Prediction of Physico-Chemical Characteristics in Batu Tangerine 55 Based on Reflectance-Fluorescence Computer Vision

Safitri D. A. Ariani[1], Inggit K. Maharsih[1], and Dimas F. A. Riza[1]

[1]Bioprocess Engineering Program, Department of Biosystems Engineering, Faculty of Agricultural Technology, Brawijaya University, Malang, Indonesia

Corresponding author: Dimas F. A. Riza (e-mail: dimasfirmanda@ub.ac.id).

ABSTRACT Oranges (Citrus sp.) are one of the most abundant agricultural commodities in Indonesia. One of the popular local citruses is Batu Tangerine 55. Harvesting tangerines begins 252 days after the flowers bloom. Conventionally, we still determine the level of maturity by observing the color, shape, and hardness. The results of manual grouping tend to be subjective and less accurate. Destructive testing could be carried out and provide objective results; however, it would require sampling and damaging the fruits. Computer vision could be used to evaluate the maturity level of the fruit non-destructively. Dual imaging computer vision, i.e., reflectance-fluorescence mode, could be used to enhance the accuracy of the prediction. This study aims to develop a classification model and predict the physico-chemical characteristics of Batu Tangerine 55. Destructive testing is still being carried out to determine the value of TPT, the degree of acidity, and the firmness of the fruit. Non-destructive testing was carried out to obtain reflectance and fluorescence images. Once we obtain the destructive and non-destructive data, we will incorporate them into the classification and prediction models. The machine learning method for maturity classification uses three models, namely KNN, SVM, and Random Forest. The best results on the reflectance data (RGB) SVM model resulted in an accuracy of 1 for training data and 0.97 for testing data. The maturity parameter prediction method uses the PLS method. The best results for the predicted Brix/Acidity ratio R2 parameter are 0.81 and RMSE 3.4.

KEYWORDS Brix/Acid ratio, Machine Learning, PLS, Tangerine

I. INTRODUCTION

Indonesia stands as one of the nations blessed with abundant agricultural commodities, with oranges holding a significant place among them. Citrus fruits, particularly tangerines (*Citrus reticulata* Blanco), enjoy immense popularity among the Indonesian populace. According to data from the Central Statistics Agency (2021), Indonesia produces approximately 2,401,064 tons of tangerines annually, with East Java Province emerging as the largest contributor, accounting for 822,260 tons per year. Among the varieties under cultivation, Batu Tangerine 55 has garnered attention for its superior quality, characterized by sweet, slightly sour, and refreshing fruit flesh [1]. These tangerines typically reach harvesting maturity 252 days after flowering, necessitating careful post-harvest handling to minimize product damage during marketing.

Assessing fruit maturity holds pivotal importance in the marketing of citrus fruits in Indonesia, significantly influencing consumer preferences. Presently, the method of determining maturity levels remains predominantly conventional. Typically, farmers gauge the maturity and physical attributes of citrus fruits by visually inspecting factors such as color, shape, and hardness [2]. The process of categorizing fruit ripeness is predominantly manual, leading to subjective and often less precise results [3].

An alternative method for determining fruit maturity involves destructive testing, which entails assessing the total dissolved solids (TDS) and acidity levels in citrus fruits [4]. However, this approach has drawbacks, as it can physically damage the fruit. Hence, there is a pressing need to predict and classify tangerine maturity without causing physical harm, utilizing non-destructive testing methods, such as digital imaging [5]. Nevertheless, the accuracy of classification using conventional computer vision systems remains inadequate, necessitating the exploration of improved methodologies to enhance model accuracy [6]. Previous studies have employed computer vision in dual reflectance-fluorescence mode, enhancing predictive models by incorporating additional features from fluorescence imagery [7]. However, these studies often rely on deep learning models, which demand substantial amounts of data. Alternatively, simpler machine learning models could achieve comparable performance with smaller datasets.

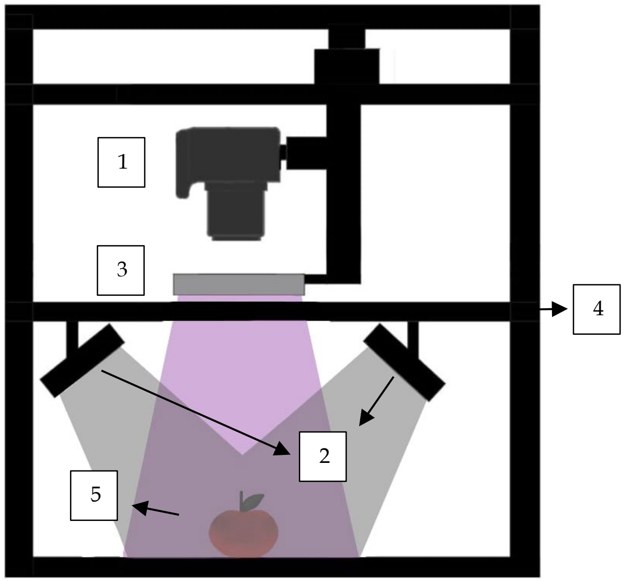
This study endeavors to develop and refine a predictive model for the physicochemical characteristics of Batu Tangerine 55 across three maturity levels. Both destructive and non-destructive testing methodologies are employed to facilitate sorting and grading processes. Destructive testing determines TDS, acidity, and fruit hardness, while non-destructive testing employs reflectance-fluorescence dual-vision computer systems. Subsequently, the data obtained from both methods are integrated into the classification and prediction model. Machine learning techniques, including KNN, SVM, and Random Forest, are applied for maturity classification. The comparative analysis of these models will reveal the most effective approach for accurately classifying Batu Tangerine 55 maturity levels.

II. RESEARCH METHODS

The experiment encompassed both non-destructive and destructive testing methodologies. Non-destructive testing involved capturing images using a mini studio setup, as depicted in Figure 1, utilizing a 700D DSLR camera. Samples comprised 55 Batu Tangerines across three ripeness levels: raw, semi-ripe, and ripe, sourced from the Research Institute of Citrus and Subtropical Plants. Image acquisition encompassed top and bottom views, facilitated by two light sources—LEDs for reflectance imaging and UV for fluorescence imaging.

Following image data collection, three non-destructive parameters were measured. First, fruit hardness was assessed using penetrometers at three equatorial points. Subsequently, the sweetness and acidity of fruit juice were gauged using the ATAGO PAL-BX|ACID 101 Brix Acidity meter.

Predictive modeling employed Partial Least-Squares Regression (PLSR) to determine the maturity parameters of Batu Tangerine 55. Widely utilized across various domains including bioinformatics, food research, medicine, and pharmacology [8]. PLSR combines principal component analysis with multiple regression analysis to predict or analyze dependent variables with multiple independent variables. This technique is particularly effective for datasets with high collinearity and numerous variables [9]. PLSR facilitated the extraction of several parameters to evaluate method accuracy, including slope values, R2 offsets, and RMSE values. R2 values indicate the proximity between real values and predictions, with higher values suggesting a stronger model relationship, ideally approaching 1. Conversely, RMSE reflects the model's error, with smaller values indicating better model performance. For predictability assessment, R2 values ≥ 0.9 are considered favorable, while values ≤ 0.64 are indicative of poorer predictability [10].



Explanation:
1. Camera
2. LED Light
3. UV LED
4. Frame
5. Sample

Figure 1. Non-destructive data retrieval scheme

Additionally, the study sought to identify an optimal machine learning model for classifying the maturity of Batu Tangerine 55. Various machine learning techniques were employed in the modeling process, including k-nearest Neighbor (k-NN), Support Vector Machine (SVM), and Random Forest. The k-NN method operates by identifying the k-nearest samples in the training dataset that closely resemble the object under consideration in the testing dataset. SVM, on the other hand, performs classification by determining an optimal hyperplane separator between different classes, particularly when the data can be linearly separated [11]. Moreover, Random Forest employs an ensemble learning approach by constructing multiple decision trees during training on the dataset. In classification tasks, it aggregates the decision outputs from individual trees, typically through a majority voting mechanism, to arrive at the final classification decision. These methods offer distinct advantages and are suitable for various types of data and classification tasks. By comparing their performance, the study aimed to identify the most effective approach for accurately classifying the maturity levels of Batu Tangerine 55 [12].

III. RESULTS AND DISCUSSION

The non-destructive testing yielded two types of images, namely reflectance images and fluorescence images, as illustrated in Figure 2. A total of 240 images were acquired for each type of light source used. These image datasets encompassed observations across three distinct maturity levels.

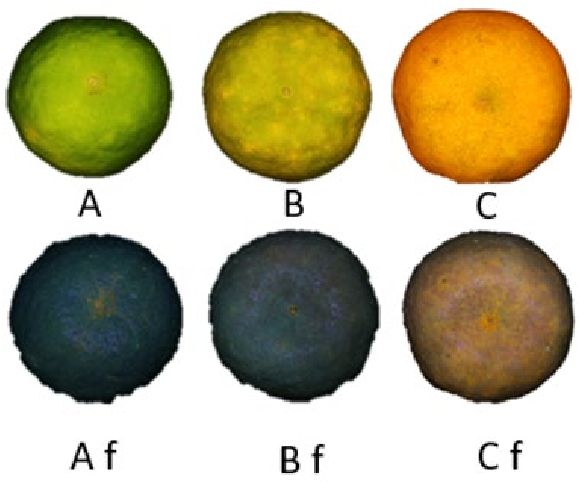


Figure 2. The acquisition result of the image of tangerine batu 55

Distinct color characteristics are observed in the images of Batu 55 tangerines across different maturity levels. The variations in color correspond to changes in the degree of maturity in the tangerine peel, attributed to the conversion of chloroplasts into chromoplasts and subsequent accumulation of carotenoids [13]. The discoloration of the fruit skin is primarily influenced by pigments such as flavonoids. Flavonoid compounds exhibit fluorescence when exposed to UV light, typically appearing yellow or blue. Notably, the use of a UV lamp as a light source results in relatively darker images compared to those obtained under white LED illumination. Nevertheless, the reflectance images capture certain colors that may not be discernible under white LED lighting. In destructive testing, measurements of firmness, Brix, and acidity values were obtained to further characterize the tangerines.

In Figure 3, it is apparent that the Brix values for partially ripe and fully ripe maturity levels exhibit insignificant differences. This suggests that tangerines classified as partially ripe are already suitable for harvesting. According to standards set by the Indonesian National Standard (SNI, 2009), tangerines are considered ready for harvest when their total dissolved solids reach 8°Brix.

Figure 4 illustrates a notable trend where the acidity value decreases as Batu Tangerine 55 progress in ripeness. This phenomenon can be attributed to the conversion of organic acids into simpler sugars, such as fructose and glucose, during the fruit ripening process [7].

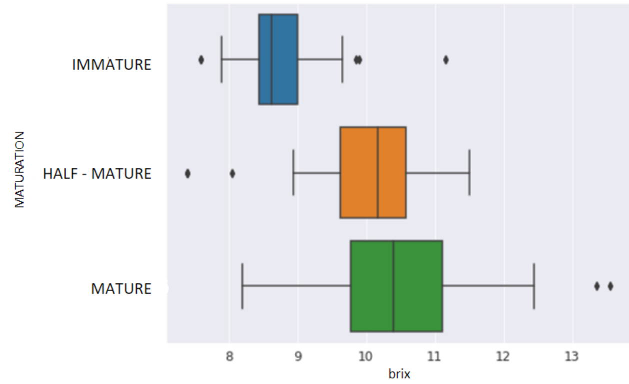


Figure 3. ***Box-plot* brix**

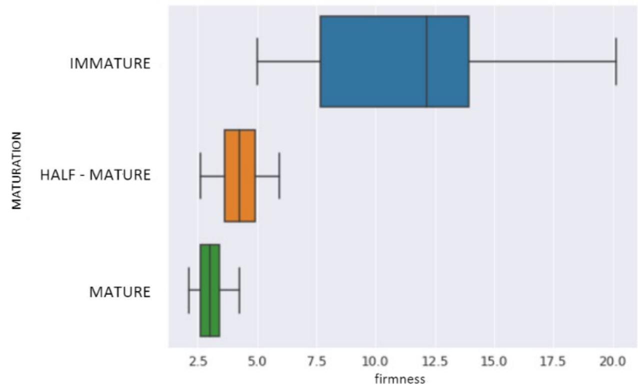


Figure 4. ***Box-plot acidity***

Observing Figure 5 reveals a clear trend: as Batu Tangerine 55 reach higher levels of maturity, their firmness values decrease. This phenomenon can be elucidated by the fact that, according to [2], the softening of oranges is a hallmark of ripening, resulting from a reduction in fruit hardness. This softening process stems from changes in the chemical composition and structural integrity of carbohydrate cell walls within fruit tissues.

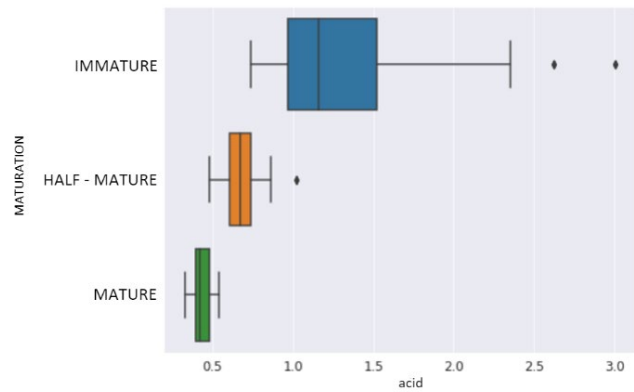


Figure 5. Box-plot firmness

Both destructive and non-destructive data were collected to create datasets for data processing. The classification task utilized three machine learning models: K-Nearest

Neighbors (KNN), Support Vector Machine (SVM), and Random Forest, to classify three levels of maturity of Batu Tangerine 55. The data variations included RGB, UV, and combined datasets.The classification process involved two phases: the training phase, where the model was constructed, and the testing phase, where the model's accuracy was evaluated using separate data. The results are summarized in table 1. In the RGB dataset, the SVM model outperformed the other models, achieving a training accuracy of 100% and a testing accuracy of 97%. Similarly, the Random Forest model performed best in the UV dataset, with a training accuracy of 100% and a testing accuracy of 92.5%.

When considering the combined RGB and UV dataset, the SVM model once again exhibited superior performance, attaining a training accuracy of 100% and a testing accuracy of 95%. Thus, the recommended model for use with RGB image data is SVM, given its robust performance across both training and testing phases.

TABLE I
MACHINE LEARNING CLARIFICATION DATA

	Model	Scaling	*Training Accuracy*	*Test Accuracy*
RGB	KNN	MinMax	1.0	0.97
	KNN	None	0.95	0.94
	Random Forest	MinMax	1.0	0.95
	Random Forest	None	1.0	0.94
	SVM (kernel: Linear)	**MinMax**	**1.0**	**0.97**
	SVM (kernel: Linear)	**None**	**1.0**	**0.97**
UV	KNN	MinMax	0.95	0.94
	KNN	None	0.91	0.95
	Random Forest	MinMax	1.0	0.93
	Random Forest	None	1.0	0.92
	SVM (kernel: Linear)	MinMax	0.95	0.97
	SVM (kernel: Linear)	None	1.0	0.88
All	KNN	MinMax	0.95	0.92
	KNN	None	0.97	0.95
	Random Forest	MinMax	1.0	0.89
	Random Forest	None	1.0	0.89
	SVM (kernel: Linear)	MinMax	1.0	0.96
	SVM (kernel: Linear)	None	1.0	0.92

TABLE II
PREDICTION RESULTS WITH PLS COMBINED DATA

Physicochemical parameters	Factor	R^2 Calibration	RMSEC	R^2 Cross-validation	RMSEC V	R^2 Prediction	RMSEP	RPD
Firmness	10	0.71	2.34	0.69	2.44	0.63	2.76	1.81
Brix	10	0.43	0.93	0.39	0.96	0.48	0.88	1.28
Acid	10	0.60	0.31	0.56	0.32	0.49	0.34	1.52
B/A	10	**0.79**	**3.58**	**0.77**	**3.70**	**0.81**	**3.48**	**2.12**

The prediction model for the maturity parameter of Batu Tangerine 55 utilized Partial Least Squares (PLS) analysis, implemented using Python software. The PLS analysis comprised two stages: calibration and validation, with 2/3 of the total 480 data points used for model training. Cross-validation, an integral part of PLS analysis, was employed to assess the accuracy of the calibration model. Subsequently, 1/3 of the total data was utilized to predict the maturity of other tangerine fruits. Table 2 presents the prediction results of the mature parameters based on physicochemical parameters derived from combined reflectance and fluorescence data. The brix/acidity ratio yielded the highest R2 value of 0.81. Statistical parameters used for model evaluation include the coefficient of determination (R2), root-mean-square error of calibration (RMSEC), and root-mean-square error of cross-validation (RMSECV). A small difference between RMSEC and RMSECV indicates model stability, with larger differences suggesting that the calibration set model does not adequately represent the validation set [14]. The accuracy achieved in this study was not superior to that of deep learning models developed in previous studies [7]. However, the machine learning model utilized herein offers advantages in terms of ease of training and implementation compared to deep learning models.

IV. CONCLUSION

A machine learning model has been developed using a reflectance-fluorescence image dataset to classify three levels of maturity of Batu Tangerine 55 fruit, employing three models: KNN, SVM, and Random Forest. The SVM model utilizing reflectance (RGB) data yielded the most favorable results, achieving a training accuracy of 100% and a testing accuracy of 97%. For the prediction of the maturity of Batu Tangerine 55 fruit using the PLS method, the brix/acidity ratio emerged as a significant parameter compared to others. Notably, the combined feature set produced the highest prediction accuracy, with an R2 value of 0.81 and an RMSE of 3.4.

ACKNOWLEDGMENTS

The author expresses gratitude to Jessica Yolanda, Abdullah Sa'id, Abdul Mulukul Ikrom, and Salsabil Lazuardi Nugroho from the Department of Biosystems Engineering, Universitas Brawijaya, for their valuable assistance during the experiment. This research was supported by funding from the 2022 Indonesian Collaborative Research through Universitas Brawijaya Rector, under decree number 6672 of 2022.

AUTHORS CONTRIBUTION

Safitri Diah Ayu Ariani: Investigation, Analysis, Visualization, Preparation of Original Drafts;
Inggit Krishna Maharsih: Validation, Review Writing and Editing;
Dimas Firmanda Al Riza: Project Administration, Software, Resources, Validation, Review Writing and Editing;

COPYRIGHT

REFERENCES

[1] A. Ashari, E. Nurcahyani, H. I. Qudus, and Zulkifli, "Analisis Kandungan Prolin Planlet Jeruk Keprok Batu 55 (*Citrus reticula blanco* Var. Crenatifolia) Setelah Diinduksi Larutan Atonik Dalam Kondisi Cekaman Kekeringan Secara In Vitro," *Analit: Analytical and Environmental Chemistry*, vol. 3, no. 1, pp. 69-78, 2018.

[2] I. Ifmalinda, K. Fahmy, and E. Fitria, "Prediction of Siam Gunung Omeh Citrus Fruit (Citrus Nobilis Var Microcarpa) Maturity Using Image Processing," Jurnal Keteknikan Pertanian, vol. 6, no. 3, pp. 335–342, Dec. 01, 2018. Doi: 10.19028/jtep.06.3.335-342.

[3] I Kadek Riastana, Ni Komang Alit Astiari, and Ni Putu Anom Sulistiawati, "Kualitas Buah Jeruk Siam (Citrus nobillis var microcarva L) Selama Penyimpanan Pada Berbagai Tingkat Kematangan Buah," GA, vol. 24, no. 1, pp. 22-28, Apr. 2020.

[4] M. G. Lieka, R. Poerwanto, and D. Efendi, "Aplikasi Ethephon dan Stiker Pascapanen untuk Perbaikan Kualitas Buah Jeruk Siam Garut (Citrus nobilis Lour)," Comm. Horticulturae Journal, vol. 2, no. 2, p. 1, Dec. 11, 2018. Doi: 10.29244/chj.2.2.1-10.

[5] R. K. Haba and K. C. Pelangi, "Pengelompokan Buah Jeruk menggunakan Naïve Bayes dan Gray Level Co-occurrence Matrix," ILKOM Jurnal Ilmiah, vol. 12, no. 1, pp. 17–24, Apr. 26, 2020. Doi: 10.33096/ilkom.v12i1.494.17-24.

[6] A. Zakiyyah , Z. Hanif , D. W. Indriani , Z. Iqbal , R. Damayanti , D. F. Al Riza , "Characterization and Classification of Citrus reticulata var. Keprok Batu 55 Using Image Processing and Artificial Intelligence," Universal Journal of Agricultural Research, Vol. 10, No. 4, pp. 397 – 404, 2022. DOI: 10.13189/ujar.2022.100409.

[7] D.F. Al Riza, A. M. Ikrom, A. A. Tulsi, Darmanto, Y. Hendrawan, "Mandarin orange (*Citrus reticulata Blanco* cv. Batu 55) ripeness parameters prediction using combined reflectance-fluorescence images and deep convolutional neural network (DCNN) regression model" Scientia Horticulturae, Volume 331, 1 May 2024, 113089, pp. 1-10, 2024

[8] A.H. Wigena "Regresi kuadrat terkecil parsial multi respon untuk statistical downscaling", Indonesian Journal of Statistics and Its Applications, Vol. 16, No. 2, pp. 12–15. 2011.

[9] V. Vijayakumar, Y. Ampatzidis, L. Costa. "Tree-level citrus yield prediction utilizing ground and aerial machine vision and machine learning", Smart Agricultural Technology, Vol. 3, February 2023, 100077, pp. 1-11. 2023.

[10] I.U. Hidayah, B. Kuswandi, L. Wulandari. "Deteksi Kemurnian Air Zamzam Menggunakan Metode Spektrofotometri Near Infra Red (NIR) dan Kemometrik" Pustaka Kesehatan, Vol. 2, No. 3, p. 439-444, sep. 2014.

[11] Y. Amrozi, D. Yuliati., A. Susilo, R. Ramadhan. "Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM" Jurnal Sisfokom, Vol. 11, No. 3, pp. 394-399, 2022.

[12] P. Rosyani, Saprudin, R. Amalia. "Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO)" Jurnal Sistem dan Teknologi Informasi, Vol. 9, No. 2, pp. 132-134, 2021

[13] A. Rahmawati and W. D. R. Putri, "Karakteristik Ekstrak Kulit Jeruk Bali Menggunakan Metode Ekstraksi Ultrasonik (Kajian Perbandingan Lama Blansing Dan Ekstraksi)," JPA, vol. 1, no. 1, pp. 26–35, Oct. 2013.

[14] A. Putri Ana, Y. A. Purwanto, and S. Widodo, "Prediksi Indeks Panen Jambu 'Kristal' secara Non Destruktif Menggunakan Portable Near Infrared Spectrometer," *Jurnal Keteknikan Pertanian*, vol. 9, no. 3, pp. 103–110, Dec. 23, 2021. doi: 10.19028/jtep.09.3.103-110.