IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 1, March 2024, pp. 319~328
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i1.pp319-328

Comparison of feature extraction and auto-preprocessing for chili pepper *(Capsicum Frutescens)* quality classification using machine learning

Jelita Asian[1], Nunik Destria Arianti[2], Ariefin[3], Muhamad Muslih[2]
[1]Department of Informatic Magister, Nusa Putra University, Sukabumi, Indonesia
[2]Department of Information System, Nusa Putra University, Sukabumi, Indonesia
[3]Department of Mechanical Engineering, Lhokseumawe State Polytechnic, Lhokseumawe, Indonesia

Article Info

Article history:

Received Mar 26, 2023
Revised May 26, 2023
Accepted Jun 3, 2023

Keywords:

Auto-preprocessing
Color feature extractor
Machine vision
ORB feature extractor
Supervised image processing

ABSTRACT

The low-cost camera for machine vision, such as a webcam, still has a problem with resolution noise. Therefore, it is important to learn strategies to reduce noise from low-cost camera images so that they can be widely used for grading machines in the future. This paper aims to compare three feature extraction methods with auto-preprocessing to classify chili pepper *(Capsicum Frutescens)* quality using a machine learning algorithm. Three extraction methods were used, including the color feature, oriented FAST and rotated BRIEF (ORB), and the combination color feature and ORB. A total of 525 image data for quality chili pepper were collected using the webcam. The auto-preprocessing strategy to classify chili peppers can improve the performance of machine-learning algorithms for all data generated by the feature extractor. The performance of the chili paper quality classification model with auto-preprocessing of the variable color feature can improve the performance of machine learning algorithms by up to 64.21%. The performance improvement of the classification model using the ORB feature variable and the auto-preprocessing of up to 4.41%. The performance improvement of the classification model using machine learning algorithms is 11.27% when using the combination color feature and ORB feature and auto-preprocessing.

Corresponding Author:

Jelita Asian
Department of Informatic Magister, Nusa Putra University
St. Cibolang 21, Sukabumi, Indonesia
Email: jelita.asian@nusaputra.ac.id

1. INTRODUCTION

Grading in the agricultural production process is needed to separate products based on quality. Generally, grading quality indicators are shape, size, color, maturity, and defection [1]–[4]. Grading can be done manually, but it is prone to inconsistencies and requires more time. An automatic grading process was developed using computer vision and machine learning to overcome this problem. In addition to speeding up the process, automatic grading with computers allows for further fruit processing utilizing an automation system. But until now, devices to support it so it can have high performance are still not affordable, especially for sensors such as cameras.

The camera is the most important part of grading machine development [5]–[7]. This part first acquire data before proceeding with processing. On the one hand, many cameras have been developed for computer vision purposes, especially for machine grading. However, most cameras are still expensive and unaffordable if they want to be used for creating a grading machine on a small industrial scale, such as for fruit and vegetable horticultural products. On the other hand, low-cost cameras (such as web cameras) can be used for machine grading but provide low-quality images. This will impact the performance of the developed grading machine if it is to be used directly. Therefore, creating a classification modeling strategy that uses source images from low-cost cameras is necessary, especially machine learning [8]–[11].

Several research results have studied the use of low-cost cameras to be used in machine grading combined with machine learning, as reported by Mahendra *et al.* [12] for strawberry products, Tuesta *et al*. [13] for citrus aurantifolia products and Nasution and Amrullah [14] for apple products. In general, the strategy proposed in this research is to prioritize machine learning methods in its modeling. In addition, they also studied various feature extraction techniques from low-quality images to obtain a robust model.

To the best of our knowledge, so far no one has studied the use of low-cost cameras to be used in the development of grading machines with all their limitations, especially chili paper. Therefore, this article proposes an additional process, namely auto-preprocessing, to improve the quality of the extracted features when modeling using machine learning algorithms. The proposed preprocessing is a strategy developed to be used automatically or is called auto-preprocessing. This study aims to compare the performance of machine learning algorithms to features extracted with and without auto-preprocessing. The feature extractor used in this study includes a color feature, feature oriented FAST and rotated BRIEF (ORB) and a combination of color and feature ORB. Also, three machine learning algorithms are used, including support vector machine (SVM), decision tree (DT), and random forest (RF).

2. METHOD

2.1. Data collection and extraction feature

A total of 525 chili paper was used in this study, consisting of 420 good quality chili paper and 105 degraded chili pepper. Their image data were acquired using a low-cost web camera (Logitech C170). Figure 1 presents an example of chili paper damage as shown in Figure 1(a) and good as shown in Figure 1(b) samples used in this study.


(a) (b)

Figure 1. Sample chili pepper with (a) damaged quality and (b) good quality

After the image data is obtained, the features of each image are extracted using three methods, including the color feature [15]–[17], oriented FAST and rotated BRIEF (ORB) feature [18]–[20] and the combination color feature and ORB feature [21]–[23]. A total of 30 variables were extracted using the color feature method. This feature consists of 10 image color identities (grey, blue, green, red, hue, saturation, value, lightness, red/green value, and blue/yellow value), ten standard deviation values for each color identity image, and ten pieces of the minimum value of each color identity for each image. A total of 200 variables extracted using the ORB extraction method were obtained. This feature consists of 20 different values for each color identity (grey, blue, green, red, hue, saturation, value, lightness, red/green value, and blue/yellow value) for each color identity for each image. The combination color feature and ORB feature uses 230 variables extracted from the two single techniques.

2.2. Auto-preprocessing of data

The auto-preprocessing used in this study is designed to consist of 4 handling groups that can run parallel (baseline handling, scatter handling, noise handling, and scaling-transformations handling) [24]. However, by considering the "None" code in each handling group as an option without preprocessing, other combinations such as single, double, triple, and quarter combination techniques from the group also have the same opportunity to be calculated in this method. In total, 600 preprocessing combinations ($5\times6\times5\times4$) will be iterated in this auto-preprocessing strategy as shown in Table 1. In this study, a Python-based programming code is developed that can run this automatically. The best preprocessing combination was born from an evaluation using a a confusion matrix using 5-fold cross-validation of each algorithm used in this experiment (SVM, DT, and RF). The results of the highest accuracy value from the cross-validation test are then sorted and automatically selected for further use as a variable to predict the quality of chili paper. If a figure contains subfigures, please mention EACH subfigures in the body-text after mentioning the main figure.

Table 1. Group and single preprocessing for auto-preprocessing strategy

Baseline handling	Scatter handling	Noise handling	Scaling-transformations handling
None	None	None	None
Detrending (2^{nd} order polynomial)	Mean scaling	Savitzky–Golay (SG) (Window: 3pt, order: 1)	Mean centering
Detrending (3^{rd} order polynomial)	Median scaling	Savitzky–Golay (SG) (Window: 5 pt, order: 2)	Auto-scaling
Detrending (4^{th} order polynomial)	Maximum scaling	Savitzky–Golay (SG) (Window: 9 pt, order: 2)	Range scaling
Arterial spin labeling (AsLS)	Standard Normal Variate (SNV)	Savitzky–Golay (SG) (Window: 11 pt, order: 2)	
	Multiplicative scatter correction (MSC)		

2.3. Machine learning algorithms

The machine learning algorithms used in this study include support vector machines (SVM), decision trees (DT), and random forests (RF). SVM is a linear classifier in a high-dimensional space. It expresses the decision boundary in terms of a linear combination of functions parametrized by support vectors, a subset of training points [25]. Decision tree analysis is a divide-and-conquer approach to classification. Decision trees can be used to discover features and extract patterns from large databases important for discrimination and predictive modeling [26]. RF is an algorithm that combines Classification and regression trees (CART) and bootstrapping aggregation (bagging) algorithms. RF applications to solve classification problems of large data and a larger sample set can generate a high computational cost in addition to requiring more time [27]. The detailed descriptions of the SVM, DT, and RF algorithms are described elsewhere and therefore not shown here. All machine learning algorithms were developed using the open source and free Scikit-learn machine learning package for Python 3.8.8, built on the scientific and numerical Python libraries SciPy and NumPy [28].

2.4. Evaluation of the classified model

The model classification was evaluated using the confusion matrix method, consisting of two chili papper quality classes: damage quality and good quality. Using the random splitting method, the total dataset will be divided by a ratio of 80/20 for training and independent data set testing. True positives, true negatives, false positives, and false negatives will be represented by true positive (TP), true negative (TN), false positive (FP), and false negative (FN), respectively. The classification algorithms used in Python 3.8.8 using a popular machine learning package from Scikit-learn. The observed parameters were precision (PRE), recall (REC), F1-score (F1S), and accuracy (ACR), each of which was calculated using (1) to (4) [29].

$$PRE = \frac{TP}{TP+FP} \tag{1}$$

$$REC = \frac{TP}{TP+FN} \tag{2}$$

$$F1S = \frac{2\times PRE\times REC}{PRE+REC} \tag{3}$$

$$ACR = \frac{TP+FN}{TP+FN+TN+FP} \tag{4}$$

3. RESULTS AND DISCUSSION

3.1. Classification model using color feature extraction

The raw data using a color feature extractor is shown in Figure 2. A total of 30 variables can be extracted from 420 image data for good chili pepper and 105 data for damaged chili pepper. In general, the features of all samples appear very similar and tricky to distinguish with the naked eye. Therefore, the use of image processing methods supported by appropriate algorithms needs to be used to be able to classify them.

The best auto-preprocessing for classifying the SVM, DT, and RF algorithms using color feature variables are presented in Figure 3. In the SVM algorithm as shown in Figure 3(a), the recommended auto-preprocessing is from the baseline handling and scaling-transformation handling preprocessing group, based on the 5-smallest fold cross-validation error. In particular, the combination of detrending preprocessing (2nd order polynomial) and range scaling is the best, giving an accuracy cross-validation of 0.898. The auto-preprocessing suggested for the DT algorithm from the color feature extractor is from the preprocessing baseline handling and scatter handling groups whose variables are presented in Figure 3(b). Combining the single preprocessing technique is detrending (4th order polynomial) and SNV with an accuracy cross-validation of 0.809. The preprocessing baseline handling group represented by detrending (3rd order polynomial) is the only best result of selecting auto-preprocessing with the RF algorithm with an accuracy cross-validation of 0.898 as shown in Figure 3(c). The auto-preprocessing evaluated by SVM and RF is equally good at providing the performance of the chili pepper classification model (damage quality and good quality) compared to the DT algorithm. In addition, although they perform similarly, the two algorithms are produced from different combination preprocessing techniques. This shows that no superior preprocessing combination technique can be used for all machine learning algorithms. All iterations of preprocessing combinations are worth trying and can be efficiently searched using an auto-preprocessing strategy.

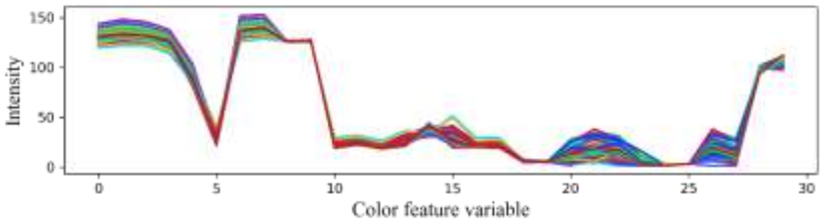
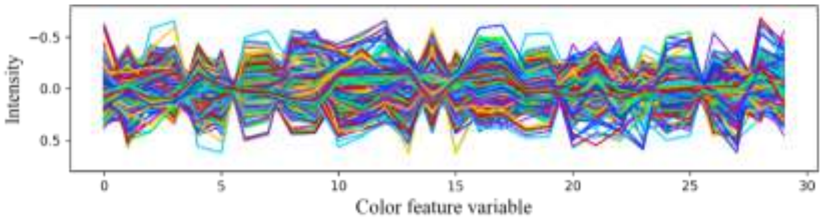
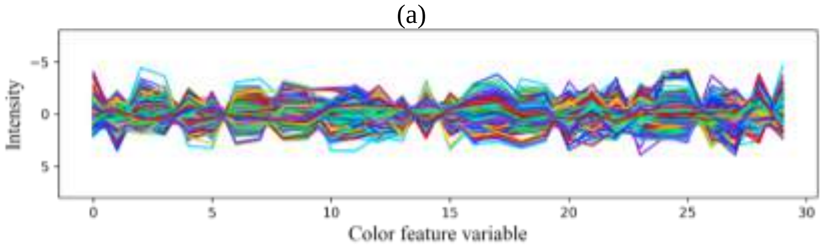


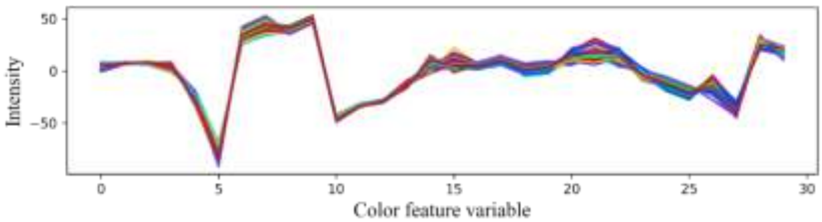
Figure 1. Raw chili pepper image variable using color feature extraction



(a)



(b)



(c)

Figure 3. Chili pepper image variable using color feature extraction after auto-preprocessing combined with (a) SVM, (b) DT, and (c) RF algorithm

The test results at the training and testing stages in classifying the quality of chili pepper using the color feature extractor as damage and good for all the machine learning algorithms tested in this study are presented in Table. The performance of the machine learning algorithm can increase in classifying chili paper quality after auto-preprocessing of the raw color feature. The performance of the chili paper quality classification model with auto-preprocessing of the variable color feature can improve the performance of all machine learning algorithms (SVM, DT, and RF) by 64.21%, 0.66%, and 0.65%, respectively. In particular, the SVM algorithm is beneficial in increasing the performance of the auto-preprocessing strategy proposed in this study from the previous accuracy of 0.60 to 0.99. This shows that the auto-preprocessing strategy is advantageous in improving the accuracy of machine vision-based classification models using a low-cost camera.

Table 2. Performance of the machine learning algorithm using color feature extraction

Algorithm	Pre-processing	Training				Testing			
		PRE	REC	F1S	ACR	PRE	REC	F1S	ACR
SVM	Original	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.60
		0.60	1.00	0.75		0.60	1.00	0.75	
	Auto-preprocessing	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.99
		1.00	1.00	1.00		0.99	0.99	0.99	
DT	Original	1.00	1.00	1.00	1.00	0.97	0.92	0.94	0.96
		1.00	1.00	1.00		0.95	0.98	0.96	
	Auto-preprocessing	1.00	1.00	1.00	1.00	0.95	0.95	0.95	0.96
		1.00	1.00	1.00		0.97	0.97	0.97	
RF	Original	1.00	1.00	1.00	1.00	0.95	0.98	0.97	0.97
		1.00	1.00	1.00		0.99	0.97	0.98	
	Auto-preprocessing	1.00	1.00	1.00	1.00	0.97	0.98	0.98	0.98
		1.00	1.00	1.00		0.99	0.98	0.98	

3.2. Classification model using ORB feature extraction

The raw data for the variable using the ORB feature extractor is shown in Figure 4. A total of 200 variables can be extracted from 420 image data for good quality and 105 data for damaged quality chili peppers. In general, the features of all samples appear very similar and challenging to distinguish from the naked eye. Also, the overlapping of each variable makes a single variable very difficult to use in developing a classification model of chili paper quality in this study.

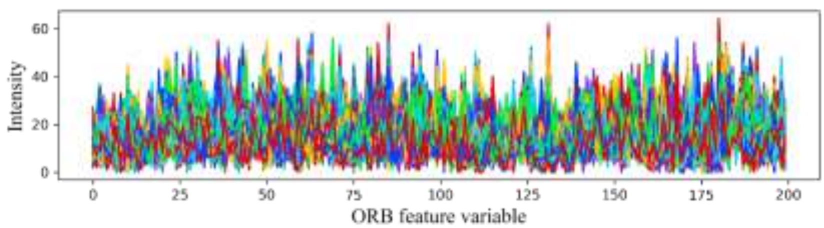


Figure 4. Raw chili pepper image variable using ORB feature extractor

The best variable auto-preprocessing for the classification of all machine learning algorithms used in this study using the ORB feature extractor is presented in Figure 5. The preprocessing baseline handling group represented by detrending (2nd order polynomial) is the only one with the best auto-preprocessing selection results with the SVM algorithm as shown in Figure 5(a) with perfect an accuracy cross-validation (1.00). In the DT algorithm as shown in Figure 5(b), the recommended auto-preprocessing comes from the group preprocessing noise handling and scaling-transformation handling, based on the smallest 5-fold cross-validation error. In particular, the combination of preprocessing S-G smoothing (window: 3 pt, order: 1) and range scaling is the best, giving an accuracy cross-validation of 0.564. The auto-preprocessing recommended for the RF algorithm from the ORB feature extractor is from the preprocessing baseline handling and scatter handling groups whose variables are presented in Figure 5(c). Combining the single preprocessing technique is detrending (4th order polynomial) and SNV with an accuracy cross-validation of 0.885. The auto-preprocessing evaluated by SVM gives the best chili pepper classification model performance (damage and good) compared to the RF and DT algorithms. In addition, no single type of preprocessing combination is superior in predicting the quality of chili paper using variable ORB features, which makes all combinations worth trying for every preprocessing.

The results of the tests at the training and testing stages to classify the quality of chili pepper using the ORB feature extractor as damage and good for all the machine learning algorithms used in this study are presented in Table 3. It can be seen that auto-preprocessing using the ORB variable can improve the performance of the paper classification model chili, specifically the DT algorithm. The model accuracy in the testing stage of the DT algorithm can increase from 0.86 before preprocessing to 0.90 after preprocessing. The performance improvement of the chili paper quality classification model using the ORB feature variable and the auto-preprocessing of the DT and RF algorithms was 4.41% and 0.64%, respectively. The SVM algorithm has no difference between before and after preprocessing. Once again, it shows that the auto-preprocessing strategy is advantageous in increasing the accuracy of machine vision-based classification models using low-cost cameras with variants derived from the ORB feature.

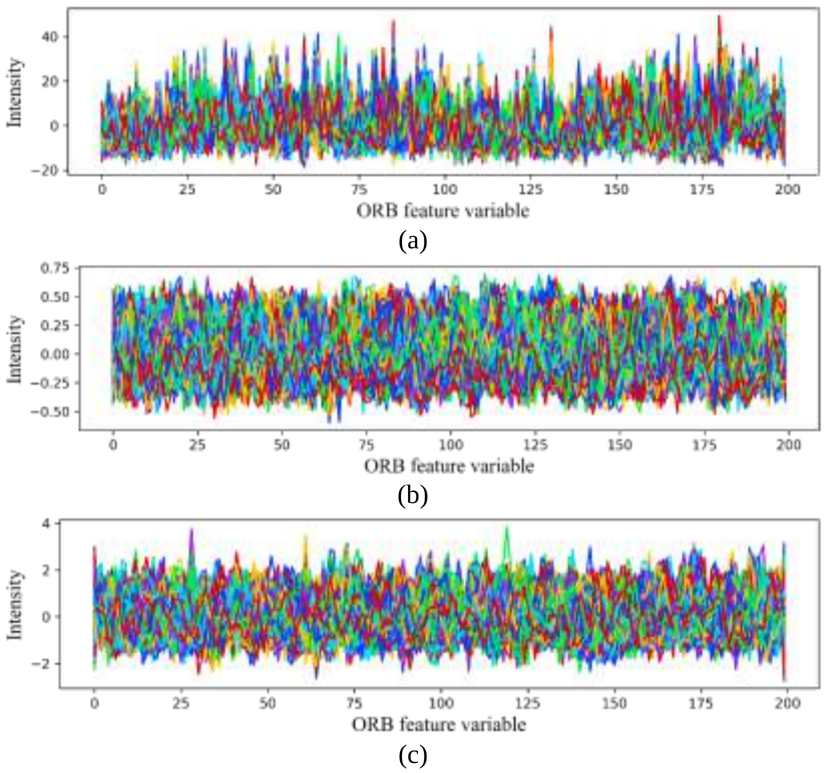


Figure 5. Chili pepper image variable using ORB feature extraction after auto-preprocessing combined with; (a) SVM, (b) DT, and (c) RF algorithm

Table 3. Performance of machine learning algorithms using ORB feature extractor

Algorithm	Pre-processing	Training				Testing			
		PRE	REC	F1S	ACR	PRE	REC	F1S	ACR
SVM	Original	1.00	1.00	1.00	1.00	0.98	1.00	0.99	0.99
		1.00	1.00	1.00		1.00	0.99	0.99	
	Auto-preprocessing	1.00	1.00	1.00	1.00	0.98	1.00	0.99	0.99
		1.00	1.00	1.00		1.00	0.99	0.99	
DT	Original	1.00	1.00	1.00	1.00	0.85	0.79	0.82	0.86
		1.00	1.00	1.00		0.87	0.91	0.89	
	Auto-preprocessing	1.00	1.00	1.00	1.00	0.86	0.89	0.88	0.90
		1.00	1.00	1.00		0.92	0.91	0.91	
RF	Original	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.99
		1.00	1.00	1.00		0.99	0.99	0.99	
	Auto-preprocessing	1.00	1.00	1.00	1.00	1.00	0.98	0.99	0.99
		1.00	1.00	1.00		0.99	1.00	0.99	

3.3. Classification model using combined color and ORB feature extraction

The raw data using combined color and ORB feature extractor are shown in Figure 6. A total of 230 variables can be extracted from 420 image data for good and 105 data for damaged chili peppers. All these variables combine the color feature and the ORB feature. It can be seen that the more variables that will be considered, the more their complexity in the analysis. Furthermore, the color feature variable has a more significant value distribution than the ORB feature in this study.

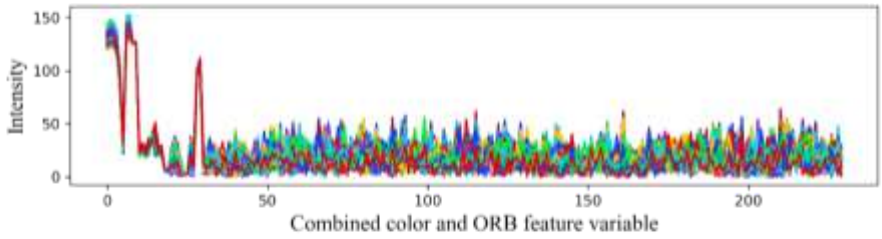


Figure 6. Raw chili pepper image variable using combined color and ORB feature extractor

The best auto-preprocessing variables for classifying all machine learning algorithms using the combined color and ORB feature extractor are presented in Figure 7. In the SVM algorithm as shown in Figure 7(a), the recommended auto-preprocessing is from the preprocessing group baseline handling, scatter handling, and noise handling, based on the smallest 5-fold cross-validation error. In particular, the combination of detrending preprocessing (4th order polynomial), SNV, and S-G smoothing (window: 3 pt, order: 1) is the best, giving an accuracy cross-validation of 1.00. The auto-preprocessing recommended for the DT algorithm using the combined color and ORB feature extractor result variables is from the preprocessing baseline handling scatter handling, noise handling, and scaling-transformation handling groups whose variables are presented in Figure 7(b). Combining single preprocessing techniques is detrending (2nd order polynomial), mean scaling, S-G smoothing (window: 3 pt, order: 1), and mean centering with an accuracy cross-validation of 0.592. The preprocessing group from baseline handling and scatter handling is represented by detrending (2nd order polynomial), and maximum scaling is the best result of choosing auto-preprocessing with the RF algorithm with an accuracy cross-validation of 0.989 as shown in Figure 7(c). The auto-preprocessing evaluated by SVM gives the best performance of the chili pepper classification model (damage quality and good quality) compared to the DT and RF algorithms. Furthermore, using variable features from the combined color and ORB feature extractor to predict the quality of chili paper did not find a superior combination of pre-processing methods. Therefore, iteration for all single preprocessing combinations using the auto-preprocessing strategy becomes crucial to achieving the best predictions.

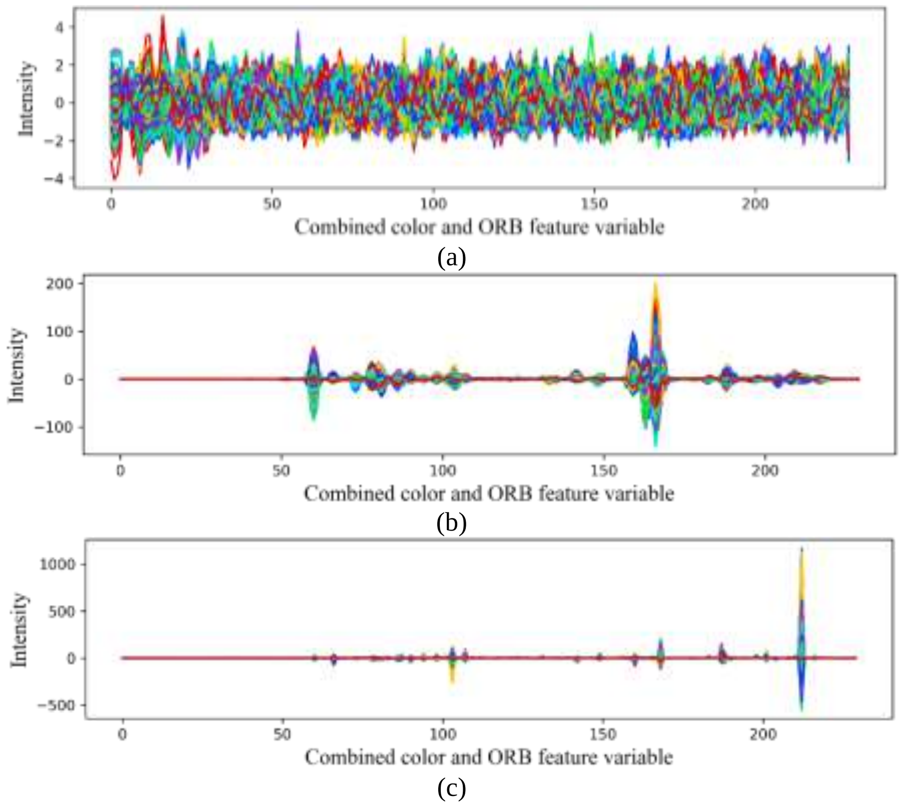


Figure 7. Chili pepper image variable using combined color and ORB feature extraction after auto-preprocessing combined with (a) SVM, (b) DT, and (c) RF algorithm

The performance of the machine learning algorithm (training and testing) in classifying the quality of chili pepper using the combined color and ORB feature extractor is presented in Table 4. It can be seen that the auto-preprocessing strategy can improve accuracy in modeling the quality classification of chili paper using variables derived from the combined color and ORB feature extraction of all machine-learning algorithms except DT. The SVM and RF algorithms can increase their accuracy to 100% in the training and testing stages. The performance improvement of the chili paper quality classification model using the SVM algorithm was 11.27% using the combination color feature and ORB feature, which was carried out by auto-preprocessing. Meanwhile, there is no difference in the accuracy of the RF algorithm before and after preprocessing. On the other hand, it differs from the DT algorithm, the accuracy at the testing stage has decreased from the previous 0.96 to 0.85. This is presumably because the number of trees used in the DT algorithm is unstable when using the auto-preprocessing result variable. This follows the results of research by Myles *et al.* [26] who stated that the parameter tree dramatically affects the performance of the DT algorithm. Apart from the performance of DT that was beyond the expectations of this study, other SVM and RF algorithms have shown very satisfactory performance in utilizing the auto-preprocessing strategy of the combined color and ORB feature variables. This further supports that the auto-preprocessing strategy is very useful in increasing the accuracy of machine vision-based classification models using low-cost cameras.

Table 4. Performance of machine learning algorithms using combined color and ORB feature extractor

Algorithm	Pre-processing	Training				Testing			
		PRE	REC	F1S	ACR	PRE	REC	F1S	ACR
SVM	Original	1.00	0.81	0.89	0.92	0.98	0.76	0.86	0.90
		0.89	1.00	0.94		0.86	0.99	0.92	
	Auto-preprocessing	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		1.00	1.00	1.00		1.00	1.00	1.00	
DT	Original	1.00	1.00	1.00	1.00	0.92	0.97	0.95	0.96
		1.00	1.00	1.00		0.98	0.95	0.96	
	Auto-preprocessing	1.00	1.00	1.00	1.00	0.83	0.79	0.81	0.85
		1.00	1.00	1.00		0.87	0.89	0.88	
RF	Original	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		1.00	1.00	1.00		1.00	1.00	1.00	
	Auto-preprocessing	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		1.00	1.00	1.00		1.00	1.00	1.00	

4. CONCLUSION

The comparison of three types of image feature extractors for the classification of chili pepper (Capsicum Frutescens) in damaged quality and good quality using a machine learning algorithm has been studied and presented in full in this study. The variables generated from the color feature extractor, ORB feature extractor, and their combinations can be used with satisfactory results in classifying the quality of chili paper captured using a low-cost camera such as a webcam camera. Surprisingly, three types of machine learning algorithms (SVM, DT, and RF) can work better with variables that have been pre-processed automatically. This study uses 420 image data for good and 105 data for damaged chili pepper, where the SVM and DT algorithms can classify them perfectly (with 100% accuracy). This study suggests using them with predictors derived from a combined color and ORB feature extractor pre-processed automatically. This research also shows that using a low-cost camera with the correct processing technique will be feasible in the future to develop a machine vision-based chili pepper grading machine.

REFERENCES

[1] X. Liming and Z. Yanchao, “Automated strawberry grading system based on image processing,” *Computers and Electronics in Agriculture*, vol. 71, no. SUPPL. 1, pp. S32--S39, Apr. 2010, doi: 10.1016/j.compag.2009.09.013.

[2] D. Ireri, E. Belal, C. Okinda, N. Makange, and C. Ji, “A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing,” *Artificial Intelligence in Agriculture*, vol. 2, pp. 28–37, Jun. 2019, doi: 10.1016/j.aiia.2019.06.001.

[3] M. R. Satpute and S. MJagdale, “Color, size, volume, shape and texture feature extraction techniques for fruits: a review,” *International Research Journal of Engineering and Technology*, vol. 3, no. 2010, pp. 2395–56, 2016.

[4] R. Bulan, Devianti, E. S. Ayu, and A. Sitorus, “Effects of moisture content on some engineering properties of arecanut (Areca Catechu L.) fruit which are relevant to the design of processing equipment,” *INMATEH - Agricultural Engineering*, vol. 60, no. 1, pp. 61–70, Apr. 2020, doi: 10.35633/INMATEH-60-07.

[5] N. Mohammadi Baneh, H. Navid, and J. Kafashan, “Mechatronic components in apple sorting machines with computer vision,” *Journal of Food Measurement and Characterization*, vol. 12, no. 2, pp. 1135–1155, Feb. 2018, doi: 10.1007/s11694-018-9728-1.

[6] N. B. A. Mustafa, S. K. Ahmed, Z. Ali, W. B. Yit, A. A. Z. Abidin, and Z. A. Md Sharrif, “Agricultural produce sorting and grading using support vector machines and fuzzy logic,” in *ICSIPA09 - 2009 IEEE International Conference on Signal and Image Processing Applications, Conference Proceedings*, 2009, pp. 391–396, doi: 10.1109/ICSIPA.2009.5478684.

[7] K. Mohi-Alden, M. Omid, M. Soltani Firouz, and A. Nasiri, “A machine vision-intelligent modelling based technique for in-line bell pepper sorting,” *Information Processing in Agriculture*, May 2022, doi: 10.1016/j.inpa.2022.05.003.

[8] H. Familiana, I. Maulana, A. Karyadi, I. S. Cebro, and A. Sitorus, “Characterization of aluminum surface using image processing methods and artificial neural network methods,” in *3rd International Conference on Computing, Engineering, and Design, ICCED 2017*, Nov. 2018, vol. 2018-March, pp. 1–6, doi: 10.1109/CED.2017.8308113.

[9] S. Gupta, M. Kumar, and A. Garg, “Improved object recognition results using SIFT and ORB feature detector,” *Multimedia Tools and Applications*, vol. 78, no. 23, pp. 34157–34171, Oct. 2019, doi: 10.1007/s11042-019-08232-6.

[10] C. Shao, C. Zhang, Z. Fang, and G. Yang, “A deep learning-based semantic filter for RANSAC-based fundamental matrix calculation and the ORB-SLAM system,” *IEEE Access*, vol. 8, pp. 3212–3223, 2020, doi: 10.1109/ACCESS.2019.2962268.

[11] N. D. Arianti, E. Saputra, and A. Sitorus, “An automatic generation of pre-processing strategy combined with machine learning multivariate analysis for NIR spectral data,” *Journal of Agriculture and Food Research*, vol. 13, p. 100625, Sep. 2023, doi: 10.1016/j.jafr.2023.100625.

[12] O. Mahendra, H. F. Pardede, R. Sustika, and R. B. S. Kusumo, “Comparison of features for strawberry grading classification with novel dataset,” in *2018 International Conference on Computer, Control, Informatics and its Applications: Recent Challenges in Machine Learning for Computing Applications, IC3INA 2018 - Proceeding*, Nov. 2019, pp. 7–12, doi: 10.1109/IC3INA.2018.8629534.

[13] V. A. Tuesta, F. D. Alcarazo, H. I. Mejia, and M. G. Forero, “Automatic classification of citrus aurantifolia based on digital image processing and pattern recognition,” in *Applications of Digital Image Processing {XLIII}*, Aug. 2020, p. 16, doi: 10.1117/12.2566888.

[14] A. M. T. Nasution and S. A. Amrullah, “Simple vision system for Apple varieties classification,” *Industria: Jurnal Teknologi dan Manajemen Agroindustri*, vol. 11, no. 1, pp. 51–63, Apr. 2022, doi: 10.21776/ub.industria.2022.011.01.6.

[15] S. Das, D. Roy, and P. Das, “Disease feature extraction and disease detection from paddy crops using image processing and deep learning technique,” in *Advances in Intelligent Systems and Computing*, vol. 1120, Springer Singapore, 2020, pp. 443–449.

[16] S. R. Maniyath *et al.*, “Plant disease detection using machine learning,” in *Proceedings - 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control, ICDI3C 2018*, Apr. 2018, pp. 41–45, doi: 10.1109/ICDI3C.2018.00017.

[17] N. Nandhini and R. Bhavani, “Feature extraction for diseased leaf image classification using machine learning,” Jan. 2020, doi: 10.1109/ICCCI48352.2020.9104203.

[18] G. Wu and Z. Zhou, “An improved ORB feature extraction and matching algorithm,” in *Proceedings of the 33rd Chinese Control and Decision Conference, CCDC 2021*, May 2021, pp. 7289–7292, doi: 10.1109/CCDC52312.2021.9602102.

[19] J. Weberruss, L. Kleeman, and T. Drummond, “ORB feature extraction and matching in hardware,” *Australasian Conference on Robotics and Automation, ACRA*, pp. 2–4, 2015.

[20] W. Fang, Y. Zhang, B. Yu, and S. Liu, “FPGA-based ORB feature extraction for real-time visual SLAM,” in *2017 International Conference on Field-Programmable Technology, ICFPT 2017*, Dec. 2018, vol. 2018-January, pp. 275–278, doi: 10.1109/FPT.2017.8280159.

[21] M. S. Nixon and A. S. Aguado, “Feature extraction and image processing for computer vision,” *Feature Extraction and Image Processing for Computer Vision*, pp. 1–609, 2012, doi: 10.1016/C2011-0-06935-1.

[22] E. Mwebaze and G. Owomugisha, “Machine learning for plant disease incidence and severity measurements from leaf images,” in *2016 15th {IEEE} International Conference on Machine Learning and Applications ({ICMLA})*, Dec. 2017, pp. 158–163, doi: 10.1109/icmla.2016.0034.

[23] K. Koyama, M. Tanaka, B. H. Cho, Y. Yoshikawa, and S. Koseki, “Predicting sensory evaluation of spinach freshness using machine learning model and digital images,” *PLoS ONE*, vol. 16, no. 3 March, p. e0248769, Mar. 2021, doi: 10.1371/journal.pone.0248769.

[24] J. Engel *et al.*, “Breaking with trends in pre-processing?,” *TrAC - Trends in Analytical Chemistry*, vol. 50, pp. 96–106, Oct. 2013, doi: 10.1016/j.trac.2013.04.015.

[25] A. I. Belousov, S. A. Verzakov, and J. Von Frese, “Applicational aspects of support vector machines,” *Journal of Chemometrics*, vol. 16, no. 8–10, pp. 482–489, 2002, doi: 10.1002/cem.744.

[26] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, “An introduction to decision tree modeling,” *Journal of Chemometrics*, vol. 18, no. 6, pp. 275–285, Jun. 2004, doi: 10.1002/cem.873.

[27] B. P. O. Lovatti *et al.*, “Different strategies for the use of random forest in NMR spectra,” *Journal of Chemometrics*, vol. 34, no. 12, Mar. 2020, doi: 10.1002/cem.3231.

[28] Pedregosa F. *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011, [Online]. Available: http://scikit-learn.sourceforge.net.

[29] R. G. Brereton, “Contingency tables, confusion matrices, classifiers and quality of prediction,” *Journal of Chemometrics*, vol. 35, no. 11, Apr. 2021, doi: 10.1002/cem.3331.

BIOGRAPHIES OF THE AUTHORS

Jelita Asian is an Assistant Professor at Nusa Putra University, Sukabumi, Indonesia. She earned her Ph.D. from RMIT University, Melbourne, Australia, in 2007. Her research interest includes data mining, machine learning, and soft computing. She can be contacted at email: jelita.asian@nusaputra.ac.id.

Nunik Destria Arianti is an Assistant Professor at Nusa Putra University, Sukabumi, Indonesia. She is presently pursuing a Ph.D. degree in the informatic system. Her research interests include data machine learning and image processing. She can be contacted at email: nunik@nusaputra.ac.id.

Ariefin is an Associate professor at the Polytechnic State of Lhokseumawe, Lhokseumawe, Indonesia. He completed his Master's in mechanical engineering in 2010 from Brawijaya University. His research interests include machine learning and image processing. He can be contacted at email: arieting62@yahoo.com.

Muhamad Muslih is an Assistant Professor at Nusa Putra University, Sukabumi, Indonesia. He completed his Master's degree in computer science in 2015. His research interests include System analysis and design, operating system testing and implementation and human and computer interaction. He can be contacted at email: muhamad.muslih@nusaputra.ac.id.