Emerging Information Science and Technology
Vol. 5, No. 1, (2024), pp. 24-29

Analysis of Cross Validation on Classification of Mangosteen Maturity Stages using Support Vector Machine

Indira Prabasari*, Afrizal Zuhri, Slamet Riyadi, Tony K. Hariadi, Nafi Ananda Utama

Universitas Muhammadiyah Yogyakarta, Jl. Brawijaya, Tamantirto, Kasihan, Bantul, Yogyakarta 55183, Indonesia
**Corresponding Author: i.prabasari@umy.ac.id*

Abstract

This study explores the efficacy of the Support Vector Machine (SVM) method in classifying mangosteen fruit images based on six ripeness levels. Employing SVM enables nonlinear data classification and simultaneous utilization of multiple feature extractions, resulting in enhanced accuracy. Analysis reveals that models integrating three feature extractions outperform those with only two. With ample training data and optimized parameters, SVM achieves detection accuracy exceeding 90%. However, algorithmic enhancements are necessary to compute RGB color index values for all pixels on mangosteen skin surfaces, possibly through circular-shaped windows approximating the fruit's contour. Moreover, comparative assessments of RGB color system calculations against alternative systems such as HSI are crucial for selecting the most suitable color model in alignment with human perception.

Keywords: *feature extraction, image classification, mangosteen fruit, RGB color system, Support Vector Machine*

1. Introduction

Mangosteen, known for its unique sweetness and sourness, fragrance, and elegant appearance, is a prestigious tropical fruit with monikers like "Queen of Fruit" and "Pearl of the Jungle." With abundant yield, it ranks second in Indonesia after bananas since 1995 [1]. The international demand for mangosteen has surged from 2014 to 2015, but its short shelf life in tropical conditions requires accurate ripeness classification. Current manual methods are labor-intensive and subjective, highlighting the need for automated techniques like the Support Vector Machine (SVM) to improve precision and efficiency [2].

The SVM algorithm, renowned for its capability to classify non-linear data and handle multiple feature extractions simultaneously, presents a promising avenue for mangosteen ripeness classification [3]. The study investigates the use of Support Vector Machine (SVM) to categorize mangosteen ripeness into six levels, aiming to improve the classification process and contribute to the advancement of automated fruit classification techniques by overcoming human-based limitations [4]. The study aimed to improve the use of digital imaging SVM to establish the non-destructive analysis on classification of mangosteen maturity stages, which is very important in agricultural practices.

2. Related Works

In research about fruit classification utilizing digital image processing, various approaches have been explored to classify fruits based on factors like form, size, skin color, and ripeness. Tomato classification using Mean Opinion Score (MOS) achieved highest accuracy rates for shape (86.17%), size (84.04%), and ripeness (80.85%) classifications [5]. Integration of RGB color feature extraction with physical factors to assess mangosteen ripeness yielded 37% accuracy with sugar content categorization and 63% accuracy with hardness-based classification, indicating moderate precision. The utilization of a fuzzy neural network for mangosteen ripeness identification faced limitations due to reliance on two input parameters only [6].

In contrast, a multi-SVM approach for tomato maturity classification achieved 86.7% accuracy with RGB mean enhancement and 76.7% with pixel index replacement enhancement [7]. Support Vector Machines are highly effective tools for multi-class classification tasks due to their resilience against irrelevant features, efficient handling of multiple feature extractions, and robust non-linear classification capabilities [8].

3. Method

The research methodology commences with data acquisition, encompassing the collection and transformation of primary data into processable formats, succeeded by labeling/grouping, preprocessing, feature extraction, and ultimately, data classification.

3.1. Data Acquisition

In the data acquisition stage, the first step involves inputting primary data in the form of photos, then counting the number of frames, iterating through frame scanning, writing frames into images, and ending with naming the images according to the frame sequence [9].

3.2. Labelling and Grouping

This study uses manual categorization of mangosteen fruit images into six ripeness classes, with each class receiving equal allocation. The data is then used for training and program accuracy evaluation. Data grouping involves folder reading, sequential file collection, total file counting, and division into six datasets corresponding to the manually determined ripeness classes [10].

3.3. Data Pre-processing

Image preprocessing involves cropping, cutting, and removing the central part of a mangosteen fruit image for accurate RGB value extraction. It ensures uniform image sizes and obtains RGB values for further processing. The feature extraction stage explains further image preprocessing interrelated steps [11].

3.4. Extraction of Image Features

3.4.1 Sum of RGB

The Sum of RGB feature extraction process involves calculating the sum of values in each R, G, and B component in mangosteen fruit images, which have different colours at different ripeness levels. This process starts with acquiring image data, extracting R, G, and B values, and then summarizing the values in each component [12].

3.4.2 Mean of RGB

The next feature extraction involves finding the mean value of each R, G, and B component from the input image. The process starts with data acquisition, followed by extracting R, G, and B values, summing them up, and counting the number of pixels in the image. Finally, the Mean of RGB is calculated by dividing each sum of R, G, and B values by the number of pixels in the image [13].

3.4.3 Standard deviation of RGB

The final feature extraction involves finding the standard deviation values of each R, G, and B component in the image. The process starts with image data acquisition, followed by extracting R, G, and B values. Finally, the standard deviation values of each R, G, and B component in the image are calculated [14].

3.5. Classification

Classification entails the practice of categorizing data by utilizing extracted Classification involves categorizing data using extracted values through training and testing. Support Vector Machine (SVM) models are created using different feature combinations. After training, the model is tested to identify misclassification and necessary modifications are made [15].

3.6. Calculation of Accuracy

Accuracy evaluation compares program-detected outcomes with manually classified outcomes, quantified as a percentage using a formula. It measures the program's achievement by dividing the number of correctly detected images by the total number of images and multiplying by 100%.

4. Results and Discussion

4.1. Data Acquisition

Data is captured using a digital camera positioned at a consistent distance of approximately 50 cm and with a resolution of 6000 x 4000. Each photo contains three mangosteen fruits of the same class to expedite data collection. Subsequently, cropping is performed for each mangosteen fruit, resizing them to 1024 x 1024 resolution.

4.2. Labelling and Grouping

Labeling and grouping are conducted post data acquisition, where image names are modified based on their classes. Mangosteen images in class 1 are labeled as sa, class 2 as sb, class 3 as sc, class 4 as sd, class 5 as se, and class 6 as sf. The naming convention follows the class designation followed by image sequence. An array representing groups is generated in the program, comprising a total of 96 entries, with 16 images for each class.

4.3. Pre-processing

Sampled images have a resolution of 1024 x 1024. To simplify the accurate retrieval of RGB values, cropping is performed on the central portion of mangosteen fruit images using Adobe Photoshop CS6, resulting in a size of 352 x 352 pixels. This process significantly impacts obtaining more accurate RGB values and ensures uniform image sizes.

4.4. Feature Extraction

The feature extraction process begins with Sum of RGB, which calculates the total values of the R, G, and B components in the image. Notably, Sum of Red extraction showed differing values among classes 1, 2, and 3, while classes 4, 5, and 6 exhibited identical values (Figure 1a). Likewise, Sum of Green extraction displayed variation between classes 1 and 2, with classes 3, 4, 5, and 6 sharing identical values (Figure 1b). The resulting data showcased unique vertical scatter patterns, denoted by different shapes and colors representing each class: red circles for class 1, yellow squares for class 2, green crosses for class 3, brown plus signs for class 4, blue stars for class 5, and purple triangles for class 6. Additionally, Sum of Blue extraction revealed shared values between classes 1 and 6, while classes 2, 3, 4, and 5 exhibited identical values (Figure 1c), reaffirming distinct class representations in the data. The result is similar with previous studies [2,3,4].

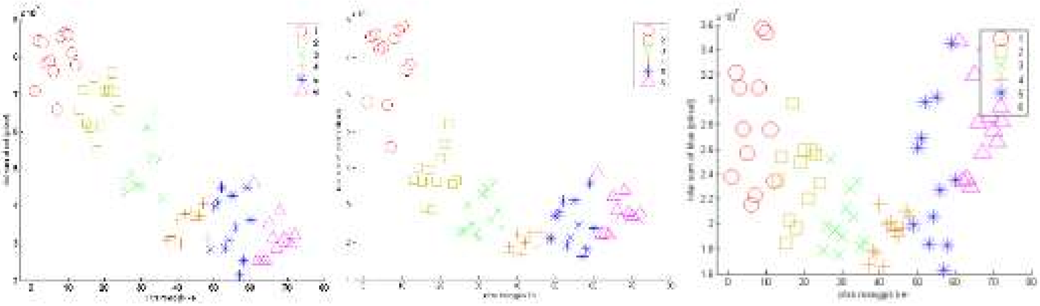

Figure 1 Sum of RGB (a) Red, (b) Green, (c) Blue

Mean of RGB represents the image based on the average values of each R, G, and B component. For Mean of Red feature extraction, differing values were observed for mangosteen image data in classes 1, 2, and 3, while classes 4, 5, and 6 shared identical values (Figure 2a). Similarly, Mean of Green feature extraction showed differing values for classes 1 and 2, whereas classes 3, 4, 5, and 6 shared the same values (Figure 2b). Regarding Mean of Blue feature extraction, all classes, including 1, 2, 3, 4, 5, and 6, exhibited identical values (Figure 3c). Each class is represented by distinct shapes and colors in the visualization.

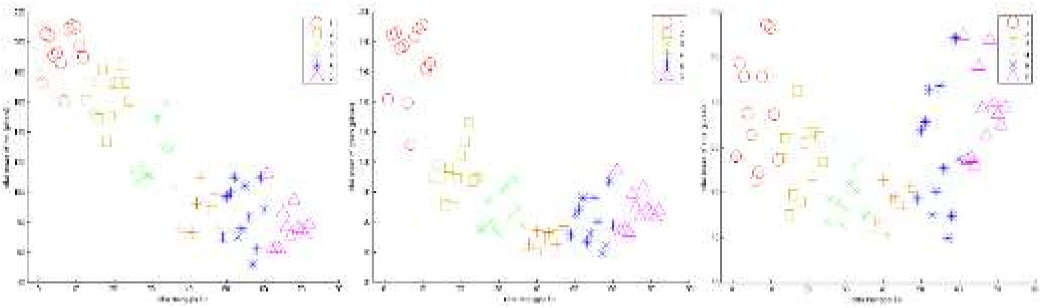

Figure 2 Mean of RGB (a) Red, (b) Green, (c) Blue

The next feature extraction is to find the standard deviation values of each R, G, and B component in the image. From the results of Standard Deviation of Red, Standard Deviation of Green, & Standard Deviation of Blue feature extraction, it is evident that all classes of mangosteen image data, including classes 1, 2, 3, 4, 5, and 6, share identical values (see Figures 1d, 1e, & 1f). Each class is represented by distinct shapes and colors in the visualization.

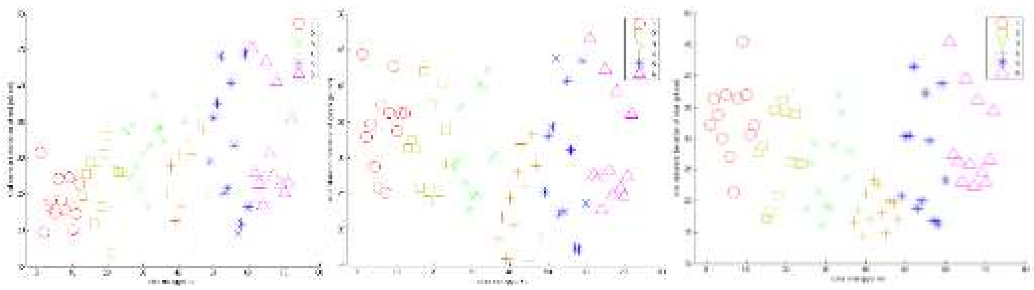

Figure 3 Standard Deviation RGB (a) Red, (b) Green, (c) Blue

4.5. Classification and Validation

The study used Support Vector Machine (SVM) for classification and 4-Fold Cross Validation for validation. The process involved training and testing with 96 images per fold. Results showed that SVM models with Sum of RGB & Standard Deviation of RGB and Mean of RGB & Standard Deviation of RGB had low accuracy due to color similarity in manually selected mangosteen fruits. However, all SVM models showed improved accuracy in Fold 2, with some misclassifications persisting. Fold 3 showed further accuracy improvement due to precise input image composition. Fold 4 experienced a slight decrease in accuracy due to color similarity in training data. Overall, models using three feature extractions achieved better accuracy, and these are consistent with previous researches [2,3,4]

5. Conclusion

The Support Vector Machine method classifies mangosteen fruit images with six ripeness levels using nonlinear data classification and multiple feature extractions. Models with three feature extractions have superior accuracy. With sufficient training data and proper parameters, it achieves detection accuracy exceeding 90%. However, algorithm refinements are needed to calculate RGB color index values for all pixels on the mangosteen skin surface and compare RGB color system calculations with other color systems like HSI. Future works needed to establish the program into a device that is applicable in agricultural practices.

Acknowledgments

The authors acknowledge the support of Universitas Muhammadiyah Yogyakarta in conducting this research on fruit classification using digital image processing techniques.

References

[1] A. Thammastitkul and T. Klayjumlang, "Mangosteen Quality Grading for Export Markets Using Digital Image Processing Techniques," *Int J Adv Sci Eng Inf Technol*, vol. 11, no. 6, pp. 2452–2458, Dec. 2021, doi: 10.18517/IJASEIT.11.6.14003.

[2] S. Riyadi *et al.*, "Classification of Mangosteen Surface Quality Based on Image Processing Using Support Vector Machine," *International Journal of Integrated Engineering*, vol. 13, no. 5, pp. 288–294, 2021, doi: 10.30880/IJIE.2021.13.05.029.

[3] M. Furqan, M. Ikhsan, and A. Dalimunthe, "Detection of Ripeness of Manggosteen Fruit Using Hsv Color Space Transformation Method," *International Journal of Information System & Technology Akreditasi*, vol. 5, no. 4, pp. 513–517, 2021.

[4] I. Prabasari, "Comparison of destructive and non destructive method in maturity index of Garcia mangostana," *Planta Tropika: Journal of Agro Science*, vol. 6, no. 2, 2018, doi: 10.18196/pt.2018.086.100-105.

[5] H. Sabrol and K. Satish, "Tomato plant disease classification in digital images using classification tree," *International Conference on Communication and Signal Processing, ICCSP 2016*, pp. 1242–1246, Nov. 2016, doi: 10.1109/ICCSP.2016.7754351.

[6] S. Tamakuwala, J. Lavji, and R. Patel, "Quality Identification Of Tomato Using Image Processing Techniques," 2018.

[7] S. KUMARI Behera, A. Mahapatra, A. Rath, and P. kumar Sethy, "Classification & Grading of Tomatoes using Image Processing Techniques," 2019. [Online]. Available: https://www.researchgate.net/publication/332653429

[8] H. Naseera, R. Mahendran, G. Mary Amirtha Sagayee, and A. Santhakumaran, "Grading of Tomatoes into Different Classes Using Digital Image Processing", Accessed: May 16, 2024. [Online]. Available: www.ijert.org

[9] T. H. Trinh, X. T. Bui, T. H. Tran, H. H. C. Nguyen, and K. D. Ninh, "Mangosteen Fruit Detection Using Improved Faster R-CNN," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 148, pp. 366–375, 2022, doi: 10.1007/978-3-031-15063-0_35.

[10] L. M. rifatul Azizah, S. F. Umayah, S. Riyadi, C. Damarjati, and N. A. Utama, "Deep learning implementation using convolutional neural network in mangosteen surface defect

detection," *Proceedings - 7th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2017*, vol. 2017-November, pp. 242–246, Jul. 2017, doi: 10.1109/ICCSCE.2017.8284412.

[11] A. S. Agung, A. F. D. SR, M. S. Hersyam, A. B. Kaswar, and D. D. Andayani, "Classification Of Tomato Quality Based On Color Features And Skin Characteristics Using Image Processing Based Artificial Neural Network," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 5, pp. 1021–1032, Oct. 2023, doi: 10.52436/1.JUTIF.2023.4.5.730.

[12] "Classification of mangosteen fruit disease detection system based on existing convolutional neural network", Accessed: May 16, 2024. [Online]. Available: https://www.researchgate.net/publication/364094660

[13] S. Riyadi, A. M. A. Ratiwi, C. Damarjati, T. K. Hariadi, I. Prabasari, and N. A. Utama, "Classification of Mangosteen Surface Quality Using Principal Component Analysis," *Emerging Information Science and Technology*, vol. 1, no. 1, 2020, doi: 10.18196/eist.115.

[14] A. D. Prabandani, N. Yudistira, A. Raisa, and K. Nisa, "Large Scale Image Classification of Exotic Fruits in Indonesia Using Transfer Learning Method with MobileNet Model," pp. 675–685, Apr. 2023, doi: 10.2991/978-94-6463-140-1_68.

[15] J. Siswantoro, H. Arwoko, and M. Widiasri, "Indonesian fruits classification from image using MPEG-7 descriptors and ensemble of simple classifiers," 2020, doi: 10.1111/jfpe.13414.