



## Classification of Fruits Based on Shape and Color using Combined Nearest Mean Classifiers

Abdullah<sup>1</sup>, Agus Harjoko<sup>2</sup>, Othman Mahmud<sup>3</sup>

<sup>1</sup>Department of Information System, Universitas Islam Indragiri Indonesia

<sup>2</sup>Department of Computer Science and Electronics, Universitas Gadjah Mada Indonesia

<sup>3</sup>Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, Malaysia

<sup>1</sup>abdialam@gmail.com, <sup>2</sup>aharjoko@ugm.ac.id, <sup>3</sup>mahmod.othman@edu.my

### Abstract

*Fruit classification is an important task in many agriculture industry. The fruit classification system can be used to identify the types and prices of fruit. Manual classification of fruit is not efficient for large amount of fruits. The advancement of information technology has made possible fruit classification be done by a machine. This research aims to propose a fruit classification methodology based on shape and color. To reduce the effect of lighting variability a color normalization is carried out prior to feature extraction. The color features used in this research are mean and standard deviation. The shape features are area, perimeter, and compactness. The classification of an unknown fruit is carried out using the nearest mean classifier. The method developed in this research is tested using 12 classes of fruits where each class is represented by a number of samples. The experimental results show that the method proposed in this research provides an accuracy of 95.83% for two samples per class and 100% for three samples per class. Experiment on small training samples has been conducted to evaluate the performance of the proposed combined nearest mean classifiers and results obtained showed that the technique was able to provide good accuracy.*

*Keywords: fruit classification, nearest mean classifier, color features, shape features*

### 1. Introduction

Fruit classification which is done manually is inefficient and inaccurate for large amount of fruits. The information technology advancement had made possible the fruit classification be done by a machine or computer. Fruit classification is meaningful to various fields i.e. industry, plantation, farming, trading and so forth [1].

Combining multiple classifier or ensemble method is considered as a general solution for pattern classification tasks. Multiple classifier combination goals to achieve the final decision by integrating the predictions of several individual classifiers to obtain comprehensive results. Experimental studies have shown that the combination of several classifiers has been very helpful in improving the classification accuracy [2],[3]

There are several ensemble and deep learning methods for fruit classification have been proposed [4]-[7]. However, deep learning in machine learning uses large amounts of data. Many deep learning algorithms use multiple layers of neural networks making

computations more complex. In this paper, a new simple method of combining multiple classifier namely combined nearest mean classifiers for fruit classification is proposed. In this method, only a few fruit samples are needed in the training process.

This research treated the fruit classification based on shape and color similarity. The classification is implemented on per fruit item, not the group of fruit. The different feature used is the color feature (mean and standard deviation of color) and shape feature (area, perimeter, and compactness). This research result will be useful in accelerating the sorting and grading process for fruit variants and make easy the fruit trading cost decision based on shape (e.g. big or small) and color (e.g. red or green).

### 2. Research Methods

The proposed multiple nearest mean classifier (NMC) for fruit classification consists of three phases namely image preprocessing, feature extraction and classifier combination as depicted in Figure 1.

The image of the fruit will be preprocessed to obtain the feature of the fruit. Operations such as background subtraction and the normalization of color will be performed on the image of the fruit. Background subtraction is performed to separate the image of the fruit from its background. Color normalization operation is then performed to eliminate the influence of different lighting.

Features in the image of the fruit are extracted and placed in feature vectors. The color features are measured by mean and standard deviation on each red, green, and blue (RGB) channel. The shape features are measured by area, perimeter and compactness. The area of fruit reflects the actual fruit size or weight. The perimeter of fruit is defined as the area that covers the boundary. The compactness of fruit is defined as the ratio of the area of a fruit to the area of a circle with the same perimeter.

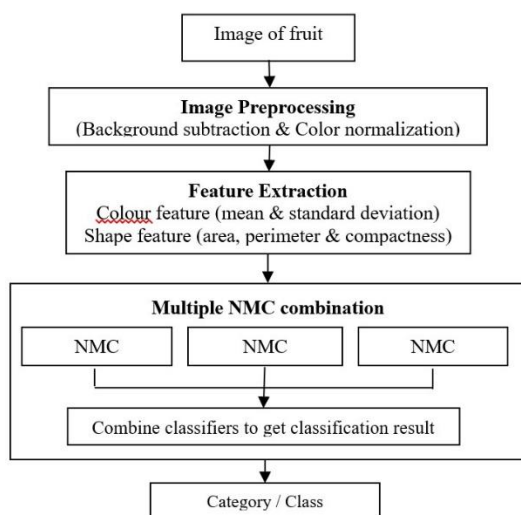


Figure 1. Phases of the proposed methodology

The multiple NMC combination consists of three nearest mean classifiers [8],[9]. The input features to the first and second classifiers are the color mean and color standard deviation respectively while the input features to the third classifier are the area, perimeter, and compactness. Output from each classifier is the similarity value between the unknown object and samples (or training patterns). The similarity value is obtained by calculating the euclidean distance between the feature vector of the unknown object and the feature vector of the sample class mean.

A sample of 84 fruit images that correspond to 12 categories has been used to form the reference values for each category. The data were divided into a training set (43%) and a testing set (57%), with one to three training samples, were used. All images were 640 x 480 pixels with 24-bit true color, 256 levels of gray and, an RGB color model. The types of fruits that were used are limited to variants of apples, mangoes, oranges, pears, and durian.

According to Sen [10],[11] that the classification used as mentioned above is named supervised classification, since the class has been noticed and the data sample has been available. To develop the supervised classification, earlier a computer system must have the knowledge that can be developed by learning the sample and recording them in a database [12].

The fruit classification system follows the structure of an introduction design system proposed by Yan and Gao [13] that includes sensor, processing feature extraction and classifier algorithm. The classification of fruits is done indirectly by capturing the fruit object's image using the sensor. The object's image that is identical with its feature as well the reality is in the same class [14].

The sensor used as the image capture in this system is a digital camera (or webcam), as shown in Figure 2.

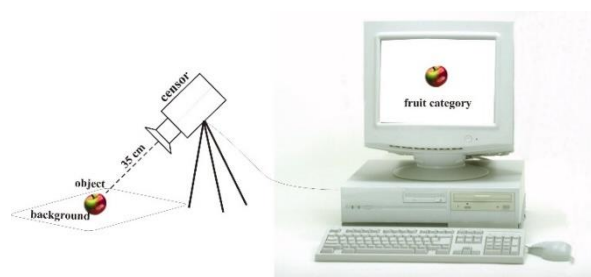


Figure 2. the hardware scheme of fruit image capture

All the images are captured in a 640 x 480 size. The images are saved as 24 true color images having 256 grayish levels with RGB color model and bitmap format file (BMP). Another supporting tool is a tripod to help capture images at the same distance of 35 cm. The background color is made same for each image of the fruit object, which is black to avoid the shadow of the fruit object. The lighting intensity is fixed and not too bright to avoid reflection effects on glossy surfaces which can cause a loss of color information.

According to [15][16] there are several mechanisms used in computer vision, one of them is the statistical design which uses two phases i.e. training phase and the testing phase's approach. Generally, the process in the classification of fruits consists of the main process i.e. class formation process (training phase) and fruit classification process (recognition phase).

The system can be arranged from sub-systems [17]. The classification of the fruit system is arranged into two sub-systems which are the class formation system (SPK), which undertakes the training process, and the class formation and fruit classification system (SKB), which undertakes the unknown classification of the fruit process into a certain class.

According to Meshram, et. al. [18], the classification of the fruit architecture system can be seen in the following Figure 3.

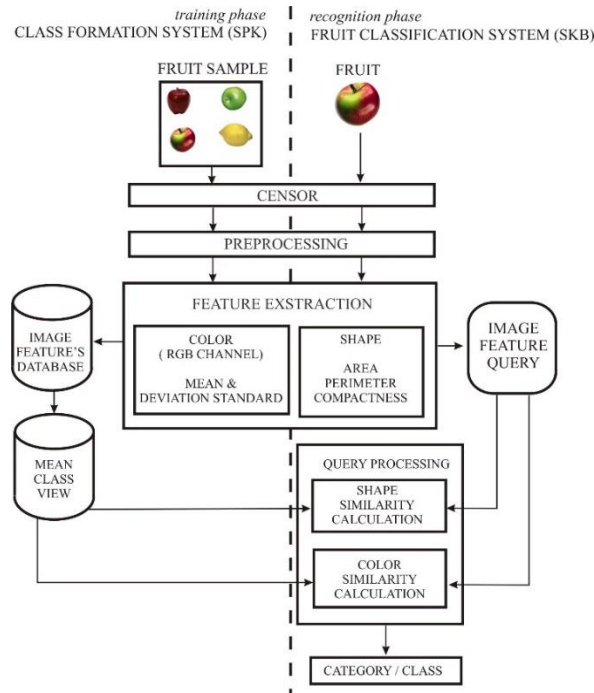


Figure 3. The classification of fruit architecture system

### 2.1. Training phase

In the training phase, the image of the fruit sample is captured through a censor. The fruit sample used consists of several fruit samples for each category/sample afterward the processing is done.

The first process is background subtraction to separate the fruit from its background by implementing pixel subtraction operation. The result is the absolute subtraction value of the fruit image with its background [19]. The pixel subtraction operation result is achieved from equation (1).

$$Q(i, j) = |P_1(i, j) - P_2(i, j)| \quad (1)$$

Where  $P_1$  is the fruit image and  $P_2$  is the background image. The intensity of R, G, and B in the pixel image is the background subtraction result in which the value is lesser than the threshold value which is 75 that is considered as the background. This value is quite ideal based on several trials that had been done.

The second process is color normalization used to disappear the influence of the different lighting [20]. The color normalization used equations (2), (3), and (4) since they are appropriate to the color feature measured in each RGB channel. The used equation in doing the color normalization in each pixel  $p$  is:

$$r(p) = \frac{R(p)}{R(p)+G(p)+B(p)} \quad (2)$$

$$g(p) = \frac{G(p)}{R(p)+G(p)+B(p)} \quad (3)$$

$$b(p) = \frac{B(p)}{R(p)+G(p)+B(p)} \quad (4)$$

with each  $R(P)$ ,  $G(P)$ , and  $B(P)$  as the color intensity in each component of R (red), G (green), and B (blue) in pixel  $p$ .

The color feature of the fruit image result's processing is extracted. The color feature can consist of statistical data based on the color histogram [21]. The equations (5), (6) and (7) are used to calculate the color mean, then equations (8), (9) and (10) are used to calculate the standard deviation of colors. If the fruit image is  $x$  and the pixel number is  $p$ , the mean color of the fruit image is:

$$x = (r_{avg}, g_{avg}, b_{avg})^T \text{ where}$$

$$r_{avg} = \frac{\sum_{p=1}^P r(p)}{P} \quad (5)$$

$$g_{avg} = \frac{\sum_{p=1}^P g(p)}{P} \quad (6)$$

$$b_{avg} = \frac{\sum_{p=1}^P b(p)}{P} \quad (7)$$

If the fruit image is  $x$  and the pixel number is  $P$ , the deviation standard of the fruit image color is:

$$x = (r_{std}, g_{std}, b_{std})^T \text{ where}$$

$$r_{std} = \sqrt{\frac{\sum_{p=1}^P (r(p) - r_{avg})^2}{P}} \quad (8)$$

$$g_{std} = \sqrt{\frac{\sum_{p=1}^P (g(p) - g_{avg})^2}{P}} \quad (9)$$

$$b_{std} = \sqrt{\frac{\sum_{p=1}^P (b(p) - b_{avg})^2}{P}} \quad (10)$$

Fruit image shape features are also extracted. Fruit image shape features are measured by area, circumference, and compactness. The fruit area reflects the size and weight of the real fruit. The perimeter is the outside part of the object that is side by side with the pixel background. Perimeter is obtained by calculating the number of pixels in the fruit image borders. If  $x_1, x_2, \dots, x_n$  border, the perimeter is measured through the equation (11).

$$perimeter = \sum_{i=1}^{n-1} d_i = \sum_{i=1}^{n-1} |x_i - x_{i+1}| \quad (11)$$

the distance  $d_i$  is equal to 1 to the border with pixel 4 line and 1 or  $\sqrt{2}$  to the border with pixel 8 line, this research perimeter definition took the pixel 8 line in the outer part of the object and  $d_i$  is equal to 1. compactness is the fruit shape compact that is measured through the equation (12) :

$$c = \frac{4 \cdot \pi \cdot a}{p^2} \quad (12)$$

The fruit image shape  $x$  is represented by  $x = (a, p, c)^T$  where  $a$  is the area,  $p$  is the perimeter, and  $c$  is compactness.

The shape and color extraction result of the fruit is kept in the image feature database, SPK creates class mean for each fruit. The class mean or centroid is measured by using the equation (13).

$$\bar{x} = \frac{1}{n_i} \sum_{j=1, n_i} x_{i,j} \quad (13)$$

where  $x_{i,j}$  is the  $j^{\text{th}}$  sample feature vector from class  $i$

or create a view from the feature mean measurement in the image feature database that is classified based on the category or class label.

## 2.2. Recognition Phase

In the recognition phase, the fruits are classified and the images will be captured through the sensor, implemented preprocessing and feature extraction as in the training phase.

The extraction result of fruit shape and color is used in the image feature query. The classification is done by measuring the shape and color similarity of the image query with the mean class as equation (13). The unknown fruit is stated as feature vector  $q$  that will be classified to class  $i$  if it is closer to vector mean class  $i$  than others.

The similarity is measured through the vector distance. Two closest vectors will possess similarities and a little bit of difference [14]. Generally, the NMC classifier used euclidean distance [22], even according to [23], NMC classifier is also well-known as Nearest Centroid Classifier [24]. Furthermore, the used distance metric in fruit classification is the L2 metric (euclidean metric). According to Malkov [25] that the euclidean of two vectors  $x$  and  $w$  is shown by the equation (14).

$$\|x - y\|_2 = \sum_{i=1}^n (x_i - y_i)^2 \quad (14)$$

The similarity measurement is done per groups' feature that can fasten and make simple the query process [26]. There are three features which are the color (mean of R, G, and B), colors' deviation standard (R, G, and B of colors's deviation standard) and, shape (area, perimeter, and compactness).

According to [27][28], the euclidean distance of two image color vectors mean can be measured. If the mean of the color fruit image query is stated as vector  $q$  and the mean class vector as  $x$ , the euclidean distance of two vectors is is shown by the equation (15).

$$d_{avg}(q, \bar{x}) = \sqrt{(r_{avgq} - r_{avg\bar{x}})^2 + (g_{avgq} - g_{avg\bar{x}})^2 + (b_{avgq} - b_{avg\bar{x}})^2} \quad (15)$$

implementing the same treatment, if the deviation standard of fruit color image query is stated as vector  $q$  and mean color in the mean class is vector  $x$ , the euclidean distance of two vectors is is shown by the equation (16).

$$d_{stdev}(q, \bar{x}) = \sqrt{(r_{stdevq} - r_{stdev\bar{x}})^2 + (g_{stdevq} - g_{stdev\bar{x}})^2 + (b_{stdevq} - b_{stdev\bar{x}})^2} \quad (16)$$

afterward, if the fruit image query shape is stated as vector  $q$  and shape in mean class is vector  $x$ , the euclidean distance of those two vectors is is shown by the equation (17).

$$d_{bentuk}(q, \bar{x}) = \sqrt{(a_q - a_{\bar{x}})^2 + (p_q - p_{\bar{x}})^2 + (c_q - c_{\bar{x}})^2} \quad (17)$$

After the above similarity measurement on each feature group, the similarity measurement is done simultaneously to those three feature groups. The way is by adding up those three groups' distances. However, the distance scale in each feature is different, it is normalized by subtracting each distance of a certain feature with the maximum distance. The normalized distance of each group is around 0-1, so the distance similarity total is around 0-3. The equation used to measure the similarity distance is the equation (18).

$$d_{sim}(q, \bar{x}) = \frac{d_{avg}(q, \bar{x})}{\max d_{avg}} + \frac{d_{stdev}(q, \bar{x})}{\max d_{stdev}} + \frac{d_{bentuk}(q, \bar{x})}{\max d_{bentuk}} \quad (18)$$

where  $d_{sim}(q, \bar{x})$  is the similarity distance,

$\max d_{mean}$  is the maximum distance of the mean color.

$\max d_{stdev}$  is the maximum distance of standard deviation color.

$\max d_{bentuk}$  is the maximum distance in shape's category.

The similarity distance measurement is done for all mean classes. The classification rule according to [14] is given to two classes  $w_1$  and  $w_2$ . The object's vectors are written as  $\{x_1, \dots, x_n\}$ , if  $x_1$  is the  $w_1$  mean class, the new object  $Z$  is represented in the space as  $Z_{x_1}$ .

- Classify  $z$  to  $w_1$  if and if only  $d^2(z, \bar{x}_{(1)}) < d^2(z, \bar{x}_{(2)})$

- Classify  $z$  to  $w_2$  if and if only  $d^2(z, \bar{x}_{(2)}) < d^2(z, \bar{x}_{(1)})$

The fruit is classified if the minimum value is  $d_{sim}(q, \bar{x}) < \epsilon$  with  $\epsilon = 0.75$ , vise versa the fruit is rejected, if the threshold value is 0.75 (scale 0-3) is achieved empirically. The similarity percentage is relative toward the distance is 100% if the distance is 0. to be classified, the minimum similarity percentage of fruit is 75% or the similarity distance is 0.75.

### 3. Results and Discussions

There are 12 fruit categories that are made as class labels i.e. Fuji Apple, Manalagi Apple, Washington Apple, Arum Manis A Mango, Arum Manis B Mango, Siam Orange, Honey Mango, Podang Mango, Sunkist Orange, Siam Orange, Pear, and Durian. Each class is represented by some of the fruits' samples. Obtained feature values a shown in Table 1.

Table 1. Reference Fruit Images Feature Values

Type of fruit	Color mean			Color standard deviation			Shape		
	Red	Green	Blue	Red	Green	Blue	Area	Perimeter	Compactness
Fuji Apple	193.93	136.28	83.73	38.71	50.96	30.03	15911	438	1.04
Manalagi Apple	167.41	180.51	66.67	33.88	34.28	25.70	12581	385	1.07
Washington Apple	184.34	75.75	64.24	37.20	39.23	30.05	16455	623	0.53
Arum Manis A Mango	131.89	143.69	51.07	27.20	39.23	30.05	31211	1449	0.19
Arum Manis B Mango	112.97	132.85	47.52	17.84	20.79	16.66	17083	920	0.25
Golek Mango	147.24	150.16	36.51	31.02	28.58	27.73	24739	1042	0.32
Honey Mango	105.27	138.55	77.66	22.91	24.81	27.03	20703	1588	0.10
Podang Mango	203.74	143.30	48.70	34.63	34.37	30.31	16436	455	1.00
Sunkist Orange	206.37	114.51	8.17	38.68	37.96	26.97	20846	618	0.69
Siam Orange	176.46	135.07	20.09	36.46	34.53	33.22	14469	498	0.73
Pear	211.77	191.14	122.68	35.30	41.74	41.80	18324	515	0.87
Durian	117.28	123.81	50.23	19.27	20.72	25.17	72276	10769	0.01

The class formation system's test is done used 3 ways: firstly, treated one fruit sample; secondly, treated two samples and thirdly, treated three samples of fruit sample for each class. The sample image is taken to look for its characteristic which is called as training image.

The class formation system forms the mean class of each class categorical label as shown in Figure 4 which is the class formation system's user face that shows the mean class formation to three samples in each class.

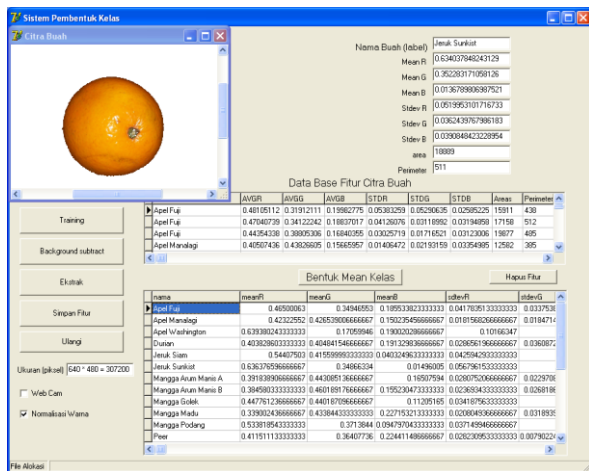


Figure 4. SPK's user interface

The fruit classification system's test is done by doing the image query. This image is called image testing used to test the system's success. The feature of the fruit classification system's user is viewed in Figure 5.

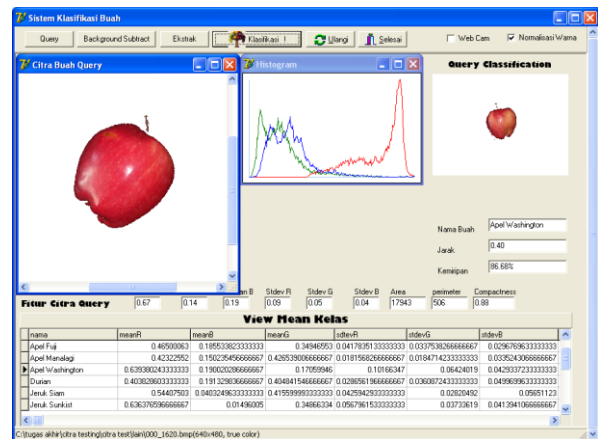


Figure 5. SKB's user interface

In evaluating the performance of the proposed method, 48 fruit images from 12 classes were used as the testing images. The first test used the image of the same fruits, however, the positions are being changed. The second test used the image of different fruits that were never used in the training. The first test result is presented in Table 2 and the second test is in Table 3.

Table1 shows that the system is able in recognizing and classifying used for training although the position is different. Table 2, shows that the system is also able in recognizing and classifying the fruit from the different samples. Generally, the successfulness level using 2 samples in each class is 95.83. If using 3 samples, the successfulness level reaches 100%. It means, this system has a good reputation.



Table 2. The Same Fruits Images Test Result

No	The image training number	The image testing number	Classification			Result %
			True	False	Rejected	
1	One sample per class	12	11	1	0	91,67
		24	23	1	0	95,83
		36	34	2	0	94,44
		48	46	2	0	95,83
2	Two samples per class	12	11	1	0	91,67
		24	23	1	0	95,83
		36	35	1	0	97,22
		48	46	1	1	95,83
3	Three samples per class	12	12	0	0	100
		24	24	0	0	100
		36	36	0	0	100
		48	48	0	0	100

Table 3. The Other Fruit Image Test Result.

No	The image training number	The image testing number	Classification			Result %
			True	False	Rejected	
1	One sample per class	12	12	0	0	100
		24	24	0	0	100
		36	36	0	0	100
		48	46	1	1	95,83
2	Two samples per class	12	12	0	0	100
		24	24	0	0	100
		36	36	0	0	100
		48	45	2	0	95,83
3	Three samples per class	12	12	0	0	100
		24	24	0	0	100
		36	36	0	0	100
		48	48	0	0	100

#### 4. Conclusion

Classification of fruits using the proposed multiple nearest mean classifier technique has shown that the technique is capable in producing high accuracy with a small sample size. The sample number in each class influences the system's ability, the system becomes better with the sample advancement's number. Up to 3 samples in each class, the system had been able in doing the classification to 48 fruits with 100% in its successfulness level or having a good reputation. The image capture process needs to be taken into account, so the color and shape of the fruit can be represented well. This way can be applied by using supplemented light and solid-state image sensor. In fact, the fruit surfaces are not always stainless, sometimes possessing stains and dust in which their colors are identical to the background thus the background subtraction result is not perfect. It needs an algorithm arrangement and image processing technique to figure out the weaknesses. It needs the fruit classification system's hardware so that the fruit sorting and grading process can be done by a machine or robotic system.

#### References

- [1] H. A. Hambali, S. L. S. Abdullah, N. Jamil, and H. Harun, "Fruit classification using neural network model," *J. Telecommun. Electron. Comput. Eng.*, vol. 9, no. 1–2, pp. 43–46, 2017.
- [2] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4. 2018, doi: 10.1002/widm.1249.
- [3] M. Astani, M. Hasheminejad, and M. Vaghefi, "A diverse ensemble classifier for tomato disease recognition," *Comput. Electron. Agric.*, vol. 198, p. 107054, 2022, doi: <https://doi.org/10.1016/j.compag.2022.107054>.
- [4] J. Kang and J. Gwak, "Ensemble of multi-task deep convolutional neural networks using transfer learning for fruit freshness classification," *Multimed. Tools Appl.*, vol. 81, Jul. 2022, doi: 10.1007/s11042-021-11282-4.
- [5] H. S. Gill, O. I. Khalaf, Y. Alotaibi, S. Alghamdi, and F. Alassery, "Fruit Image Classification Using Deep Learning," *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 5135–5150, 2022, doi: 10.32604/cmc.2022.022809.
- [6] C. C. Ukwuoma, Q. Zhiguang, M. B. Bin Heyat, L. Ali, Z. Almaspoor, and H. N. Monday, "Recent Advancements in Fruit Detection and Classification Using Deep Learning Techniques," *Math. Probl. Eng.*, vol. 2022, p. 9210947, 2022, doi: 10.1155/2022/9210947.
- [7] P. Sumari *et al.*, "A Precision Agricultural Application: Manggis Fruit Classification Using Hybrid Deep Learning," *Rev. d'Intelligence Artif.*, vol. 35, no. 5, pp. 375–381, 2021, doi: 10.18280/ria.350503.
- [8] Z. Mai, R. Li, H. Kim, and S. Sanner, "Supervised Contrastive Replay: Revisiting the Nearest Class Mean Classifier in Online Class-Incremental Continual Learning," *CoRR*, vol. abs/2103.13885, 2021, [Online]. Available: <https://arxiv.org/abs/2103.13885>.
- [9] E. Santucci, "Quantum Minimum Distance Classifier," *Entropy*, vol. 19, no. 659, pp. 1–14, 2017, doi: 10.3390/e19120659.
- [10] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised Classification Algorithms in Machine Learning: A Survey and Review BT - Emerging Technology in Modelling and Graphics," 2020, pp. 99–111.
- [11] K. D. Copsey, *Statistical Pattern Recognition*. Wiley, 2011.
- [12] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.
- [13] X. Yan and L. Gao, "A feature extraction and classification algorithm based on improved sparse auto-encoder for round steel surface defects," *Math. Biosci. Eng.*, vol. 17, no. 5, pp. 5369–5394, 2020, doi: 10.3934/MBE.2020290.
- [14] I. Reppa, K. E. Williams, W. J. Greville, and J. Saunders, "The relative contribution of shape and colour to object memory," *Mem. Cognit.*, vol. 48, no. 8, pp. 1504–1521, 2020, doi: 10.3758/s13421-020-01058-w.
- [15] M. H. Guo *et al.*, "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, 2022, doi: 10.1007/s41095-022-0271-y.
- [16] M. Han, H. Wu, Z. Chen, M. Li, and X. Zhang, "A survey of multi-label classification based on supervised and semi-supervised learning," *Int. J. Mach. Learn. Cybern.*, 2022, doi: 10.1007/s13042-022-01658-9.
- [17] K. . Kendall and J. . Kendall, *System Analysis and Design*, 8th ed. New Jersey: Prentice-Hall, 2020.
- [18] V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S. D. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," *Artif. Intell. Life Sci.*, vol. 1, no. September, pp. 1–11, 2021, doi: 10.1016/j.ailesci.2021.100010.
- [19] D. N. Arulnathan, B. C. W. Koay, W. K. Lai, T. K. Ong, and L. L. Lim, "Background Subtraction for Accurate Palm Oil Fruitlet Ripeness Detection," in *2022 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, 2022, pp. 48–53, doi: 10.1109/I2CACIS54679.2022.9815275.
- [20] N. Phuangsaichai, J. Jakmunee, and S. Kittiwachana, "Investigation into the predictive performance of colorimetric sensor strips using RGB, CMYK, HSV, and CIELAB coupled with various data preprocessing methods: a case study on an analysis of water quality parameters," *J. Anal. Sci. Technol.*, vol. 12, no. 1, 2021, doi: 10.1186/s40543-021-00271-9.
- [21] T. Gevers and A. Smeulders, "Foreword," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes*

- Bioinformatics*), vol. 9909 LNCS, p. V, 2016, doi: 10.1007/978-3-319-46493-0.
- [22] N. Ali, D. Neagu, and P. Trundle, "Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets," *SN Appl. Sci.*, vol. 1, no. 12, pp. 1–15, 2019, doi: 10.1007/s42452-019-1356-9.
- [23] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 2019, pp. 1255–1260, doi: 10.1109/ICCS45141.2019.9065747.
- [24] S. Johri *et al.*, "Nearest centroid classification on a trapped ion quantum computer," *npj Quantum Inf.*, vol. 7, no. 1, 2021, doi: 10.1038/s41534-021-00456-5.
- [25] Y. A. Malkov, "Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 31–33, 2018, [Online]. Available: [https://github.com/nmslib/nmslib%0Ahttp://ann-benchmarks.com/hnsw\(nmslib\).html](https://github.com/nmslib/nmslib%0Ahttp://ann-benchmarks.com/hnsw(nmslib).html).
- [26] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, no. 52, 2020, doi: 10.1186/s40537-020-00327-4.
- [27] J. Lv and J. Fang, "A Color Distance Model Based on Visual Recognition," *Math. Probl. Eng.*, vol. 2018, 2018, doi: 10.1155/2018/4652526.
- [28] R. Manisha, "Content Based Image Retrieval using Color and Texture Feature with Distance Matrices," *Int. J. Sci. Res. Publ.*, vol. 7, no. 8, pp. 512–523, 2012, doi: 10.5121/sipij.2012.3104.