ISSN 2278-3091
Volume 10, No.2, March - April 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse651022021.pdf
https://doi.org/10.30534/ijatcse/2021/651022021

Ripeness Level Classification of Banana Fruit Based on Hue Saturate Value (HSV) Color Space Using K-Nearest Neighbor Algorithm

Lia Kamelia[1], Mufid Ridlo Effendi[2], Nisa Hafidzatul Adila[3]
[1]Department of Electrical Engineering UIN Sunan Gunung Djati Bandung, Indonesia,lia.kamelia@uinsgd.ac.id
[2] Department of ElectricalEengineering UIN Sunan Gunung Djati Bandung, Indonesia,mufid.ridlo@uinsgd.ac.id
[3] Department of Electrical Engineering UIN Sunan Gunung Djati Bandung, Indonesia

ABSTRACT

Many types of bananas are cultivated locally in Indonesia, including the Muli Banana or Musa Acuminata Linn. During the post-harvest period of banana fruit, there is a problem in the sorting process of bananas based on their level of maturity. The fruit sorting process manually uses the human eye, but it is ineffective due to decreased vision and the large quantity of fruit. Therefore, we need a system that can quickly classify the ripeness of the banana fruit. This study aims to create a system that can organize the maturity level of the banana fruit. The classification system designed using the HSV color feature extraction method and the K-Nearest Neighbor classification algorithm. After going through the testing phase, the system can classify bananas into three classes: unripe, ripe, and rotten. System testing used 30 test data images, and the results show 2 test images whose classification results are wrong and 28 other test images whose classification results are correct. Based on calculations, the accuracy achieved by the system is 93.333%.

Key words: ***banana, HSV, KNN, ripeness***.

1. INTRODUCTION

Banana is a type of fruit often found in Indonesia. It is easy to find this fruit because it is sold in various places ranging from small stalls, traditional markets to modern markets. One type of banana that is most commonly found in Indonesia is the muli banana. Muli banana (Musa acuminata Linn) has a small size with a length of 9 cm and a diameter of 10.5 cm. The color of the fruit skin is yellow, the taste of the fruit is sweet and fragrant, and it has a high content of vitamins, minerals, and carbohydrates. Because of the many benefits, bananas are consumed by many people and are widely cultivated[1]. One of the post-harvest problems for large-scale fruit products is the sorting process for the following distribution process[2].

Image processing is a technique that can be used to process images by converting them into the desired digital image data to obtain specific information [3]. Image processing is a method or technique to process pictures or images by manipulating them into the selected image data to get precise information. Image processing applications make it easy to process an image.

The use of proven image processing technology can improve accuracy in the sorting process of fruit ripeness. There have been several studies that utilize image processing for the classification process of fruit maturity[4],. Reference [5] proposes a system for automatic grading of oranges using a pattern recognition technique applied to a single color image of the fruit. The other research in the ripeness of fruit held by Opena and Yusiong. In research [6] introduced an automatic tomato classification system using Artificial Neural Network (ANN) classifiers that were trained using the Artificial Bee Colony (ABC) algorithm. The maturity level also studied in citrus fruit using the Particle Swarm Optimization (PSO) algorithm to optimize an Artificial Neural Network. The research could categorize between the ripe and unripe levels of C*itrus Suhuensis*. The algorithm would adjust the network connections weights and adapt its values during training for the best output results [7].

The ripeness of fruit usually can be seen on its color[8]. The color space transformation method is one way of image processing carried out to obtain the color space variety of an image in a particular color coordinate system [9]. Although the RGB base is good for displaying color information, it is not suitable for image processing applications. In object recognition applications, it is easier to identify objects with the difference in hue is by providing a threshold value in the range of hue values (spectrum wavelength) that surrounds the thing [10],[11]. The HSV color space is very effective and takes the most similar images for color extraction, and this will increase the speed of the search system. The HSV color space is very close to human visual perception. The hue component is more dominant than the saturation/value component so that the extraction of HSV colour features is better compared to other colour spaces [12].

The research uses a classic-algorithm, kNN because it has several advantages. KNN is compatible with high-dimensional objects. The data represented is adapted to a simple structure and is easy to apply, especially for geometric learning[13]. The Nearest Neighbor algorithm is the simplest of all machine learning algorithms. The principle is to memorize the training data and then predict the class/level of each new instance based on the class of its closest neighbors in the training data. The method is based on the assumption that the features used to describe a point domain are relevant to labeling to make the closest points more likely to have the same label/class. Also, in some situations, even when training data is plentiful, finding the nearest neighbors can be done very quickly[14]. The kNN algorithm widely used for the classification in various sector. kNN algorithm has proved as the simplest and accurate methods to classified the object into several level / class, such as for water quality level[15], detection of the natural deficiency level in the plants[16], the detection of animal tracking system[17], the detection of pesticide residues [18]and the classification of orange varieties[19].

The research proposed a simple classification for the maturity level of banana fruit using kNN algorithm to obtain the experiment's speed test based on the explanation above. The extraction process base on the HSV color space because it's performance better than other color space.

2. RESEARCH METHODOLOGY

The classification system for the maturity level of the banana fruit created using the MATLAB 2017a software by utilizing the Graphical User Interface (GUI) feature. GUI features a display application from the MATLAB software that contains tasks, commands, or program components that make it easier for users to run a MATLAB program. Some of the commands stored in the GUI include controls for extracting training data, for removing test data, for displaying the selected image, for displaying the values of classification variables in the image, and for showing the classification results of the maturity level of the banana fruit.

The input image has the RGB color space. In this research, the value of the HSV color space is used for the classification process. Images that are still in the RGB color space are first carried out by the separation process of each color component, namely Red, Green, and Blue so that the system can determine the value of each color component. After separating the values from the Red, Green, and Blue components, the next step is to convert the values from the RGB color space to the HSV color space. Mathematical calculations do value conversion to find each component's rate in the HSV color space, namely Hue, Saturation, and Value. In designing the system of the classification system for the level of maturity of a banana, several stages of image processing were carried out to classify the ripeness level of the banana fruit.

Classification results are obtained from the KNN algorithm process and compare training data with test data. The value of training data will be a benchmark in determining a level. The system will look for the value of the training data that is closest to the test data. It is from the nearest value that the classification results are obtained. Two types of data needed in the research process are training data and test data. Training data will be stored in a database whose value is used as a differentiator for each class. The test data is used as a test sample to see whether the system is functioning correctly or not. The test data and training data each consist of 30 images of bananas. The 30 images are divided into three levels: the unripe level, the mature level, and the rotten level so that each data consists of 10 images. **Figure 1** shows the process of the entire system.

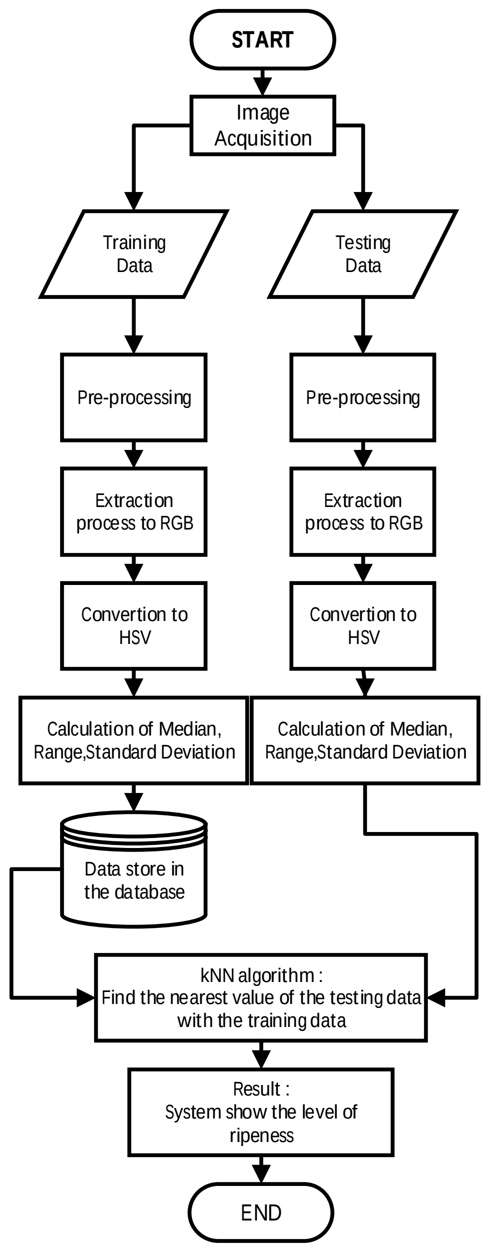
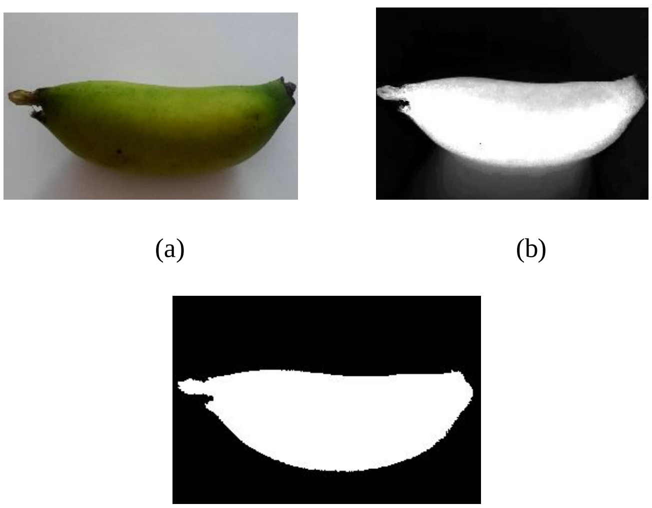


Figure 1: The Flowchart of the Research

3. RESULTS AND DISCUSSION

The first image processing is the segmentation process. The purpose of the segmentation process is to separate the banana image from its background so when calculating the color value, it will not mix with the background color. The segmentation process needs a saturation component of the HSV color space so that the image needs to be converted first from the RGB color space. There are several stages in the segmentation process. Initially, the image in the RGB color space is transformed into the HSV color space to obtain a picture with the saturation component. The original image in the RGB color space can be seen in Figure 2 (a), while the HSV component image can be seen in Figure 2 (b). After the system obtains the saturation component image, it will transform into a binary image. The binary image is an image that only consists of 2 color values, namely 0 (white) and 1 (black), as can be seen in Figure 2 (c).



(a) (b)

(c)

Figure 2 : Image with (a) RGB volour (b) HSV colour (c) biner colour

Binary images can distinguish objects and their backgrounds. The item (banana) is a color that has a value of 0. At the same time, the background is a color that has a value of 1.After the object is separated from its background, the next step is to return the image to the RGB color space again for the final segmentation results, as can be seen in Figure 3.

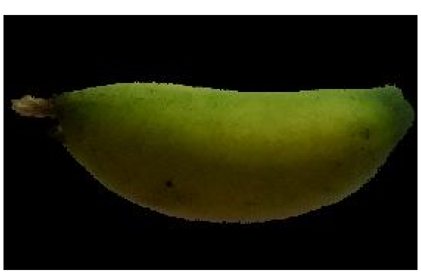
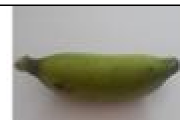
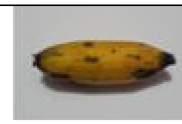
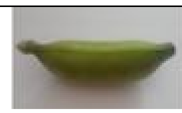
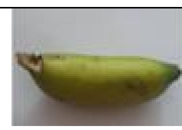
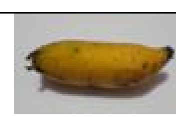
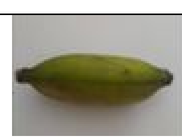


Figure 3: Segmented image

After the banana image has been segmented, the next process is to find the hue, saturation, and values of the segmented image. The hue, saturation, and value values are obtained from the calculation using the red, green, and blue value components of the image. The source code to get the hue, saturation, and value values is obtained the mathematical equations as in displayed (1) **-** (3).

Table 1 : Training data

No	Unripe	Ripe	Rotten
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

$$H = \tan\left(\frac{3\,(G-B)}{(R-G)+(R-B)}\right) \qquad (1)$$

$$S = 1 - \left(\frac{\min(R,G,B)}{V}\right) \qquad (2)$$

$$V = \frac{R+G+B}{3} \qquad (3)$$

The total number of training data is 30 images consisting of 3 categories of unripe, ripe, and rotten, as shown in **Table 1**.

24 types of data are stored and used in the classification process. 24 This type of data is the data that will be compared in value between the value of the training data and the test data using the KNN algorithm. 24 The types of data are:

1. Mean of the components R, G, B, H, S, and V.
2. the variance of the components R, G, B, H, S, and V
3. range of components R, G, B, H, S, and V
4. the standard deviation of the components R, G, B, H, S, and V.

Table 2: Mean value of the training data

Level	Result	*Mean* R	*mean* G	*Mean* B	*Mean* H	*Mean* S	*Mean* V
Unripe	Min	0.021586295	0.021645601	0.002816837	0.01312055	0.062830361	0.015935734
	Max	0.041945478	0.039441386	0.010113199	0.018389598	0.101332058	0.030500021
	Average	0.028350648	0.027145469	0.005750089	0.015692623	0.083107845	0.020415402
Ripe	Min	0.030042677	0.016567531	0.000957889	0.007145004	0.083056306	0.015856032
	Max	0.059865568	0.035567226	0.003091485	0.012703067	0.12979458	0.032841426
	Average	0.045133986	0.027436545	0.001816244	0.00970571	0.097593806	0.024795592
Rotten	Min	0.022405507	0.010730467	0.001736601	0.003930059	0.054236381	0.012034787
	Max	0.066799851	0.039854827	0.009757405	0.010495173	0.101348414	0.03873626
	Average	0.039910134	0.020790066	0.004910788	0.006560536	0.083855791	0.021870329

Table 3: Variance of the training data

Level	Result	*Variance* R	*Variance* G	*Variance* B	*Variance* H	*Variance* S	*Variance* V
Unripe	Min	0.006067633	0.006099146	0.000248617	0.001983955	0.049051088	0.003252649
	Max	0.015364692	0.015038628	0.001889088	0.002501919	0.081670307	0.009016474
	Average	0.008854375	0.008812363	0.000911412	0.002236898	0.065093294	0.005059669
Ripe	Min	0.0113447	0.00389027	0	0.000589731	0.072338967	0.003256375
	Max	0.025158819	0.01097062	0.000264297	0.001044377	0.104308782	0.008011202
	Average	0.020378686	0.008199104	0.000148721	0.000861493	0.082244433	0.00624825
Rotten	Min	0.721568627	0.509803922	0.317647059	0.166666667	1	0.496732026
	Max	1	1	1	0.5	1	1
	Average	0.853333333	0.694901961	0.529019608	0.440555926	1	0.675424837

The results from the extraction of 30 training data images that have been extracted and counted 24 types of data can be seen in Table 2 - 5.

The KNN algorithm's principle is to look for the closest value or the most similar tendency between the test data and the training data. The system will look for which class of training data has the most relative value of the 24 data types. From testing the test image, the system notifies that the test image's banana belongs to the raw class. The classification results correspond to the actual condition of the banana in its unripe state.

Table 4: Range of the training data

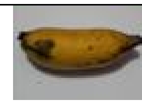
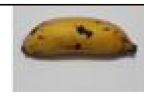
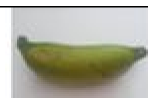
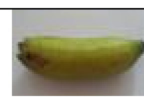
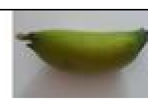
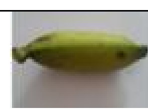
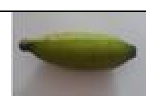
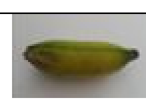
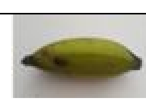
Level	Result	*Range* R	*Range* G	*Range B*	*Range* H	*Range* S	*Range* V
Unripe	Min	0.525490196	0.509803922	0.294117647	0.246718745	1	0.411764706
	Max	0.760784314	0.768627451	0.552941176	0.475143679	1	0.674509804
	Average	0.643137255	0.649019608	0.388235294	0.35498895	1	0.552418301
Ripe	Min	0.631372549	0.450980392	0.266666667	0.386407237	1	0.405228758
	Max	1	0.968627451	0.752941176	0.475143679	1	0.907189542
	Average	0.761960784	0.592941176	0.383137255	0.435116326	1	0.536339869
Rotten	Min	0.721568627	0.509803922	0.317647059	0.166666667	1	0.496732026
	Max	1	1	1	0.5	1	1
	Average	0.853333333	0.694901961	0.529019608	0.440555926	1	0.675424837

Table 5: Standard deviation of the training data

Level	Result	SD R	SD G	SD B	SD H	SD S	SD V
Unripe	Min	0.07789501	0.078097032	0.015767595	0.044541615	0.221474802	0.057032002
	Max	0.123954395	0.122632083	0.043463644	0.050019186	0.285780173	0.094955117
	Average	0.093171505	0.093036104	0.029267022	0.047252186	0.254504532	0.070309555
Ripe	Min	0.106511501	0.062372029	0.007678564	0.024284373	0.268959043	0.05706466
	Max	0.158615319	0.104740726	0.016257225	0.032316819	0.322968701	0.08950532
	Average	0.141970266	0.089772684	0.011890357	0.029272357	0.286386488	0.078499912
Rotten	Min	0.091389662	0.045358516	0.012828138	0.015756491	0.208758942	0.048930205
	Max	0.186203013	0.118977453	0.039978114	0.028506851	0.272719536	0.111595885
	Average	0.130108097	0.073298567	0.024200173	0.021224477	0.256087761	0.073533681

From the test results shown in Table 6, it can be seen that there are two images of test data whose classification results do not match the actual condition of the banana fruit. 2 The test data image is the 5th test data and the 13th test data. In the 5th test data, when the extraction and classification process is carried out in the system, the result is unripe, even though it is an image of banana fruit in a rotten state. Whereas in the 13th test data, when the extraction and classification process is carried out, the result is rotten, when in fact, the image is an image of a banana in a ripe state.

Table 6: Result of testing data

No.	Testing data image	Result of the classification	Information
1		Rotten	True
2		Rotten	True
3		Rotten	True
4		Rotten	True
5		Unripe	False
6		Rotten	True
7		Rotten	True
8		Rotten	True
9		Rotten	True
10		Rotten	True
11		Ripe	True
12		Ripe	True
13		Rotten	False
14		Ripe	True
15		Ripe	True
16		Ripe	True
17		Ripe	True
18		Ripe	True
19		Ripe	True
20		Ripe	True
21		Unripe	True
22		Unripe	True
23		Unripe	True
24		Unripe	True
25		Unripe	True
26		Unripe	True
27		Unripe	True
28		Unripe	True
29		Unripe	True
30		Unripe	True

The fifth banana image is an image of the rotten category, but the system's results are raw. To analyze the classification error location, it is necessary to compare the values in the test data with unripe category training data and rotten category test data. The 24 types of test data values need to be categorized one by one to determine the image class.

Based on testing 24 data values in the 5th test data, six test data types (Variance R, Variance B, Variance S, Variance V, standard deviation V and standard deviation S) whose values can be classified into unripe class. There are five test data types whose values belong to the rotten class, namely Mean G, Mean H, Range H, Standard Deviation G, and Standard Deviation H. The other types of test data whose values can be categorized into unripe or rotten classes, and two types of data. Test data shows Variance H and Variance G showed the value does not belong to the unripe or rotten class.

Table 7: Result of Euclidean distance fo K=5

No.	Training data	Distance to testing data
1	Rotten-9	0,096456715
2	Rotten-7	0,118859741
3	Ripe-1	0,131604174
4	Rotten-8	0,171201307
5	Ripe-8	0,17363607
6	Rotten-6	0,189158539
7	Ripe-7	0,197169181
8	Rotten-2	0,205494293
9	Ripe-2	0,226465903
10	Rotten-3	0,227887828
11	Ripe-4	0,235300172
12	Ripe-5	0,245297991
13	Ripe-6	0,297880646
14	Ripe-9	0,329481775
15	Rotten-5	0,33342593
16	Rotten -4	0,364069974
17	Ripe-10	0,517746128
18	Ripe-3	0,656720342
19	Rotten -10	0,808607818
20	Rotten -1	0,888057115

Based on Table 7, after the distance value between all ripe and rotten category training data is determined, we focused on the top 5 values of the sorting results. The K value specified is 5. In the maximum five values, 3 of them are rotten class, and two others are mature class. Thus, the KNN classification calculation result is that the 13th test data is included in the rotten category because of the five values of the distance between the training data and the closest test data is the rotten class training data.The final step is to calculate the accurate percentage. 2 data are false from 30 data so that the system has an accuracy of 93.3%.

4. CONCLUSION

The proposed system can classify bananas' ripeness by using the GUI feature in the MATLAB 2017a application. The system can classify the muli banana fruit's ripeness into three levels: unripe, ripe, and rotten. The accuracy achieved by the system as a whole is 93.333%. These results were obtained by testing 30 images of bananas, and 2 of them showed incorrect classification results.

REFERENCES

[1] FAO, "Citrus Fruit Statistics 2015," Rome, 2015.

[2] A. Bhargava and A. Bansal, "Fruits and vegetables quality evaluation using computer vision: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, 2018, doi: 10.1016/j.jksuci.2018.06.002.

[3] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation —A review," *Inf. Process. Agric.*, no. xxxx, 2019, doi: 10.1016/j.inpa.2019.09.006.

[4] F. Liantoni and H. Nugroho, "Perbaikan Kontras Citra Dengan Ekualisasi Histogram Dan Gaussian Pada Klasifikasi Semangka," *J. Inform. Upgris*, vol. 5, no. 1, pp. 35–39, 2019, doi: 10.26877/jiu.v5i1.3016.

[5] J. Jhawar, "Orange Sorting by Applying Pattern Recognition on Colour Image," *Phys. Procedia*, vol. 78, no. December 2015, pp. 691–697, 2016, doi: 10.1016/j.procs.2016.02.118.

[6] H. J. G. Opeña and J. P. T. Yusiong, "Automated Tomato Maturity Grading Using ABC-Trained Artificial Neural Networks," *Malaysian J. Comput. Sci.*, vol. 30, no. 1, pp. 12–26, 2017.

[7] A. D. Rosli, N. S. Adenan, H. Hashim, N. E. Abdullah, S. Sulaiman, and R. Baharudin, "Application of Particle Swarm Optimization Algorithm for Optimizing ANN Model in Recognizing Ripeness of Citrus," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 340, no. 1, pp. 0–10, 2018, doi: 10.1088/1757-899X/340/1/012015.

[8] J. Lado, "Maturity indicators and citrus fruit quality," *Stewart Postharvest Rev.*, vol. 2, no. May 2016, 2014.

[9] C. Zhang, W. Yang, Z. Liu, D. Li, Y. Chen, and Z. Li, "Color Image Segmentation in RGB Color Space Based on Color Saliency," in *Computer and Computing Technologies in Agriculture VII*, 2014, pp. 348–357.

[10] M. Kuncan and H. M. Ertunç, "Olive separation machine based on image processing," *Mechanika*, vol. 22, no. 5, pp. 438–443, 2016, doi: 10.5755/j01.mech.22.5.13726.

[11] M. Loesdau, S. Chabrier, and A. Gabillon, "Hue and Saturation in the RGB Color Space," in *Lecture Notes in Computer Science*, 2014, vol. 8509, pp. 203–212, doi: 10.1007/978-3-319-07998-1_23.

[12] D. S. Kalel, P. M. Pisal, and R. P. Bagawade, "A

Study of Color, Shape and Texture Feature Extraction for Content Based Image Retrieval System," *Int. J. Adv. Res. Comput. Commun.*, vol. 5, no. 4, pp. 303–306, 2016, doi: 10.17148/IJARCCE.2016.5477.

[13] D.-W. Sun, *Computer Vision Technology for Food Quality Evaluation*, 2nd ed. Elsevier, 2016.

[14] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. New York: Cambridge University Press, 2014.

[15] A. Danades, D. Pratama, D. Anggraini, and D. Anggriani, "Comparison of accuracy level K-Nearest Neighbor algorithm and support vector machine algorithm in classification water quality status," *Proc. 2016 6th Int. Conf. Syst. Eng. Technol. ICSET 2016*, no. December 2016, pp. 137–141, 2017, doi: 10.1109/FIT.2016.7857553.

[16] M. Vassallo-Barco, L. Vives-Garnique, V. Tuesta-Monteza, H. I. Mejía-Cabrera, and R. Y. Toledo, "Automatic Detection of Nutritional Deficiencies in Coffee Tree Leaves Through Shape and Texture Descriptors," *J. Digit. Inf. Manag.*, vol. 15, no. 1, pp. 7–18, 2017.

[17] N. Manohar, Y. H. S. Kumar, and G. H. Kumar, "An Approach for the Development of Animal Tracking System," *Int. J. Comput. Vis. Image Process.*, vol. 8, no. 1, pp. 15–31, 2018, doi: 10.4018/ijcvip.2018010102.

[18] B. Jiang *et al.*, "Fusion of machine vision technology and AlexNet-CNNs deep learning network for the detection of postharvest apple pesticide residues," *Artif. Intell. Agric.*, vol. 1, pp. 1–8, 2019, doi: 10.1016/j.aiia.2019.02.001.

[19] S. Sabzi and Y. Abbaspour-gilandeh, "A New Approach for Visual Identification of Orange Varieties Using Neural Networks and Metaheuristic Algorithms," *Inf. Process. Agric.*, vol. 5, pp. 162–172, 2018, doi: 10.1016/j.inpa.2017.09.002.