for golden delicious apples based on surface features', Inf.

[5] K. Padmavathi, 'Investigation and monitoring for leaves disease detection and evaluation using image processing', Int. Res. J. Eng. Sci. Technol. Innov., vol. 1, no. 3, pp. 66–70, 2012.

[6] M. R. Satpute and S. MJagdale, 'Color, Size, Volume, Shape and Texture Feature Extraction Techniques for Fruits: A Review', Int. Res. J. Eng. Technol., no. 2010, pp. 2395–56, 2016.

[7] N. Abdellahhalimi, Roukhe, A., Abdenabi, B., & El Barbri, 'Sorting Dates Fruit Bunches Based on Their Maturity Using Camera Sensor System', J. Theor. Appl. Inf. Technol., 2013.

[8] M. A. (University of G. Wirth, 'Shape Analysis & Measurement Shape Analysis & Measurement', Image Process., pp. 1–49, 2004.

[9] X. Yang, H. Hong, Z. You, and F. Cheng, 'Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification', Sensors (Switzerland), vol. 15, no. 7, pp. 15578–15594, 2015, doi: 10.3390/s150715578.

[10] X. Li, B. Dai, H. Sun, and W. Li, 'Corn classification system based on computer vision', Symmetry (Basel)., vol. 11, no. 4, 2019, doi: 10.3390/sym11040591.

[11] X. Chen, Y. Xun, W. Li, and J. Zhang, 'Combining discriminant analysis and neural networks for corn variety identification', Comput. Electron. Agric., vol. 71, no. SUPPL. 1, pp. 48–53, 2010, doi: 10.1016/j.compag.2009.09.003.

[12] K. Markham, 'Simple guide to confusion matrix terminology', Data Sch., 2015.

[13] E. Hari Rachmawanto, G. Rambu Anarqi, D. R. I. Moses Setiadi, and C. Atika Sari, 'Handwriting Recognition Using Eccentricity and Metric Feature Extraction Based on K-Nearest Neighbors', Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018, pp. 411–416, 2018, doi: 10.1109/ISEMANTIC.2018.8549804.

[14] L. Boyd, S., & Vandenberghe, Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares Book. 2017.

[15] Y. Ji, Q. Zhao, S. Bi, and T. Shen, 'Apple Grading Method Based on Features of Color and Defect', Chinese Control Conf. CCC, vol. 2018-July, pp. 5364–5368, 2018, doi: 10.23919/ChiCC.2018.8483825.

[16] M. Jhuria, A. Kumar, and R. Borse, 'Image processing for smart farming: Detection of disease and fruit grading', 2013 IEEE 2nd Int. Conf. Image Inf. Process. IEEE ICIIP 2013, pp. 521–526, 2013, doi: 10.1109/ICIIP.2013.6707647.

[17] H. Toylan and H. Kuscu, 'A real-time apple grading system using multicolor space', Sci. World J., vol. 2014, 2014, doi: 10.1155/2014/292681.

[18] A. Mizushima and R. Lu, 'An image segmentation method for apple sorting and grading using support vector machine and Otsu's method', Comput. Electron. Agric., vol. 94, pp. 29–37, 2013, doi: 10.1016/j.compag.2013.02.009.

[19] M. P. Arakeri and Lakshmana, 'Computer Vision Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry', Procedia Comput. Sci., vol. 79, pp. 426–433, 2016, doi: 10.1016/j.procs.2016.03.055.

[20] Y. Liu, A. Ouyang, J. Wu, and Y. Ying, 'An automatic method for identifying different variety of rice seeds using machine vision technology', Opt. Sensors Sens. Syst. Nat. Resour. Food Saf. Qual., vol. 5996, no. Icnc, p. 59961H, 2005, doi: 10.1117/12.631004.

[21] F. F. Zain and Y. Sibaroni, 'Effectiveness of SVM Method by Naïve Bayes Weighting in Movie Review Classification', Khazanah Inform. J. Ilmu Komput. dan Inform., vol. 5, no. 2, pp. 108–114, 2019, doi: 10.23917/khif.v5i2.7770.