Buletin Ilmiah Sarjana Teknik Elektro
Vol. 5, No. 4, December 2023, pp. 539-551
ISSN: 2685-9572, DOI: 10.12928/biste.v5i4.9145

Optimizing Banana Type Identification: An Support Vector Machine Classification-Based Approach for Cavendish, Mas, and Tanduk Varieties

Aji Pamungkas, Abdul Fadlil
Program Studi Teknik Elektro, Universitas Ahmad Dahlan, Indonesia

ARTICLE INFORMATION

Article History:

Submitted 14 October 2023
Revised 12 December 2023
Accepted 29 December 2023

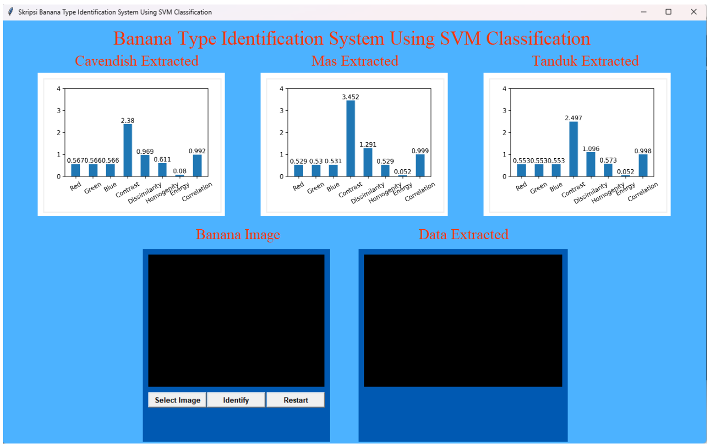
Keywords:

Banana;
RGB;
GLCM;
SVM

Corresponding Author:

Abdul Fadlil,
Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia.
Email: fadlil@mti.uad.ac.id

ABSTRACT



This research focuses on addressing the need for improved efficiency in the agricultural sector, particularly in banana processing in Indonesia, where the demand for bananas is consistently high. To improve the efficiency of banana processing, the research proposes the development of a machine learning based solution for automatic banana type selection. This solution uses image data of three banana types (Cavendish, Mas, and Tanduked) captured by a microscopic camera. The images are subjected to feature extraction, and a Support Vector Machine (SVM) algorithm is used to train the model. The results are implemented in a graphical user interface (GUI). The experimental results show promising results, with an accuracy of 86.67%, a precision of 87.78%, and an error rate of 13.33%, achieved with SVM parameters of C = 1000 and a linear kernel. This automated approach provides a practical and sustainable solution to the labor-intensive manual banana variety selection process, thus increasing the efficiency of the banana processing industry.

Document Citation:

A. Pamungkas and A. Fadlil, “Optimizing Banana Type Identification: An Support Vector Machine Classification-Based Approach for Cavendish, Mas, and Tanduk Varieties,” *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 5, no. 4, pp. xx-xx, 2023, DOI: 10.12928/biste.v5i4.9145.

1. INTRODUCTION

The rapid advancement of technology has led to increased effectiveness and efficiency in various sectors, and agriculture is no exception [1],[2]. In the context of agriculture, technological progress plays a pivotal role in facilitating numerous activities, particularly in the processing of agricultural and plantation products [3]. Among the sectors experiencing significant growth due to technological innovation is the banana processing industry in Indonesia. Bananas enjoy immense popularity among consumers [4], both as direct consumption and as processed products such as banana chips [5], banana bread, and banana jam.

Data compiled by the Directorate General of Horticulture and the Central Statistics Agency (Badan Pusat Statistik Indonesia) over the past three years show a steady upward trend in Indonesia's banana production. In 2022, banana production reached a remarkable 9.24 million tons, an increase of 5.77% (504 thousand tons) compared to the previous year. Household consumption of bananas also increased to 2.42 million tons, representing a growth of 1.35% (32.14 thousand tons) from the previous year [6]. These statistics underscore the growing demand and popularity of bananas in the market, indicating Indonesia's considerable potential in banana production.

However, despite the robust banana production in Indonesia, the banana processing industry remains characterized by inefficiencies. This research seeks to support the nascent but promising banana processing sector. The banana processing pipeline consists of several stages, with a critical stage being the selection of bananas from plantations and agricultural sources suitable for processing [7]. This is essential because each banana variety requires different processing techniques, resulting in different processing outcomes.

Traditionally, bananas are selected based on their physical characteristics, including shape [8], size [9], and color [10]-[13]. Color assessment includes evaluation of RGB standard deviation and RGB mean [14], while size criteria include parameters such as perimeter, area, and width [15]. Shape criteria consider slenderness and roundness [16]. The differentiation process aims to categorize bananas according to their variety. Manual selection by banana farmers or selectors relies mainly on size and skin color because of their observable nature. However, manual selection proves to be an inefficient approach due to discrepancies in human perception regarding color, shape, and size components within the same images.

In addition, with the expected increase in banana production in Indonesia, more banana processing facilities are expected to emerge. However, the expansion of banana processing faces challenges, particularly in the manual selection of banana varieties. Relying on manual labor for this task would result in a significant increase in the number of workers and, consequently, in rising labor costs. This would lead to dissatisfaction among communities and processing facilities, as the demand for processed bananas would remain unmet. Therefore, there is an urgent need for alternative solutions to mitigate this challenge.

This research seeks to develop a machine learning-based solution that efficiently automates the selection of banana varieties. It is envisioned that this solution will significantly improve the efficiency and effectiveness of the banana processing industry, providing a practical and sustainable remedy to the prevailing problems.

2. METHODS

This research method aims to find out how to carry out the process of collecting banana data using a microscopic camera, to know the feature extraction process used to identify banana types, to understand the classification process using the SVM method to implement banana type identification, and to apply a GUI as an implementation of banana type identification to make it easier to use.

2.1. Block Diagram

Figure 1 shows the block diagram describing the architecture of the banana identification system using SVM classification for the differentiation of Cavendish, Mas and Tanduk banana varieties. The research apparatus is structured around a microscopic camera, which acts as the primary data acquisition tool, and a laptop computer connected via serial communication, which provides the critical link between the data source and the processing unit. Within the laptop's computational framework, the SVM classification algorithm operates to distinguish banana types based on feature extraction from the captured images.

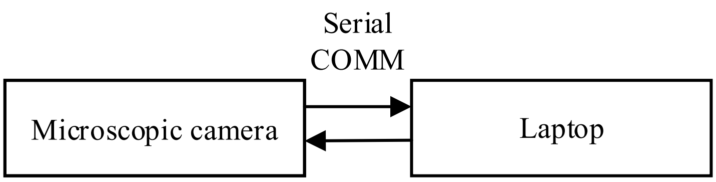


Figure 1. System block diagram

This block diagram succinctly encapsulates the core components and workflow of the banana identification system, underscoring the critical integration of hardware (microscopic camera) and software (SVM classification) to achieve automated and accurate banana type discrimination, with profound implications for optimizing banana processing and agricultural practices.

2.2. Flowchart

In this research, there are several flow diagrams used to identify types of bananas. This flowchart starts with data collection shown in Figure 2, feature extraction and training data shown in Figure 3 (a), and test data shown in Figure 3 (b).

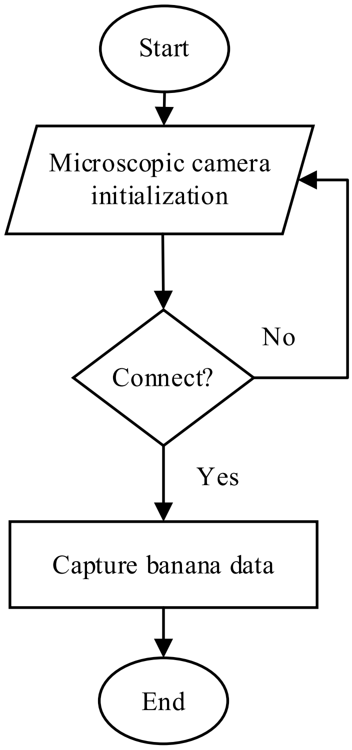


Figure 2. Data collection diagram

Figure 2 is a flow chart representation of the data collection process that will be carried out in this research. Before the data collection process was carried out, three types of bananas that would be used in this research were prepared, namely Cavendish bananas, Mas bananas, and Tanduk bananas. For each type of banana, data was collected one hundred times as research samples. This data collection process is carried out using predetermined methods and following predetermined procedures.

Based on Figure 3 (a) the first step in this research is to extract features using RGB and GLCM methods. Then, the feature extraction results are combined into an array and the data are labeled according to the predetermined class, namely Cavendish bananas with label 1, Mas bananas with label 2, and Tanduk bananas with label 3. Then, the data is divided into two parts, namely training data and test data. This data division is done with a ratio of 80% for training data and 20% for test data. The training data will be used to train a classification model using the SVM algorithm. The SVM parameters used to train the training data are the kernel and C parameters. Classification results are evaluated using precision, recall and accuracy metrics to determine the extent to which the trained model is successful in classifying the data well.

Based on Figure 3 (b), the first step is to input the image to be identified using a model that has been trained based on the flowchart in Figure 3 (a), then the feature extraction process is performed on the image according to the feature extraction used in the training model. Then, the test data is combined into test data, the test data is processed for identification using the training model, and the prediction results using the training model are displayed. By using this method, it is hoped to obtain relevant and accurate information about the performance of the identification model in identifying banana varieties. This will provide a better understanding of the effectiveness of the developed system in identifying bananas based on the features extracted using the RGB and GLCM methods.

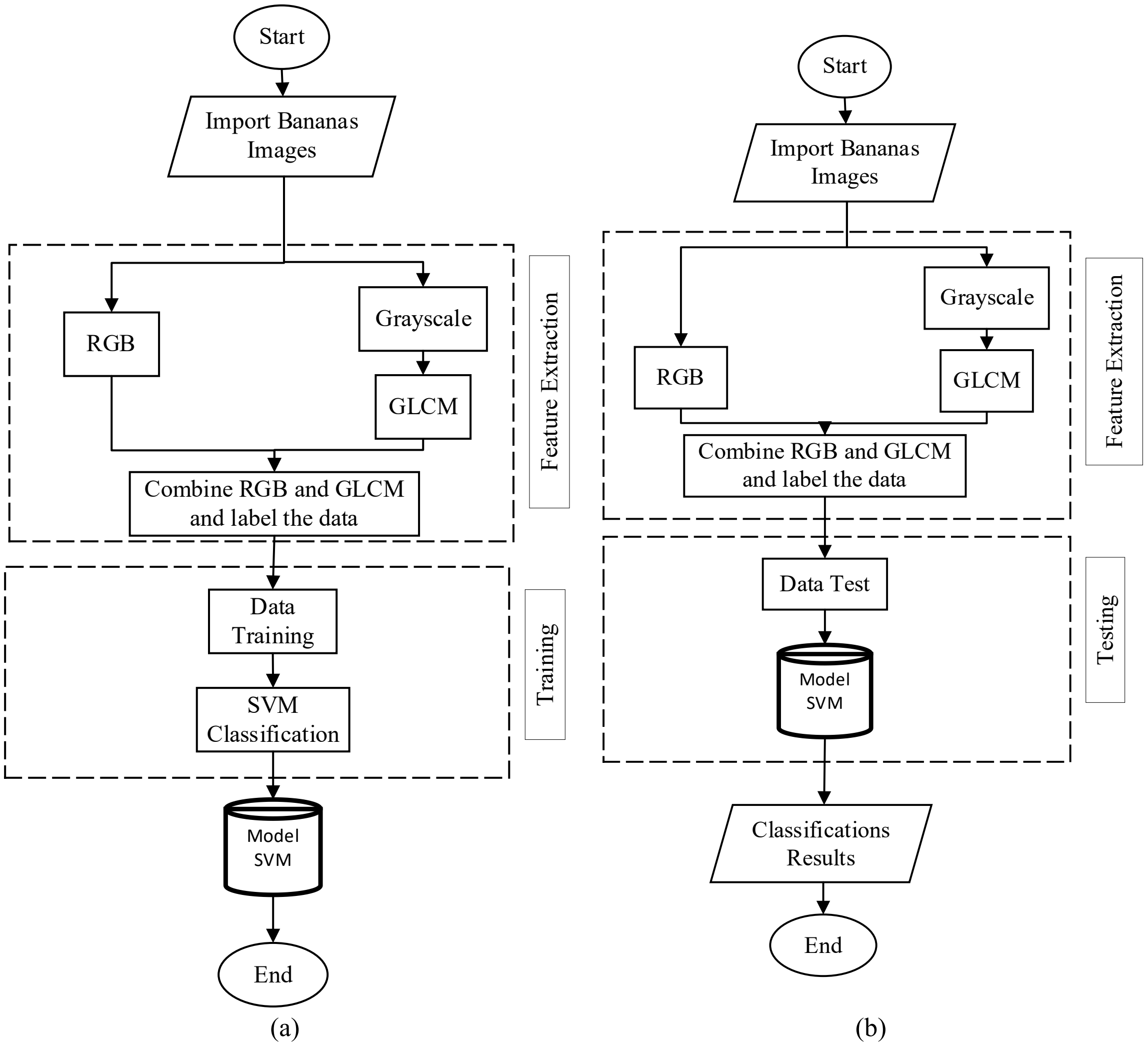


Figure 3. (a) Flowchart of data training (b) data testing

2.3. Equation Formula

This research has several equations to find the values that will be needed in the banana identification process, here is the equation formula.

a. RGB

The development of the RGB color model was based on human understanding of trichromatic theory before the era of electronic devices. To find the RGB value [17], use the formula available in the In Python library, which is shown in equations (1), (2), and (3).

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

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

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

In finding RGB values, there is a basic formula that is used. R refers to the value of the red component, G refers to the value of the green component, and G refers to the value of the blue component. To determine the brightness level of each component, the mean or middle value function can be used.

b. Gray Level Co-occurrence Matrix (GLCM)

GLCM is a feature extraction method that uses second-order texture computations by considering pairs of two pixels in the original image [18]-[20]. The basic formula for calculating the GLCM extraction values is shown in equations (4), (5), (6), (7), and (8).

$$energy = \sum_i \sum_j p(i,j)^2 \; r = \frac{R}{R+G+B} \tag{4}$$

$$contrast = \sum_i \sum_j p(i,j)^2 . p(i,j) \tag{5}$$

$$correlation = \frac{\sum_i \sum_j (i - \mu_x).(j - \mu_y).p(i,j)}{\sqrt{\delta_x . \delta_y}} \tag{6}$$

$$homogenity = \sum_i \sum_j \frac{p(i,j)}{1 + (1-j)^2} \tag{7}$$

$$dissimilarity = \sum_i \sum_j |i - j| p(i,j) \tag{8}$$

c. Support Vector Machine (SVM)

SVM is a machine learning algorithm used for classification and regression problems. SVM works by constructing an optimal hyperplane that separates two classes of data from each other [21]-[23]. Mathematically, SVM works by solving optimization problems. Given some training data $\{(x1, y1), (x2, y2), \dots, (xn, yn)\}$, where xi is a feature vector and yi is the associated class label ($y1 = -1 \text{ or } y1 = 1$) [24], SVM searches for the optimal hyperplane to maximize the margin. The margin is defined as the shortest distance between the hyperplane and the support vector. The hyperplane can be expressed as a linear equation of the form $w - x + b = 0$, where w is the normal vector perpendicular to the hyperplane, x is the feature vector, and b is the bias [25]. The dimensions of the SVM are represented in equation (9) using the hyperplane equation.

$$w \cdot x + b = 0 \tag{9}$$

For data that is in the negative class, the equation is (10).

$$w \cdot x_a + b \leq -1 \tag{10}$$

For data that is in the positive class, the equation is (11).

$$w \cdot x_b + b \geq 1 \tag{11}$$

d. Evaluation of Classification (Confusion Matrix)

The evaluation of classification methods involves a rigorous examination of their performance through the application of specific metrics. This evaluation is performed using the formulations described in equations (12), (13), (14), and (15). Among these key metrics, accuracy [26] serves as a basic indicator of overall correctness, reflecting the proportion of correctly classified instances in the total dataset. Precision [27], on the other hand, measures the proportion of true positive classifications relative to all positive predictions, indicating the ability of the method to avoid false positives. At the same time, recall [28] measures the proportion of true positive predictions relative to the actual positive instances, highlighting the method's ability to capture all relevant results. Error, as a complementary metric, quantifies the rate of misclassification in the classification process, shedding light on the efficiency of the method. Together, these metrics provide a comprehensive evaluation of the classification method's effectiveness, ensuring a thorough understanding of its performance in practice. These equations are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN), each of which is related to each class. The higher the TP (True Positive) and TN (True Negative) values, the better the accuracy the model can achieve.

$$Accuracy(\%) = \frac{TN + TP}{TN + TP + FN + FP} x100\% \tag{12}$$

$$Precision(\%) = \frac{TP}{TP + FP} x100\% \tag{13}$$

$$Recall(\%) = \frac{TP}{TP + FN} x100\% \tag{14}$$

$$Error(\%) = \frac{FP}{TP} x100\% \tag{15}$$

2.4. Research Tools and Materials

The hardware design in this research involves integrating a microscopic camera with a laptop using a USB cable to initialize direct communication between the two. The communication between the camera and the laptop uses a serial method, which allows the laptop to retrieve the data captured by the microscopic camera. The microscopic camera used in this research is shown in Figure 4, and the GUI display on the laptop is shown in Figure 5. In the application shown in Figure 5, there is a display of the sample extraction values for each type of Cavendish, Mas and Tanduk banana. The Banana Identification application has 3 functions, namely the selection of the banana image to be identified, then the classification button is a command to identify the selected banana image, and the restart button is used to reset the identification results.

Figure 4. Microscopic camera

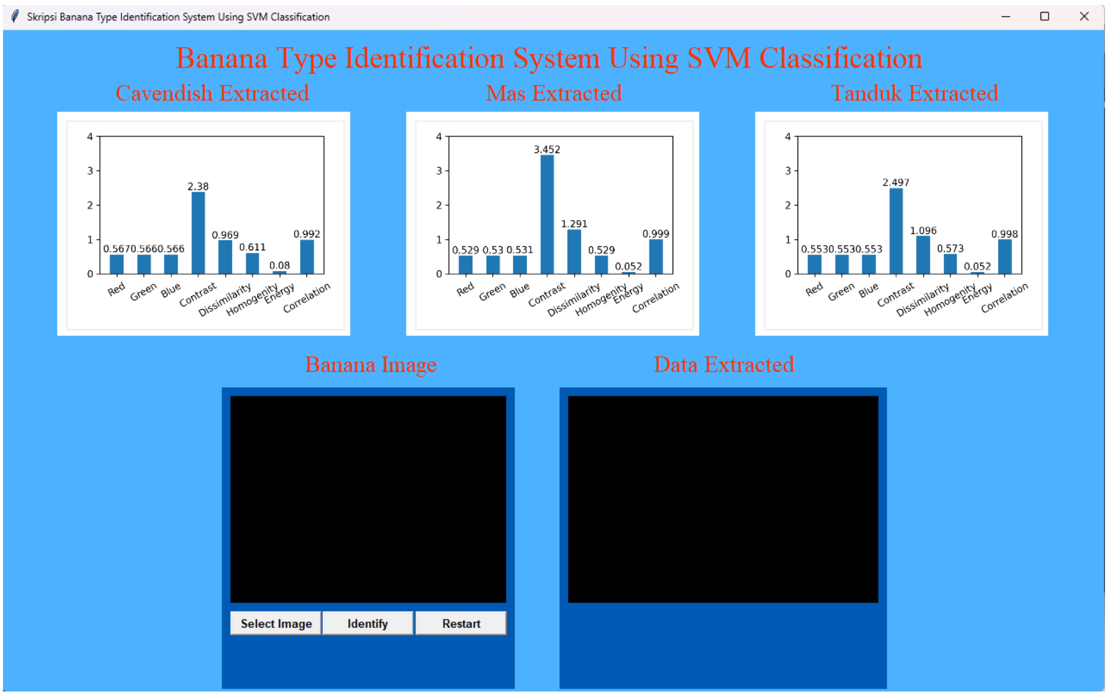


Figure 5. GUI display

3. RESULT AND DISCUSSION

This research carried out several tests on three types of banana, namely Cavendish, Mas, and Tanduk, which were taken using a microscopic camera. Some of these tests are described as follows.

3.1. Data Collection Results

In this research, the examination encompassed the analysis of image data pertaining to three distinct banana types: Cavendish, Mas, and Tanduk. The images were recorded in RGB (Red, Green, and Blue) format, featuring a resolution of 640 x 480 pixels and a density of 96 dots per inch (dpi). Each banana category was represented by a comprehensive dataset, consisting of precisely 100 images, thus amounting to a total of 300 images for the entire research. Subsequently, the image data was systematically segregated into two subsets, specifically designated as training data and test data. The division of data was executed utilizing three distinct partitioning ratios, which included a 50% allocation to training data and 50% to test data, an 80% training data and 20% test data distribution, and finally, a 90% training data and 10% test data distribution.

Figure 6 serves as an illustrative representation, providing a glimpse of the sample data acquired for each banana type as an integral component of this research. These meticulously collected datasets and their division into training and test subsets form the foundation upon which subsequent analyses and discussions regarding the efficacy and performance of the banana type identification system are predicated.

(a) (b) (c)

Figure 6. Sample banana data (a) Cavendish, (b) Mas, (c) Tanduk

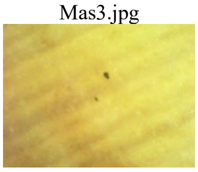
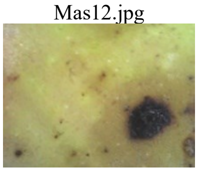
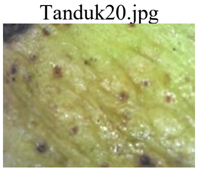
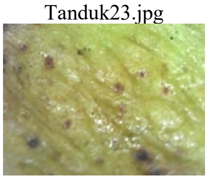
3.2. Feature Extraction Results

In this research, researchers have processed the banana image data that has been taken to carry out feature extraction using the RGB (Red, Green and Blue) and GLCM (Gray-Level Co-occurrence Matrix) color methods. The feature extraction process is carried out to obtain relevant information from each banana image. The information obtained from feature extraction will be an important basis for developing an accurate and reliable banana type identification system. This feature extraction is shown in Table 1.

An examination of the data presented in Table 1 reveals notable similarities between Mas and Tanduk bananas, particularly concerning attributes such as Green, Blue, Energy, and Correlation characteristics. These shared characteristics are evident from the closely aligned values, implying a degree of resemblance between the two banana varieties in these specific aspects. However, it is imperative to underscore those significant disparities were observed in the values of other traits under consideration.

Table 1. Sample results from extracting characteristics of three types of bananas Cavendish, Mas, Tanduk

Original Image	Red	Green	Blue	Contrast	Dissimilarity	Homogenity	Energy	Correlation
 Cavendish3.jpg	0.567	0.566	0.566	2.38	0.969	0.611	0.08	0.992
 Cavendish6.jpg	0.567	0.566	0.567	2.229	0.959	0.613	0.082	0.992

Original Image	Red	Green	Blue	Contrast	Dissimilarity	Homogenity	Energy	Correlation
 Cavendish27.jpg	0.568	0.585	0.585	2.052	0.942	0.617	0.072	0.996
 Mas3.jpg	0.629	0.628	0.63	3.075	1.187	0.553	0.054	0.997
 Mas11.jpg	0.624	0.621	0.621	3.473	1.207	0.55	0.051	0.997
 Mas12.jpg	0.596	0.593	0.592	3.632	1.328	0.52	0.044	0.998
 Tanduk20.jpg	0.404	0.405	0.406	2.335	1.067	0.577	0.057	0.998
 Tanduk23.jpg	0.585	0.586	0.586	2.61	1.123	0.566	0.053	0.996
 Tanduk45.jpg	0.537	0.534	0.535	2.516	1.115	0.565	0.057	0.999

3.3. Classification Results

In the course of this investigation, a curated dataset comprising banana type data that has undergone a rigorous feature extraction process was amalgamated into a unified and comprehensive dataset. Each individual data sample was meticulously annotated with a numerical label corresponding to the specific banana type it represented, facilitating computational discernment and classification. Specifically, a numerical encoding scheme was adopted, wherein the label "0" denoted the Cavendish banana, "1" designated the Mas banana, and "2" signified the Tanduk banana. This numerical labeling system was instrumental in enabling the computer-based classification to accurately identify and classify the diverse banana types under scrutiny.

The outcomes of the classification process, as informed by the test results, are succinctly presented in Table 2. This tabulated data serves as a vital repository of information delineating the effectiveness and precision of the banana type identification system. The results encapsulated in Table 2 are instrumental in evaluating the performance and accuracy of the Support Vector Machine (SVM) classifier within the framework of this research, thus underpinning the discussions and conclusions to follow in this research.

Table 2. Parameter testing of the SVM algorithm

No	C	Kernel	Accuracy (%) Training data 50%	Accuracy (%) Training data 80%	Accuracy (%) Training data 90%
1	0,1	rbf	63.33	75.00	61.11
2	0,1	*linear*	60.00	79.20	63.00
3	0,1	*poly*	63.33	77.00	59.33
4	1	rbf	66.67	77.10	64.80
5	1	*linear*	66.67	81.20	64.80
6	1	*poly*	70.00	79.20	61.11
7	10	rbf	66.67	79.20	72.20
8	10	*linear*	70.00	81.20	68.50
9	10	*poly*	70.00	83.00	63.00
10	100	rbf	70.00	81.20	74.80
11	100	*linear*	70.00	83.00	70.40
12	100	*poly*	73.33	85.00	70.40
13	1000	rbf	73.33	83.33	79.60
14	1000	*linear*	73.33	86.67	83.33
15	1000	*poly*	73.33	85.40	81.50

The investigation into banana type identification hinged upon the utilization of the Support Vector Machine (SVM) algorithm, marked by an exhaustive exploration of varying parameter configurations, including C values (C=0.1, 1, 10, 100, and 1000), kernel functions (rbf, linear, and poly), and diverse data allocation strategies for training purposes (50%, 80%, and 90%). The paramount objective of this battery of tests was to discern the influence of these variable combinations on the accuracy of the identification outcomes.

Upon identifying the most optimal parameter values, the researcher further delved into the impact of data partitioning between training and test sets. This entailed a meticulous analysis of three specific data allocation ratios, specifically: 90% training data and 10% test data, 80% training data and 20% test data, and 50% training data and 50% test data. Table 3, presented below, serves as an encapsulation of the results stemming from this assessment, delineating the ramifications of varied data ratios on the identification of banana types. These findings are fundamental to the ensuing discussions and conclusions, which underscore the optimal parameter configurations and data partitioning strategies to enhance the effectiveness of the SVM classifier in the context of banana type identification.

Table 3. Results of testing the features used on banana type data

No	Testing	Feature Extraction	Accuracy (%)
1	Training data 90%; test data 10%	GLCM	80
2	Training data 90%; test data 10%	RGB + GLCM	83.33
3	Training data 80%; test data 20%	GLCM	73.33
4	Training data 80%; test data 20%	RGB + GLCM	86.67
5	Training data 50%; 50% test data	GLCM	72.67
6	Training data 50%; 50% test data	RGB + GLCM	79.33

The presented results in the Table 3 illustrate the significant influence of feature extraction methods and the partitioning of training and test data on the accuracy of banana type identification. Notably, when utilizing the Gray-Level Co-occurrence Matrix (GLCM) feature extraction method, a higher percentage of training data, such as 90%, consistently leads to superior accuracy, as exemplified by the 80% accuracy achieved. Intriguingly, the integration of both RGB (Red, Green, Blue) and GLCM features within the feature set further bolsters accuracy to 83.33% under the same data allocation conditions, underscoring the synergistic potential between these feature types.

In contrast, reducing the proportion of training data to 80% demonstrates a reduction in accuracy with GLCM as the sole feature extraction method, resulting in 73.33% accuracy. Nevertheless, incorporating both RGB and GLCM features under the same conditions substantially elevates accuracy to 86.67%, manifesting the complementarity of these feature types.

When the data allocation is balanced at 50% for both training and test data, regardless of the feature extraction method used, there is a minor decline in accuracy. For instance, GLCM alone yields 72.67% accuracy, whereas the inclusion of both RGB and GLCM features elevates accuracy to 79.33% within this data partitioning framework.

3.4. Classification Test Results

Based on the results of classification and parameter testing Table 2 and features used Table 3, researchers chose the highest accuracy value to carry out classification tests using the confusion matrix. Confusion matrix is used to evaluate accuracy, precision, recall and error values using equation (12) – equation (15). Table 4 shows the confusion matrix.

Based on data analysis in this research, calculation results were obtained which showed that the accuracy value reached 86.667%. Accuracy is a measure to measure the extent to which the identification model can provide correct predictions. Furthermore, the precision results for each class of Cavendish Banana, Mas Banana, and Tanduk Banana were 90.476%, 92.857%, and 80%, respectively. The overall precision in this research reached 87.777%.

Table 4. Confusion Matrix

		Actual		
		Cavendish	**Mas**	**Tanduk**
Predicted	**Cavendish**	19	1	0
	Mas	2	13	5
	Tanduk	0	0	20

3.5. Model Testing with PCA

The dataset that has undergone initial extraction is then subjected to a dimension reduction process using the PCA method with the aim of reducing the dimensions of the original eight characteristics to only two main characteristics. The data obtained from PCA are shown in Table 5.

Table 5. PCA characteristic data based on GUI testing

Types of Bananas	PC1	PC2
Cavendish	-1.06857715	-0.03350558
Cavendish	0.13391553	0.08295698
Mas	-0.52129645	-0.04295051
Mas	-0.67741068	0.07899853
Tanduk	2.04562957	-0.00423230
Tanduk	0.08773917	-0.08126712

Table 5 displays the results of the process of reducing the characteristic dimensions (R, G, B, Contrast, Dissimilarity, Homogeneity, Energy and Correlation) into two main components (PC1 and PC2) with different values. Figure 7 shows the visualization of the SVM classification with PCA features reduction.

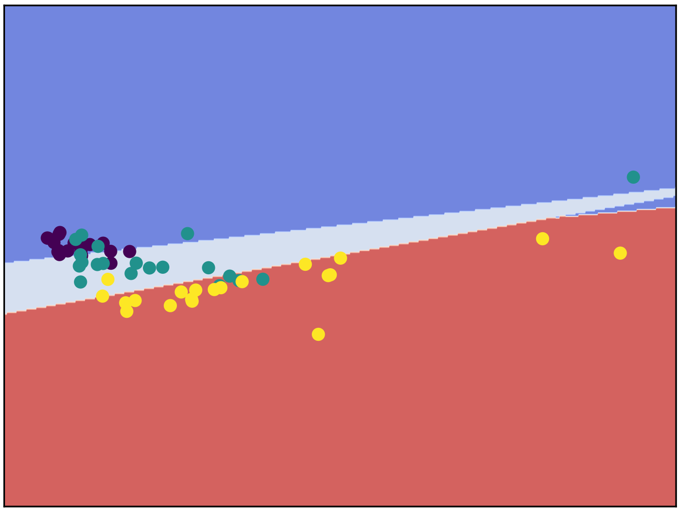


Figure 7. SVM classification visualization

In Figure 7, visual observation shows that the purple data represents Cavendish banana in the blue region, while the green data represents Mas banana in the white background, and the yellow data represents Tanduk banana in the red region. The classification results using the Support Vector Machine (SVM) method with the application of Principal Component Analysis (PCA) on the same dataset as in the classification without PCA resulted in different evaluation metrics; 73.333% accuracy, 72.643% precision, and 73.333% recall, in contrast to the classification without PCA which had 86.667% accuracy, 87.777% precision, and 86.667% recall. Overall, the integration of the PCA method does not significantly improve the classification performance, although it provides an advantage in the visualization of the classification results.

3.6. Classification Test Results using GUI

The results of this identification were performed using a model trained on the original image of Pisang Mas. Table 6 shows the results of the identification test using the GUI.

Table 6. Feature extraction test results using GUI

No	R	G	B	Contrast	Dissimilarity	Homogen	Energy	Corr	Types of Bananas
1	0.574	0.573	0.574	1.625	0.880	0.629	0.083	0.994	Cavendish
2	0.567	0.568	0.569	2.823	1.036	0.597	0.071	0.994	Cavendish
3	0.564	0.563	0.562	2.153	1.020	0.590	0.057	0.998	Mas
4	0.641	0.638	0.639	2.018	0.948	0.615	0.068	0.997	Mas
5	0.489	0.490	0.492	4.661	1.541	0.476	0.038	0.998	Tanduk
6	0.524	0.522	0.522	2.740	1.167	0.554	0.048	0.999	Tanduk

Based on Table 6, tests were performed on a sample of 6, images consisting of 2 banana varieties, namely Cavendish, Tanduk and Mas bananas. Next, feature extraction is performed on these images, the results of which are then processed using the SVM identification algorithm with a previously trained model. The implementation of the SVM identification algorithm is integrated into a GUI-based interface, and the identification is performed using Equation (3). Table 7 shows the results of the identification process using the GUI.

Table 7. Identification results

No	Types of Bananas	Y1	Y2	Y3	Results Identification
1	Cavendish	2.260348	0.877386	-0.24969	Cavendish
2	Cavendish	2.195209	1.02063	-0.19888	Cavendish
3	Mas	-0.10418	2.120787	0.965978	Mas
4	Mas	2.215098	1.170699	-0.24717	Cavendish
5	Tanduk	-0.30111	1.236235	2.291207	Tanduk
6	Tanduk	-0.25896	1.163889	2.238486	Tanduk

Overall, this analysis shows the effective ability of the SVM model in distinguishing banana types based on various prediction patterns. The variability of predicted values for each test reflects the model's capacity to detect differences in the dataset. Although some challenges were encountered in classifying tests with close predictive values, the model overall was able to recognize the typical characteristics of different types of banana well, resulting in accurate classifications in the majority of cases.

4. CONCLUSIONS

In conclusion, the comprehensive analysis conducted in this research utilizing microscopic camera testing has shed light on the distinct characteristics of various banana varieties, including the Cavendish, Mas, and Tanduk types. Notably, our investigation revealed that the color intensity of Cavendish bananas, as indicated by higher RGB values, set them apart from the other banana varieties, underscoring the significance of color attributes in distinguishing these types. To further refine our banana type identification methodology, we employed both RGB and Gray-Level Co-occurrence Matrix (GLCM) feature extraction techniques, which brought to the fore unique inherent patterns in each banana type. Importantly, the integration of GLCM features significantly improved the accuracy of our testing results, emphasizing the utility of GLCM features in elucidating intricate patterns and subtleties within banana images. Furthermore, the implementation of the Support Vector Machine (SVM) algorithm with well-tuned polynomial kernel parameters (C = 1000) yielded an impressive accuracy rate of 86.667% in identifying the extracted data, highlighting SVM's robustness as a reliable choice for banana type identification. Nevertheless, the Principal Component Analysis (PCA) method, when applied with two characteristics, did not enhance accuracy and resulted in diminished system performance. Despite this limitation, PCA proved valuable for visualizing SVM classification results and enhancing the interpretability of the outcomes. Additionally, the development of a user-friendly graphical interface (GUI) for emotion identification based on brain waves represents a significant advancement in practicality and accessibility, enabling individuals with varying technical backgrounds to use this technology effectively. These collective insights underscore the efficacy of a multifaceted approach in advancing banana type identification systems, with broader implications for the utilization of similar technologies in the fields of agriculture and image analysis. In summary, our research contributes to the understanding of banana type differentiation, primarily focusing on the visual attributes and image analysis techniques. The incorporation of GLCM features and SVM algorithms demonstrates the potential for enhancing accuracy in banana type identification, while the PCA method, though not successful in this context, aids in result interpretation. Furthermore, the development of a user-friendly GUI facilitates broader adoption of this technology, offering a practical solution for banana type classification. These findings hold promise for the broader utilization of analogous technologies in agricultural and image analysis domains, showcasing the multifaceted nature of the approach to improving identification systems in agriculture and related fields.

ACKNOWLEDGEMENT

The author would like to thank everyone who helped in this writing, all assistance in the form of accompanying in writing and so forth.

REFERENCES

[1] S. I. Hassan, M. M. Alam, U. Illahi, M. A. Al Ghamdi, S. H. Almotiri and M. M. Su'ud, "A Systematic Review on Monitoring and Advanced Control Strategies in Smart Agriculture," in *IEEE Access*, vol. 9, pp. 32517-32548, 2021, https://doi.org/10.1109/ACCESS.2021.3057865.

[2] M. A. Goralski and T. K. Tan, "*Artificial intelligence and sustainable development. The International Journal of Management Education*, vol. 18, no. 1, p. 100330, 2020, https://doi.org/10.1016/j.ijme.2019.100330.

[3] M. Mahfud, I. K. Yudiana, and S. Sariyanto, "HISTORY OF BANYUWANGI KALIKLATAK PLANTATION AND ITS IMPACT ON SURROUNDING COMMUNITIES," *International Journal of Educational Review, Law And Social Sciences (IJERLAS)*, vol. 3, no. 1, pp. 91–104, https://doi.org/10.54443/ijerlas.v3i1.492.

[4] R. Chase, M. Dita, B. A. Omondi, B. Ekesa, S. J. Zheng, and N. Roux, "Tapping into the wealth of local banana diversity for pest and disease resistance and consumer acceptability: a catalogue of the most popular cultivars in local markets across the world," *Acta Hortic*, vol. 1367, pp. 47-54, 2023, https://doi.org/10.17660/ActaHortic.2023.1367.5.

[5] M. E. Purbaya, D. Putra Rakhmadani, M. Puspa Arum and L. Zian Nasifah, "Comparison of Kernel Support Vector Machines in Conducting Sentiment Analysis Review of Buying Chips on the Shopee E- Marketplace in Indonesian," *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, pp. 435-440, 2022, https://doi.org/10.1109/ICIMCIS56303.2022.10017546.

[6] L. E. Walsh, *et al*. "Potential of urban green spaces for supporting horticultural production: A national scale analysis," *Environmental Research Letters*, vol. 17, no. 1, p. 014052, 2022, https://doi.org/10.1088/1748-9326/ac4730.

[7] V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S. D. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," *Artificial Intelligence in the Life Sciences*, vol. 1, p. 100010, 2021, https://doi.org/10.1016/j.ailsci.2021.100010.

[8] J. Zhuang *et al.*, "Assessment of External Properties for Identifying Banana Fruit Maturity Stages Using Optical Imaging Techniques," *Sensors*, vol. 19, no. 13, p. 2910, 2019, https://doi.org/10.3390/s19132910.

[9] J. I. Larada, G. J. Pojas, and L. V. V. Ferrer, "Postharvest classification of banana (Musa acuminata) using tier-based machine learning," *Postharvest biology and technology*, vol. 145, pp. 93-100, 2018, https://doi.org/10.1016/j.postharvbio.2018.06.004.

[10] M. Fiallos-Cárdenas *et al.*, "Bacterial Nanocellulose Derived from Banana Leaf Extract: Yield and Variation Factors," *Resources*, vol. 10, no. 12, p. 121, 2021, https://doi.org/10.3390/resources10120121.

[11] P. Marimo *et al.*, "Gender and Trait Preferences for Banana Cultivation and Use in Sub-Saharan Africa: A Literature Review," *Econ Bot*, vol. 74, pp, 226–241, 2020, https://doi.org/10.1007/s12231-020-09496-y.

[12] X. Chu *et al.*, "Green Banana Maturity Classification and Quality Evaluation Using Hyperspectral Imaging," *Agriculture*, vol. 12, no. 4, p. 530, 2022, https://doi.org/10.3390/agriculture12040530.

[13] A. H. Saputro, S. D. Juansyah and W. Handayani, "Banana (Musa sp.) maturity prediction system based on chlorophyll content using visible-NIR imaging," *2018 International Conference on Signals and Systems (ICSigSys)*, pp. 64-68, 2018, https://doi.org/10.1109/ICSIGSYS.2018.8373569.

[14] V. B. Raju, M. H. Imtiaz, and E. Sazonov, "Food Image Segmentation Using Multi-Modal Imaging Sensors with Color and Thermal Data," *Sensors*, vol. 23, no. 2, p. 560, 2023, https://doi.org/10.3390/s23020560.

[15] A. C. S. I. Mumthas, G. L. D. Wickramasinghe, and U. S. Gunasekera, "Effect of physical, chemical and biological extraction methods on the physical behaviour of banana pseudo-stem fibres: Based on fibres extracted from five common Sri Lankan cultivars," *Journal of Engineered Fibers and Fabrics*, vol. 14, p. 1558925019865697, 2019, https://doi.org/10.1177/1558925019865697.

[16] A. Derossi, R. Caporizzi, M. O. Oral, and C. Severini, "Analyzing the effects of 3D printing process per se on the microstructure and mechanical properties of cereal food products," *Innovative Food Science & Emerging Technologies*, vol. 66, 2020, https://doi.org/10.1016/j.ifset.2020.102531.

[17] G. M. Fernandes, W. R. Silva, D. N. Barreto, R. S. Lamarca, P. C. F. L. Gomes, J. F. da S Petruci, and A. D. Batista, "Novel approaches for colorimetric measurements in analytical chemistry–A review," *Analytica Chimica Acta*, vol. 1135, pp. 187-203, DOI: https://doi.org/10.1016/j.aca.2020.07.030.

[18] S. Marianingsih and F. Utaminingrum, "Comparison of Support Vector Machine Classifier and Naïve Bayes Classifier on Road Surface Type Classification," *2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, pp. 48-53, 2018, https://doi.org/10.1109/SIET.2018.8693113.

[19] S. Gayathri, A. K. Krishna, V. P. Gopi and P. Palanisamy, "Automated Binary and Multiclass Classification of Diabetic Retinopathy Using Haralick and Multiresolution Features," in *IEEE Access*, vol. 8, pp. 57497-57504, 2020, https://doi.org/10.1109/ACCESS.2020.2979753.

[20] R. A. Surya, A. Fadlil, A. Yudhana, "Ekstraksi Ciri Metode Gray Level Co-Occurrence Matrix (GLCM) dan Filter Gabor Untuk Klasifikasi Citra Batikk Pekalongan," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 02, no 2, 2017, https://doi.org/10.30591/jpit.v2i2.520.

[21] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6308-6325, 2020, https://doi.org/10.1109/JSTARS.2020.3026724.

[22] X. Feng et al., "Online State-of-Health Estimation for Li-Ion Battery Using Partial Charging Segment Based on Support Vector Machine," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 8583-8592, 2019, https://doi.org/10.1109/TVT.2019.2927120.

[23] A. Aruraj, A. Alex, M. S. P. Subathra, N. J. Sairamya, S. T. George and S. E. V. Ewards, "Detection and Classification of Diseases of Banana Plant Using Local Binary Pattern and Support Vector Machine," *2019 2nd International Conference on Signal Processing and Communication (ICSPC)*, pp. 231-235, 2019, https://doi.org/10.1109/ICSPC46172.2019.8976582.

[24] I. Ahmad, M. Basheri, M. J. Iqbal and A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," in *IEEE Access*, vol. 6, pp. 33789-33795, 2018, https://doi.org/10.1109/ACCESS.2018.2841987.

[25] A. Chakraborty et al., "Determining Protein–Protein Interaction Using Support Vector Machine: A Review," in *IEEE Access*, vol. 9, pp. 12473-12490, 2021, https://doi.org/10.1109/ACCESS.2021.3051006.

[26] Y. Shuai, Y. Zheng and H. Huang, "Hybrid Software Obsolescence Evaluation Model Based on PCA-SVM-GridSearchCV," *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, pp. 449-453, 2018, https://doi.org/10.1109/ICSESS.2018.8663753.

[27] A. Wibowo Haryanto, E. Kholid Mawardi and Muljono, "Influence of Word Normalization and Chi-Squared Feature Selection on Support Vector Machine (SVM) Text Classification," *2018 International Seminar on Application for Technology of Information and Communication*, pp. 229-233, 2018, https://doi.org/10.1109/ISEMANTIC.2018.8549748.

[28] N. Hameed, A. M. Shabut and M. A. Hossain, "Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine," *2018 12th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*, pp. 1-7, 2018, https://doi.org/10.1109/SKIMA.2018.8631525.

AUTHOR BIOGRAPHY

Aji Pamungkas is completed his undergraduate education at the Ahmad Dahlan University electrical engineering research program in 2023

Abdul Fadlil is received the B.Eng. in Physic – Electrical & Insrumentation and M.Eng. in Electrical Engineering from Universitas Gadjah Mada in 1992 and 2000, respectively. He also received his Ph.D. in Electrical Engineering from Universiti Teknologi Malaysia in 2006. He is currently as Associate Professor with the Electrical Engineering Department, Universitas Ahmad Dahlan (UAD), Yogyakarta, Indonesia. His current research interests include pattern recognition, image processing, and artificial intelligence