AFSSAAE

Advances in Food Science, Sustainable Agriculture and Agroindustrial Engineering 2023, 6(3), 243-252
ISSN 2622-5921

ORIGINAL RESEARCH Open Access

Development of tomato maturity level prediction model based on portable visible spectrometer and machine learning

Dimas Firmanda Al Riza*, Nughi Arie Nugraha and Darmanto

Department of Agricultural and Biosystem Engineering, Faculty of Agricultural Technology, Universitas Brawijaya, Malang, Indonesia

KEYWORDS

Acid
Ripeness
Sensor
Tomato
Total dissolved solids

ABSTRACT

A tomato is classified as a fruit, which level of maturity is determined by its color. Upon distribution, tomatoes require sorting based on their ripeness level. Generally making improvements done conventionally with the human eye. This method has the disadvantage that the results are subjective. One way that can be used to measure the ripeness level of tomatoes is using a spectroscopic sensor. Spectroscopic sensors can predict the level of ripeness and its contents automatically. This study uses machine learning to create a model to classify ripeness level and predict firmness, total dissolved solids (TDS), and total acid in tomatoes. This study used tomatoes with 3 categories of maturity. Tomatoes were tested non-destructively, namely measuring firmness, total dissolved solids content, and total acid. The data obtained were processed using the Partial Least Square Regression method to predict firmness, TDS, and total acid, while the maturity level used the Naive Bayes method. The data processing results to predict the level of maturity using Naive Bayes obtained a success rate of 100%. While for the predictions of firmness, TDS, and total acid had R^2 training and R^2 testing, namely 0.685 and 0.678, 0.534 and 0.521, and 0.352 and 0.349, respectively.

Introduction

Tomatoes are a plant that grows in tropical regions and is widely cultivated in both highland and lowland areas (Agussabti et al., 2019). Tomatoes are classified as a staple food and an agricultural product of Indonesian farmers. In Indonesia, tomato production has increased from 2015 to 2020, reaching 1,084,993 tons in 2020 (Haikal et al., 2022). In daily life, tomatoes are a versatile food ingredient, as they can be consumed fresh, used as a seasoning for cooking, or processed as a raw material for food industries such as fruit juice and tomato sauce (Aydoğan-Coşkun et al., 2018).

Tomatoes are a fruit that have a specific level of ripeness. The process of measuring the ripeness level of tomatoes is crucial to be conducted. Measuring the ripeness level is done to separate the tomatoes that will be distributed to faraway places to prevent spoilage or for specific processing purposes (Abdelhamid et al., 2020). The maturity level of tomatoes is categorized into six maturity stages: Green stage, Breakers stage, Turning stage, Pink stage, Light Red stage, and Red stage (Choi et al., 2022). The surface color of the tomatoes is analyzed to determine the level of ripeness. This is generally done manually using the human eye directly (Ifmalinda et al., 2022). The separation of tomatoes' ripeness levels is generally done conventionally (i.e., directly using human eye), which may subject to people condition, causing inconsistent results (Opeña and Yusiong, 2017). Conventional ripeness measurement methods can only differentiate tomato's ripeness levels based on skin color, but not the firmness level, total dissolved solids (TDS), and total acid content (Zhu et al., 2022). A way to determine the firmness, TDS, and total acid is through destructive methods. However, this method is ineffective because it may damage the tomato, making it unsellable. Therefore, a technology is needed to predict the tomato ripeness level, TDS, and titrated total acid without damaging the tomato fruit itself.

*Corresponding author
E-mail address: dimasfirmanda@ub.ac.id

Received on 13 May 2023, Revised on 6 September 2023, Accepted on 14 September 2023

Nowadays, the rapid development of time accompanied by technological advancements allows for automatic measurement of ripeness levels and their content without damaging the fruit (non-destructive). Previous research has been conducted on predicting ripeness levels and vitamin C content in tomatoes using the RGB TCS3200 sensor, and an accuracy of 93% was obtained (Sandra et al., 2020). In addition, research has also been conducted on the prediction of the water and soluble solids content of *manalagi* apples using near-infrared spectroscopy and multivariate analysis. The results showed that water content prediction displayed R^2 calibration, R^2 prediction, RMSEC, RMSEP, and RPD of 0.81, 0.61, 0.009, 0.020, and 1.62. While soluble solids content prediction displayed R^2 calibration, R^2 prediction, RMSEC, RMSEP, and RPD of 0.79, 0.85, 0.474, 0.420, and 2.69 (Kusumiyati et al., 2021). Based on previous research, development was carried out using the AS7262 Spectral Sensor instead of the RGB sensor and laboratory-level spectrometry. This sensor works by measuring 6 (six) visible light colors ranging from 450 nm, 500 nm, 550 nm, 570 nm, 600 nm, and 650 nm, where each of these wavelengths represents the colors purple, blue, green, yellow, orange, and red, respectively (Suder et al., 2021). This study aimed to create a model that can classify the ripeness level and predict firmness, TDS, and total acid in tomatoes using machine learning.

Research Methods

Tools and materials

The tools used in this study include the Spectral Sensor AS7262 Visible functioned to capture spectra data with a wavelength of 450-650 nm, and an Arduino Uno microcontroller functioned as a controller for the sensor in capturing spectra data. The Arduino IDE software is a medium for creating programs in capturing and reading spectra data, while a laptop functioned as a medium for connecting Arduino IDE with the microcontroller. The test setup can be seen in Figure 1. The AMTAST GY-4 penetrometer was used to measure the firmness of tomato fruit flesh, and the ATAGO Brix-Acidity was used to measure TDS and total acid in tomato fruit. The materials utilized include Gustavi tomato variety (as the testing sample) and aquadest (for cleaning experimental tools and making tomato juices).

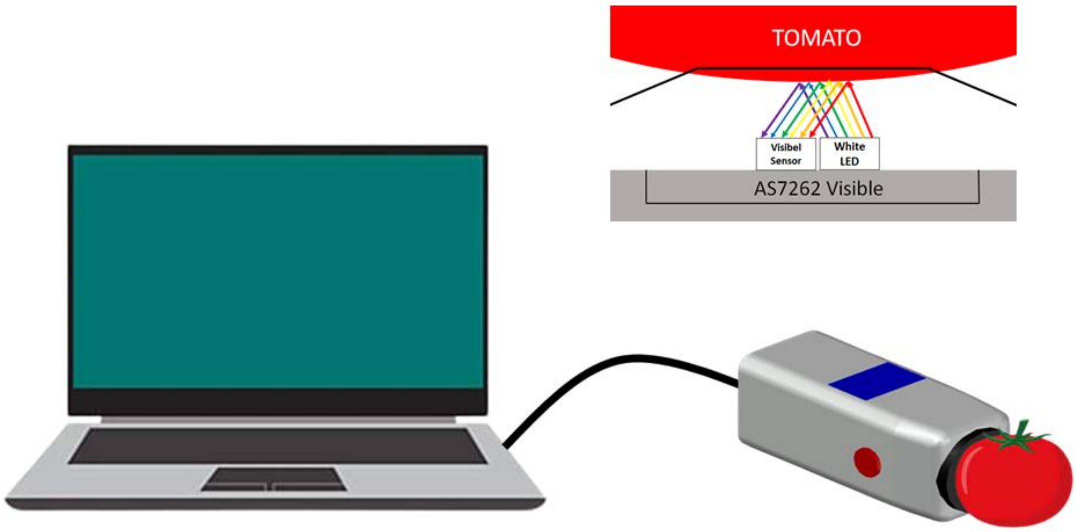


Figure 1. Illustration set up

Category	
1	
2	
3	

Figure 2. Maturity level of tomatoes

Research design

The experiment used visible spectroscopy as an independent variable with 6 (six) channel outputs. The dependent variables measured consist of firmness, total dissolved solids, total acid, and maturity level of Gustavi tomato fruits, divided into 3 (three) maturity categories. The study will measure the output value of the visible spectroscopy sensor to observe tomatoes' content and maturity level. The research data were processed using Partial Least Square Regression, Naive Bayes, and Support Vector Machine analysis methods using Google Collaboratory.

Research stages

a. Sample Preparation

Tomato fruits (Gustavi variety) from Batu City, East Java, Indonesia, were used as sample. The fruits were harvested and categorized into 3 (three) groups. In category 1, the tomatoes consist of those with ripeness indices 1 and 2, while category 2 includes tomatoes with ripeness indices 3 and 4. Category 3 comprises tomatoes with ripeness indices 5 and 6. The ripeness level of the tomato fruit, based on the ripeness indices, can be seen in Figure 2. The ripeness level of tomatoes can be determined by observing color changes on the tomato skin, with transitions from full green to full red. Tomatoes have six levels of ripeness based on their color, starting from mature green, breakers, turning, pink, light red, and red. The harvested tomato samples were immediately subjected to testing (i.e., on the same day). Prior testing, the tomatoes were cleaned and labeled for ease of data collection during the testing process.

b. Visible spectrum data acquisition

After preparing the tomato, visible spectrum data was taken three times. Before measuring the visible spectrum, the AS7262 Visible Spectral Sensor, integrated with the Arduino Uno microcontroller, was connected to a laptop to run the Arduino IDE program for reading. Once it confirmed that it could read the visible spectrum, the tomato fruit was placed close to the sensor, and the reading was taken. The indication of spectrum reading is the LED light turns on and the spectrum value appears on the Arduino IDE software on the laptop. In the process of taking visible spectrum data from the tomato fruit, the sensor works by the LED light from the sensor hitting the tomato fruit, being reflected, and then received by the visible spectrum at wavelengths of 450-650 nm (or 6 channels), namely violet, blue, green, yellow, orange, and red. This spectrum data were read by the Arduino IDE as the sensor output value, then stored as a database in Microsoft Excel.

c. Destructive testing

Destructive testing on tomatoes consists of firmness, TDS, and total acid. Firmness testing was carried out using an AMTAST GY-4 penetrometer. The device's measurement includes N, kg, and lb, with a maximum firmness measurement limit of 20 kg. A flat probe with a length of 2 cm and a diameter of 8 mm is used for the firmness testing of tomatoes. The firmness measurement result is automatically read at the highest value with a unit of kgF. Then, the TDS and total acid testing were performed using an ATAGO Brix-Acidity meter. Before the TDS measurement, the tomato was crushed and taken its juice. Then, the juice was poured onto the sample lens of the device, and the start button was pressed. The TDS value automatically appeared on the ATAGO Brix-Acidity meter screen with the unit of °Brix. After measuring the TDS, total acid measurement was carried out. Unlike the TDS measurement, the scale selection to be used on the device must be determined and adjusted according to the type of fruit used. Prior total acid measurement, the juice sample was diluted with aquadest at a 1:50 ratio (on a weight basis). The diluted juice was poured onto the sample lens. Then, the total acid was measured. The total acid value appeared automatically on the ATAGO Brix-Acidity meter screen as acid unit.

d. Building a prediction model

The output data from the sensor consisting of 6 channels and the test results (i.e., firmness, TDS, and total acid) were compiled into a data set to build a prediction model. The sensor output data becomes the independent variable, while the destructive test results become the dependent variable. Out of 90 tomato samples obtained, 80% of the data was used as training data and 20% was used as testing data. The prediction method used was Partial Least Square (PLS) Regression. the PLS Regression is one of the methods used to solve multiple linear regression problems when specific problems occur in the data (such as small sample size, missing data, and multicollinearity), and can be applied to different types of data. This method is considered powerful because it can be applied to all data scales, does not require many assumptions, and does not require a large sample size (Larbi et al., 2020).

e. Building a classification model

In addition to building a predictive model, a classification model was also developed. In forming the classification model, 90 tomato fruit samples were divided into 3 (three) categories of ripeness levels. These ripeness levels were become the classification labels for the tomato fruit. The data was split into 80% for training and 20% for testing. The algorithms used for classifying the ripeness levels of tomato fruit were Naive Bayes and Support Vector Machine (SVM). Naive Bayes is a classification method that uses probability and statistics to predict future probabilities based on past experiences. Naive Bayes has several advantages, such as being simple and easy to apply, requiring only a small amount of training data, and could handle both quantitative and discrete data (Achmad and Girsang, 2020). SVM is a method used for prediction in classification and regression cases. The basic principle of SVM is linear classification, meaning the classification case can be linearly separated. However, SVM has also been developed to work on non-linear problems by incorporating the concept of kernels in high-dimensional feature space. By using kernels, SVM can map data into a higher feature space, thus data can be separated more effectively (Saikin et al., 2021).

Results and Discussion

Visible spectroscopy

The average value of visible spectrum acquisition results on tomatoes can be seen in Table 1. In tomatoes with maturity category 1, the highest spectrum value was at 550 nm wavelength or in the green color. Then, in maturity category 2, the highest spectrum value was at 600 nm wavelength or in the orange color. Furthermore, in maturity category 3, the highest spectrum value was at 650 nm wavelength or in the red color. Based on Figure 3, it can be seen that there are differences in the visible spectrum of tomatoes in each maturity category. Therefore, it can be observed that each maturity category has different values at each wavelength.

Firmness, total dissolved solids, and total acid

The difference in maturity levels in tomatoes caused changes in visible spectrum firmness, TDS, and total acid. As shown in Table 2, the firmness values of tomatoes with maturity level 1 ranged from 3.73 to 7.36 KgF, while for maturity level 2 and 3, the firmness values ranged from 1.26 to 4.25 KgF and 1.44 to 4.25 KgF, respectively. The TDS values for maturity level 1, 2, and 3 were ranged from 2.80 to 4.90 °Brix, 3.60 to 5.50 °Brix, and 3.60 to 6.30 °Brix. Furthermore, the total acid values for maturity level 1, 2, and 3 were in the range of 0.41 - 0.80 acid, 0.34 - 0.50 acid, and 0.33 - 0.61 acid. Figure 4 shows that the tomato's firmness decreased as its maturity level increased. One of the signs that a tomato has entered the ripening phase is its softer texture or lower firmness value. This occurs due to the degradation of water-insoluble pectin (i.e., protopectin) and transformed into water-soluble pectin, causing a decrease in the cohesive strength of the cell walls that bind one cell wall to another (Roiyana et al., 2012).

Figure 5 shows that the TDS values in tomatoes increase with their ripeness level. Thus, as tomatoes become ripen, their TDS values also increase. The increase may be caused by the hydrolysis of carbohydrates into glucose and fructose during ripening (Angelia, 2017).

Figure 6 shows that the total acid content decreased as the maturity level of the tomato increased. The total acid value of the tomato was related to the TDS, where an increase in TDS is parallel to a reduction in total acid value, and vice versa. The changes in total acid value in maturity categories 2 and 3 were not significant. This is because, at the early stage of maturity, tomatoes have high levels of organic acids, resulting in high total acid values. Previous study also reported the highest total acid content in tomatoes at the early stage of maturity with no significant change in further maturity levels (Novita et al., 2012).

Table 1. Average value of spectroscopy data acquisition results

Category	Wavelength (nm)					
	450	500	550	570	600	650
1	3746	4214	6367	6468	5996	4512
2	3605	3607	5831	6453	7550	6780
3	3012	2699	4103	4491	6271	7493

Table 2. Data on the results of tomato destructive testing

Destructive Test	Maturity Category	Minimal	Maximal	Average
Firmness (KgF)	1	3.73	7.36	5.33
	2	1.26	4.25	2.59
	3	1.44	4.25	2.48
Total Dissolved Solids (°Brix)	1	2.80	4.90	4.18
	2	3.60	5.50	4.53
	3	3.60	6.30	5.28
Total Acid (acid)	1	0.41	0.80	0.54
	2	0.34	0.50	0.42
	3	0.33	0.61	0.45

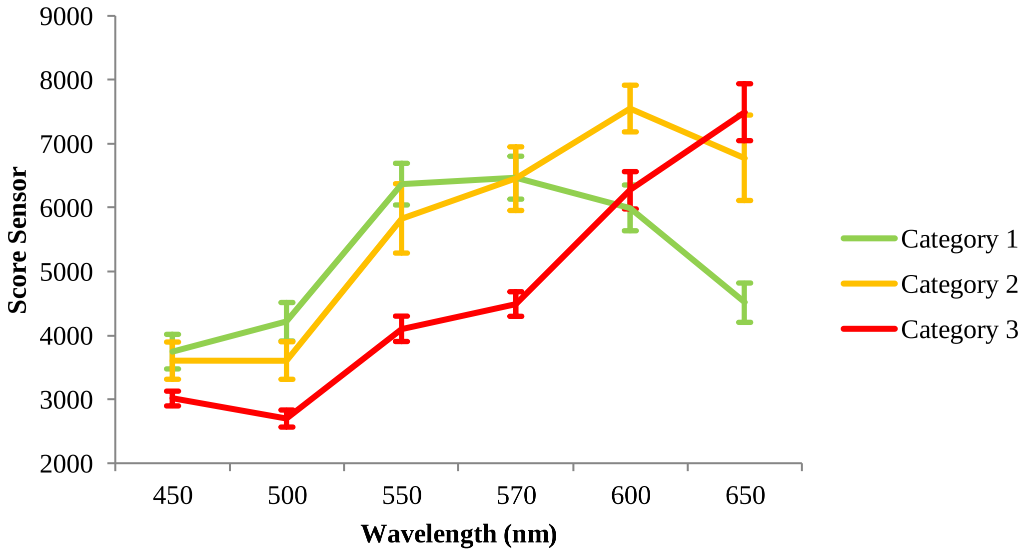


Figure 3. Visible spectroscopy sensor reading result

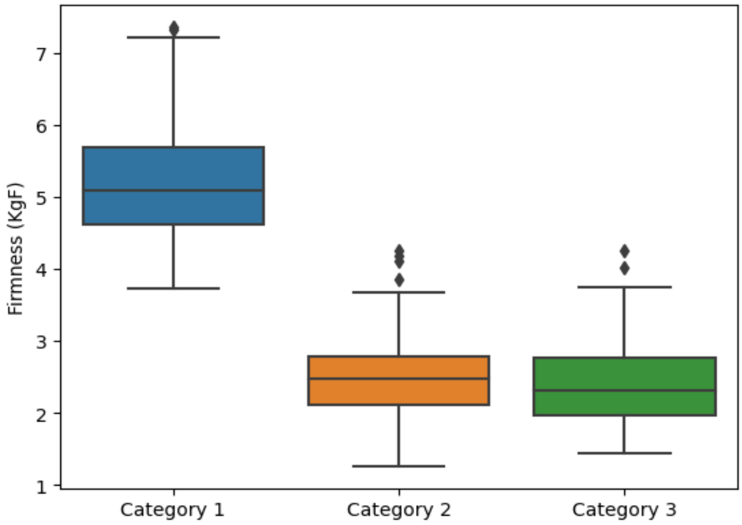


Figure 4. Firmness distribution of the sample

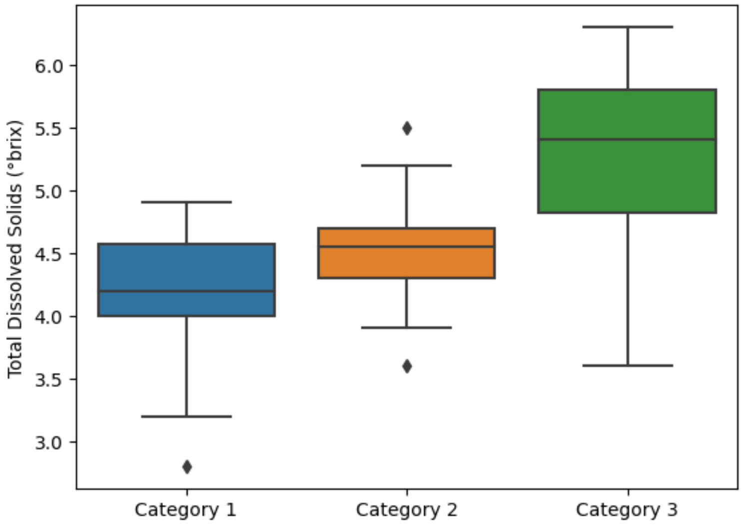


Figure 5. Total Dissolved Solids distribution of the sample

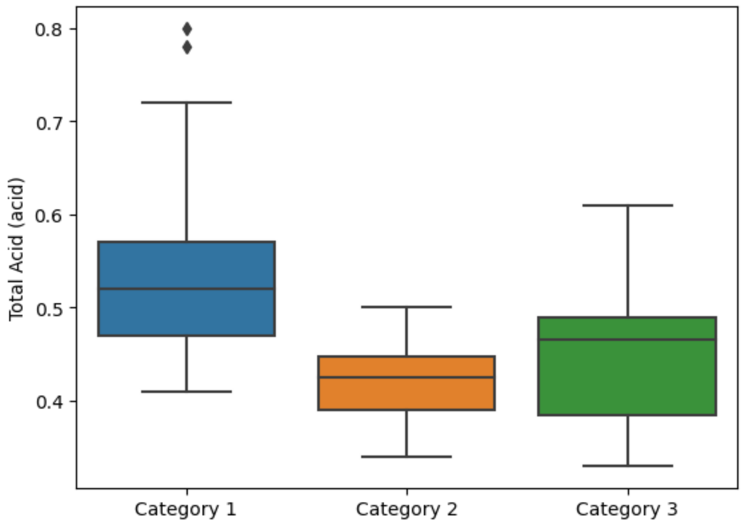


Figure 6. Total Acidity distribution of the sample

Table 3. Result training and testing prediction model using partial least square regression

	Training		**Testing**	
	$\mathbf{R^2}$	**RMSE**	$\mathbf{R^2}$	**RMSE**
Firmness	0.685	0.803	0.678	1.009
Total Dissolved Solid	0.534	0.451	0.521	0.548
Total Acid	0.352	0.071	0.349	0.078

Predictive model results

Table 3 shows the R^2 and Root Mean Square Error (RMSE) values used to determine the quality of a regression prediction model. The coefficient of determination, or R^2, is a value that indicates the strength of the relationship between variables. Generally, R^2 measures a model's ability to predict the dependent variable. The range of R^2 values is from 0 to 1, where a higher R^2 value indicates a better ability of the model to explain the variation in the dependent variable. Conversely, if the R^2 value is low, it means that the ability of the independent variable to explain the variation in the dependent variable is limited (Lu, 2004). The RMSE is a common measure to assess the difference between predicted values generated by a model or estimator and the actual observed values. A low RMSE value indicates that the variation in the predicted values generated by the forecasting model is close to the actual observed values. Therefore, RMSE can help evaluate the accuracy of a forecasting model in predicting values that match the actual observed values (Khalid et al., 2020). The results indicate that the tomatoes firmness obtained a training R^2 value of 0.685 with an RMSE of 0.803, and a testing R^2 value of 0.678 with an RMSE of 1.009, respectively. The training and testing graphs of the firmness prediction model can be seen in the Figure 7.

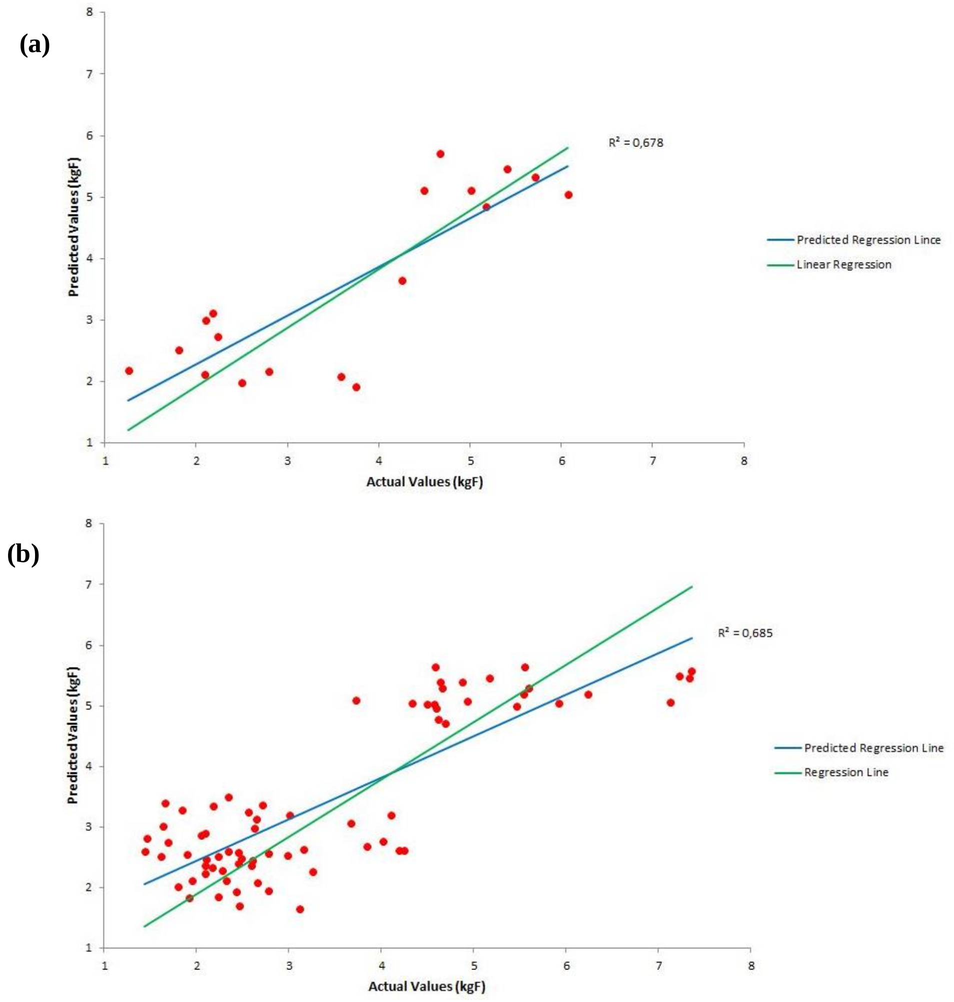


Figure 7. Graph training (a) and testing (b) prediction firmness

Furthermore, the prediction model for TDS in tomatoes obtained a training R^2 value of 0.534 with an RMSE of 0.451 and a testing R^2 value of 0.521 with an RMSE of 0.548. The training and testing graphs of the total dissolved solids prediction model is shown in the Figure 8. Additionally, the prediction model for total acid in tomatoes obtained a training R^2 value of 0.352 with an RMSE of 0.071 and a testing R^2 value of 0.349 with an RMSE of 0.078, as can be seen in Figure 9. From the obtained R^2 values, the model's performance can be inferred by examining the R^2 values of both training and testing datasets. During training and testing, If the R^2 value obtained was greater than 0.7, it falls into the strong category. A value between 0.5-0.7 is categorized as moderate, 0.3-0.5 is categorized as weak, and less than 0.3 indicates very weak (Moore et al., 2013). Therefore, the prediction models for the tomatoes' firmness and TDS have moderate accuracy , while weak accuracy for total acid.

Classification model results

The classification model results, obtained using the Naive Bayes and SVM methods, were then tested on a testing dataset comprising 18 tomato fruit samples across categories 1, 2, and 3, each containing 6 samples. Out of the 18 samples tested using the Naive Bayes and SVM algorithms, all samples yielded readings consistent with their original data. These findings confirm that the Naive Bayes and SVM algorithms for classifying tomato fruit maturity level exhibit an accuracy rate of 100%.

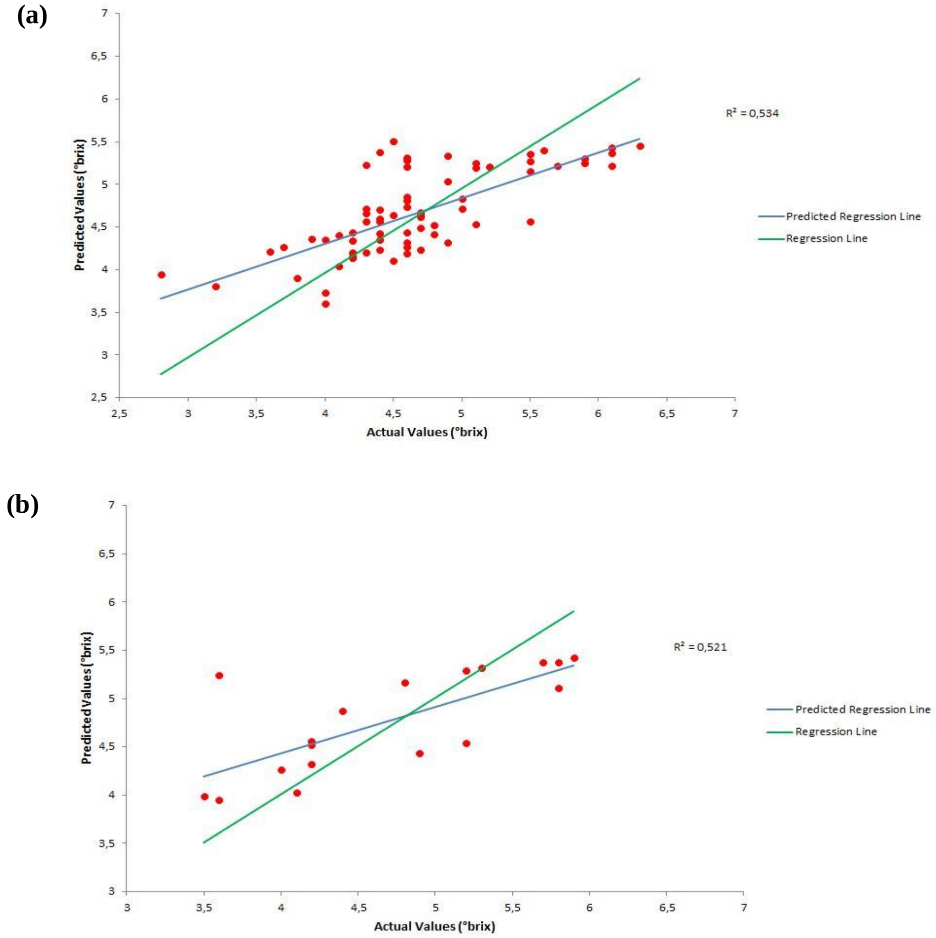
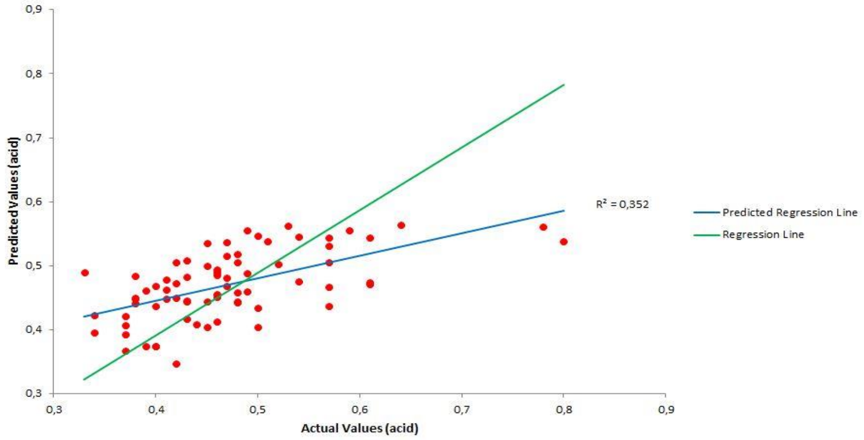


Figure 8. Graph training (a) and testing (b) total dissolved solids

(a)



(b)

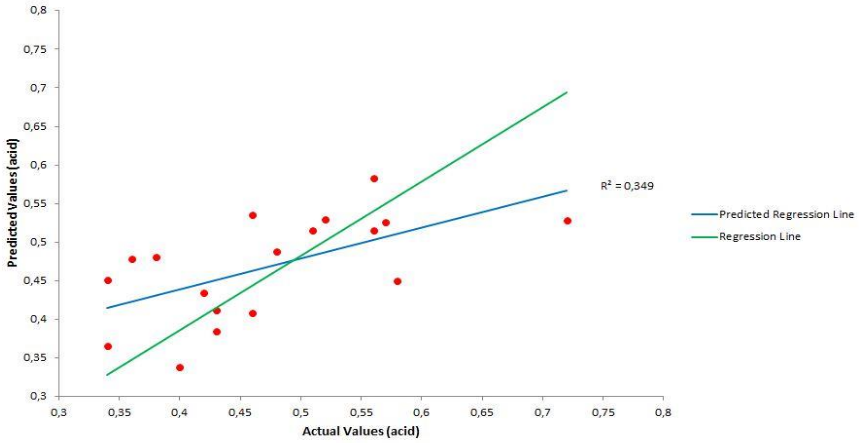


Figure 9. Prediction graph (a) firmness, (b) total dissolved solids, (c) total acid

Conclusions

In this study, visible spectroscopy can be effectively utilized to classify the ripeness levels of tomato fruits. Both the Naive Bayes and SVM algorithms demonstrated the ability to classify tomato ripeness levels with an accuracy of 100%. In predicting the firmness of tomatoes, the visible spectroscopy sensor with the PLS Regression prediction model has strong accuracy with training and testing R^2 values of 0.685 and 0.678, respectively. Meanwhile, predicting the TDS content has moderate accuracy with training and testing R^2 values of 0.534 and 0.521. Predicting total acid has weak accuracy with training and testing R^2 values of 0.352 and 0.349. Future study is required to improve the accuracy of predicting TDS and total acid values using Near-Infrared spectroscopy sensors. Near-Infrared spectroscopy sensors have longer wavelengths than visible spectroscopy, allowing to measure the interaction between light and material in the fruit, such as chemical contents (i.e., sugar, acid, protein, and fat).

Acknowledgement

This work was supported by Penelitian PNBP (OPF) Faculty of Agricutlural Technology, Universitas Brawijaya Year of 2023 (Contract number 1074/UN10.F10.06/TU/2023).

Declarations

Conflict of interests The authors declare no competing interests.

References

Abdelhamid, M. A., Sudnik, Y. A., Alshinayyin, H. J., and Shaaban, F. (2020) 'Chlorophyll fluorescence for classification of tomato fruits by their maturity stage', *E3S Web of Conferences*, 193, pp. 1–6

Achmad, R., and Girsang, A. S. (2020) 'Implementation of naive bayes classifier algorithm in classification of civil servants', *Journal of Physics: Conference Series*, 1485, pp. 1–7

Agussabti, Rahmaddiansyahromano, and Afridila, S. (2019) 'Aceh tomato farmers and the application of tomato cultivation technology', *IOP Conference Series: Earth and Environmental Science*, 365, pp. 1-9

Angelia, I. O. (2017) 'Kandungan ph, total asam tertitrasi, padatan terlarut dan vitamin c pada beberapa komoditas hortikultura (pH content, total acidified acid, dissolved solids and vitamin c in some horticultural commodities)', *Journal of Agritech Science*, 1, pp. 68–74

Aydoğan-Coşkun, B., Çoklar, H., and Akbulut, M. (2018) 'Effect of sunlight on fatty acid composition of tomato seed oil stored in different colored bottles', *GIDA the Journal of Food*, 43, pp. 122–128

Choi, H. G., Park, D. Y., and Kang, N. J. (2022) 'The fruit proteome response to the ripening stages in three tomato genotypes', *Plants*, 11(4), pp. 1-18

Haikal, R., Wahyudi, M., Pujiastuti, L., and Solikhun, S. (2022) 'Fletcher-reeves algorithm for predicting the quantity of production tomato plants in Indonesia', *International Journal of Mechanical Computational and Manufacturing Research*, 11, pp. 109–120

Ifmalinda, Andasuryani, and Rasinta, I. (2022) 'Classification of tomato *(Lycoersicon Esculentum Miil*) ripeness levels based on HSV color using digital image processing', *IOP Conference Series: Earth and Environmental Science*, 1116, pp. 1–9

Khalid, S., Goldenberg, M., Grantcharov, T., Taati, B., and Rudzicz, F. (2020) 'Evaluation of deep learning models for identifying surgical actions and measuring performance', *JAMA Network Open*, 3(3), pp. 1-10

Kusumiyati, K., Hadiwijaya, Y., Suhandy, D., and Munawar, A. A. (2021) 'Prediction of water content and soluble solids content of "*manalagi*" apples using near infrared spectroscopy', *IOP Conference Series: Earth and Environmental Science*, 922, pp. 1–5

Larbi, M., Trinchero, R., Canavero, F. G., Besnier, P., and Swaminathan, M. (2020) 'Analysis of parameter variability in integrated devices by partial least squares regression', *SPI 2020 - 24th IEEE Workshop On Signal and Power Integrity, Proceedings*, pp. 9–12

Lu, Z. Q. J. (2004) 'The elements of statistical learning: data mining, inference, and prediction', *Journal of the Royal Statistical Society*, 173, pp. 693–694

Moore, DS., Notz, WI., and Flinger, MA. (2013) *The Best Practice of Statistics*. United States: W. H. Freeman and Company.

Novita, M., Satriana, Martunis, Rohaya, S., and Hasmarita, E. (2012) 'Pengaruh pelapisan kitosan terhadap sifat fisik dan kimia tomat segar (*Lycopersicum pyriforme*) pada berbagai tingkat kematangan (Effects of chitosan coating on physico-chemical characteristics of fresh tomatoes (*Lycopersicum Pyriforme*) in different maturity stages)', *Jurnal Teknologi dan Industri Pertanian Indonesia*, 4(3), pp. 1–32 [In Indonesian]

Opeña, H. J. G., and Yusiong, J. P. T. (2017) 'Automated tomato maturity grading using ABC-trained artificial neural networks', *Malaysian Journal of Computer Science*, 30(1), pp. 12–26

Roiyana, M., Izzati, M., and Prihastanti, E. (2012) 'Potensi dan efisiensi senyawa hidrokoloid nabati sebagai bahan penunda pematangan buah (Potential and efficiency of vegetable hydrocolloid compounds as delaying fruit ripening agents)', *Buletin Anatomi dan Fisiologi*, 20(2), pp. 40–50 [In Indonesian]

Saikin, Fadli, S., and Ashari, M. (2021) 'Optimization of support vector machine method using feature selection to improve classification results', 4(1), pp. 22–27

Sandra, Damayanti, R., Hendrawan, Y., Susilo, B., and Oktavia, S. (2020) 'Prediction of tomatoes maturity using TCS3200 color sensor', *IOP Conference Series: Earth and Environmental Science*, 475, pp. 1–9

Suder, J., Podbucki, K., Marciniak, T., and Dąbrowski, A. (2021) 'Spectrum sensors for detecting type of airport lamps in a light photometry system', *Opto-Electronics Review*. 29, pp. 133–140

Zhu, Y., Gu, Q., Zhao, Y., Wan, H., Wang, R., Zhang, X., and Cheng, Y. (2022) 'Quantitative extraction and evaluation of tomato fruit phenotypes based on image recognition', *Frontiers in Plant Science*, 13, pp. 1–12