Pertanika J. Trop. Agric. Sci. 49 (5): 2565 - 2580 (2026)

PERTANIKA JOURNALS

TROPICAL AGRICULTURAL SCIENCE

Journal homepage: http://www.pertanika.upm.edu.my/

Classification and Prediction of Chilli Powder (*Capsicum Annuum* L.) Colour Based on Near-Infrared Spectra and Random Forest

Nafis Khuriyati[1]*, Anggoro Cahyo Sukartiko[1], Ririn Nur Alfiani[1], Tyasto Prima Ahmadi[2], and Avicenna Nur Kasih[3]

[1]Department of Agro-industrial Technology, Faculty of Agricultural Technology, Universitas Gadjah Mada, 55281 Yogyakarta, Indonesia
[2]Department of Industrial Engineering, Faculty of Engineering and Food Security, Universitas Indonesia Mandiri, 35592 Lampung, Indonesia
[3]Department of Agroindustrial Technology, Politeknik Negeri Tanah Laut, A. Yani Street, Tanah Laut, 70815 South Kalimantan, Indonesia

ABSTRACT

Colour measurements play a crucial role in the chilli industry for establishing colour standards. Limited research has been conducted on predicting colour characteristics using Near Infrared (NIR) measurements, yet none of them has been applied to chilli powder. This study aimed to classify and predict the colour intensity of chilli powder using chemometric models based on NIR spectroscopy and the random forest algorithm. This study used 84 chilli powder samples from four colour groups. NIR spectra were measured with NIRFlex N-500 at 1000 to 2500 nm. Three pretreatments were applied: Multiplicative Scatter Correction (MSC), first derivative Savitzky-Golay, and detrending. Random forest was used for classification and prediction of L*, a*, and b*. Data were split into 75% training and 25% testing, evaluated with accuracy score and root mean square error (RMSE). The machine learning method of random forest has been successfully applied for classifying and predicting the colour of chilli powder based on NIR spectroscopy. Classification of four colour groups achieved the best accuracy of 0.9524 with first derivative Savitzky-Golay, while raw spectra and MSC reached 0.7619. L* was predicted most accurately with MSC (RMSE 2.68), meanwhile a* and b* were predicted most accurately with first derivative Savitzky-Golay (RMSE 2.86 and 3.0821). The findings indicate that the most effective pretreatment varied according to colour index.

Keywords: Chilli powder, near infrared, random forest, colour, Savitzky-Golay

ARTICLE INFO

Article history:
Received: 19 August 2025
Accepted: 25 October 2026
Published: 30 October 2026

DOI: https://doi.org/10.47836/pjtas.49.5.23

E-mail addresses:
nafis.khuriyati@ugm.ac.id (Nafis Khuriyati)
cahyos@ugm.ac.id (Anggoro Cahyo Sukartiko)
ririn.nur.alfiani@mail.ugm.ac.id (Ririn Nur Alfiani)
tyasto@uimandiri.ac.id (Tyasto Prima Ahmadi)
avicennanurkasih@politala.ac.id (Avicenna Nur Kasih)
* Corresponding author

e-ISSN: 2231-8542

© Universiti Putra Malaysia Press

INTRODUCTION

Colour measurements play a crucial role in the chilli industry for establishing colour standards, whether chilli powder is used as a seasoning or as a food coating (Aşkın & Yazici, 2021). In addition to providing a distinctive spicy taste, chilli powder also provides an attractive colour to food. The colour of chilli powder can vary from green, yellow, orange, red, brown, aniline/white, and purple depending on the type and level of maturity (Swammy, 2023). The chemical content of chilli peppers, such as carotenoids, is responsible for the chilli colour. The colouration is attributed to more than 30 types of carotenoids; for example, capsanthin and capsorubin contribute to the red colour, while xanthophyll and carotene are responsible for the yellow hue (Batiha et al., 2020).

A decrease in colour values is anticipated in dried samples and powders, with reductions across all colour measurements. The colour values were assessed using L*, a*, and b* parameters. The decrease in colour values in powder form is due to the red pepper's colour darkening, likely from changes in carotenoids and the formation of browning compounds. Generally, L* values of dried pepper samples were significantly higher ($p < 0.05$) than fresh chilli (Jiménez et al., 2021). This increase is likely attributed to the varying ways in which light is reflected on wet and porous surfaces, as observed by Nowacka and Wedzik (2016). Meanwhile, the results for the a* coordinate values indicate a statistically significant decrease ($p < 0.05$) in the dried samples compared to the fresh samples before drying samples (Jiménez et al., 2021). This decline can be attributed to the oxidation and degradation of carotenoids, specifically capsanthin and capsorubin, in red pepper (Konishi et al., 2019), likely caused by the effects of heat (Liu et al., 2014).

As a result, colour plays an important role in food perception, where the brighter the red colour in food, the greater a person's arousal and appetite (Foroni et al., 2016). Existing methods for evaluating chilli colour attributes are inefficient, destructive, and often subject to inconsistency and variability. There is a need to develop fast, non-destructive, and accurate techniques. NIR spectroscopy holds promise as a foundation for such techniques, offering advantages in speed, ease of use, and minimal interference.

Over decades, NIR has been successfully applied considerably in chemical analysis of the content of chilli peppers, such as capsaicin (Jiang et al., 2018), antioxidant activity (Khuriyati et al., 2022), vitamin C and total carotene (Khuriyati et al., 2023). Those chemical contents corresponded with the colour of chilli powder. There are limited studies available on predicting colour characteristics using NIR measurements (Elmasry et al., 2012; Gabriëls et al., 2020; Wang et al., 1999). Theanjumpol et al. (2023) investigated the use of visible and near-infrared spectroscopy to predict ASTA-colour and pungency in dry red chilli powder using partial least squares (PLS) models. These studies demonstrate the potential of spectroscopic techniques for evaluating chilli quality.

NIR operates on the principle that different chemical bonds absorb or emit light at specific wavenumbers, with prominent absorption bands corresponding to hydrogen-based functional groups like C-H, C-O, O-H, and N-H (Li et al., 2019). The wavelength absorption pattern is influenced by the chemical content of the chilli, including pigments such as carotenoids and anthocyanins that determine the colour of the chilli. This was proven by Khuriyati et al. (2023) who stated that a calibration model can be developed to determine the total carotene content in chilli powder using NIR spectra. These findings suggest that NIR spectral information can also be useful for classifying chilli colour groups and predicting colour intensity. However, further investigation is needed to evaluate the use of NIR spectroscopy combined with the Random Forest algorithm for these two tasks. Accordingly, the present study evaluates both classification and colour-intensity prediction within the same analytical framework, while examining the influence of spectral pre-treatments on model performance.

Machine learning is widely used for data analysis to build the best classification and prediction models (Vidyarthi et al., 2020). Researchers commonly apply support vector machines, neural networks, and random forests for product quality evaluation (Alfiani et al., 2024; Chen et al., 2021; Younis et al., 2019). Neural networks excel at predicting non-linear data, parallel processing, and tolerance for errors, but they require large training datasets and are susceptible to overfitting (Fei et al., 2009). Support vector machines can help mitigate overfitting, although selecting optimal parameters can be challenging (Kim et al., 2019). In this context, random forests are a suitable method for measuring NIRS data because they are more robust with respect to noise as an ensemble learning method composed of decision trees (Breiman, 2001).

Therefore, this study aimed to classify and predict the level of colour in chilli powder using a combination of chemometric models based on Near Infrared (NIR) and random forest. In this research, four different levels of colour were used as samples: light green, dark green, red tinge, and red, which also indicate the level of maturity. Colour prediction in chillies has important implications for the food industry, as it can be used to optimise the quality and consistency of chilli-based products (Jiang et al., 2018). In addition, the use of NIR for chilli colour analysis can help to decrease the need for destructive and time-consuming analytical techniques (Hong & Chia, 2021). Overall, this research will contribute to the development of new and innovative methods for chilli powder colour classification with potential applications in the food industry and beyond.

MATERIAL AND METHODS

Materials and Methods

Fresh chillies were obtained from Patrowangsan Village, Candibinangun, Pakem District, Sleman Regency, Special Region of Yogyakarta, Indonesia. A total of 84 chilli powder

samples were classified into 4 colour groups of fresh chillies, using a grouping method similar to that developed by Jiang et al. (2018), including light green, dark green, red tinge, and red. The fresh chilli samples from the field were sorted, washed and dried. The drying process was carried out using a cabinet dryer at a 60°C for 10-15 hours, ground using a food processor, and sieved through a 60-mesh sieve. Then, the chilli powder was packed using a small ziplock plastic bag for further analysis.

Measurement of NIR Spectra

A 0.5 g sample was prepared in a vial tube to ensure uniformity in both surface and thickness. NIR measurements were taken across a wavelength range of 1000 to 2500 nm. The measurements were performed using the NIRFlex N-500 Fiber Optic Solids Cell (Buchi Labortechnik AG, Flawil, Switzerland). Spectra were acquired for each chilli powder sample at four different colours using near-infrared spectroscopy.

Measurement of Chilli Powder Colour

In this study, we deal with two types of colour data: qualitative and quantitative. The qualitative data originated from visual colour assessment, which categorised the colours into four groups. In contrast, the quantitative data regarding chilli powder colour were obtained through measurements using a Chroma Meter Konica Minolta CR-400 (CIELAB coordinates) in order to obtain the values of L* (lightness), a* (redness), and b* (yellowness). The calculation results were then analysed using ANOVA and further analysed using the Least Significant Difference Test (LSDT).

Spectra Analysis

The spectral data were analysed using Unscrambler X 11.0 software from Camo Analytics. Pre-treatments, including first derivative Savitzky-Golay, MSC, and detrending, were applied during data processing to enhance the detection of chemical content associated with functional groups, which serve as indicators of chilli colour characteristics, such as carotenoids or anthocyanins.

Random Forest Model

The machine learning analysis was carried out with the random forest method through Google Colab with Google Compute Engine Python 3 backend, 12.7 GB system RAM, and 107.7 GB disk. Random Forest (RF) is a machine learning model used for both regression and classification tasks. It consists of an ensemble of multiple decision tree (DT) models (Vidyarthi et al., 2020). In this instance, qualitative colour data was utilised for classification, whereas quantitative data served the purpose of prediction. Random forest was performed on the entire spectral data (input) and colour data (output), where each

factor corresponds to the potential characteristics linked to different wavelengths of the NIR data. The data were divided randomly into 75% for training and 25% for testing out of the total sample, and then the prediction results were evaluated using the accuracy score for the classification task and Root Mean Square Error (RMSE) for the regression task.

RESULTS AND DISCUSSION

Analysis of Raw Spectra

When chilli powder samples are scanned with light, the energy from the light is absorbed and reflected by the sample. The reflected light is then captured by a spectrometer and displayed on a computer. Spectra of 84 samples were recorded from 1000 nm to 2500 nm. Each spectrum displayed distinct peaks, indicating the presence of chemical bonds. The spectrum plot in Figure 1 highlights valleys (inverted peaks), which serve as signatures of specific materials in chilli powder. These valleys are observed at wavelengths of 1190-1240 nm, 1410-1490 nm, 1700-1730 nm, 1950-1960 nm, 2060-2080 nm, and 2420-2500 nm. Additionally, absorption peaks for β-carotene were identified at approximately 1235 nm, 1770 nm, 2176 nm, and 2376 nm (Chen et al., 2009). (Hagos et al., 2022) reported that the β-carotene content is associated with a wavelength of 1415 nm, which corresponds to the first overtone band of C-H stretching.

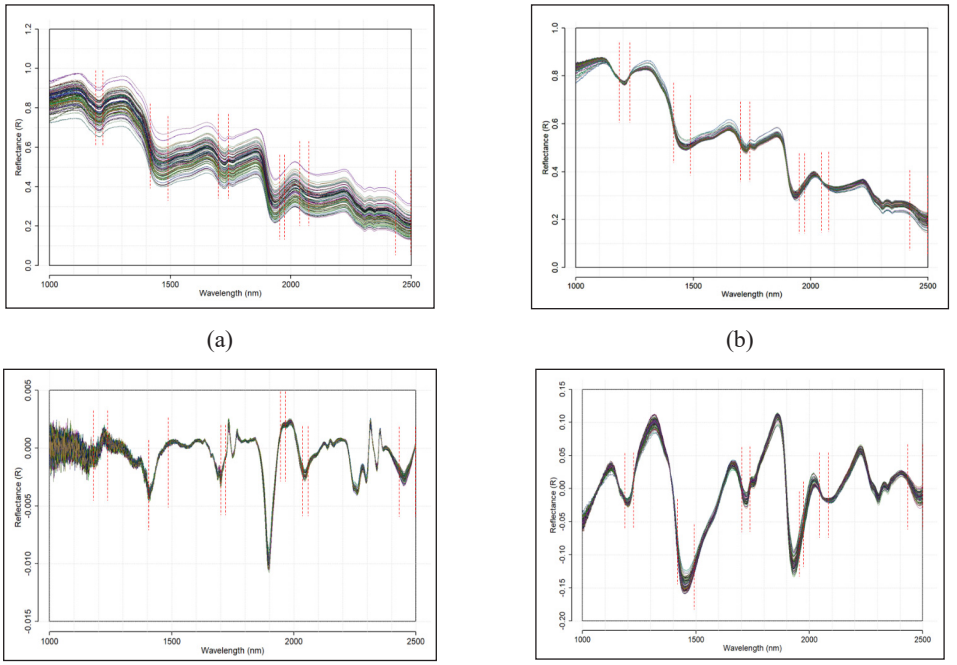

Figure 1. (a) raw spectra; (b) MSC; (c) first derivative savitzky-golay; and (d) de-trending

Pre-Treatment of Raw Spectra

Pre-treatment or preprocessing of raw NIR data serves as the initial step in model development and optimisation. This process is essential for minimising unwanted sample-to-sample variability in the measured NIR spectral data, thereby improving data quality and reliability. Spectral pre-treatment of this data is required to remove all irrelevant information such as noise, uncertainty, variability, interactions, and undetected features. Three spectral pre-treatments were applied to remove nonlinear trends in the data and reduce radio interference: MSC, first derivative Savitzky-Golay, and detrending. The differences between raw spectra and the three pre-treatment spectra are shown in Figure 1.

MSC is a widely used normalisation technique in NIR spectroscopy and other spectroscopic methods. It is designed to correct variations in both scale and offset within a set of spectra, thereby improving data consistency and accuracy (Windig et al., 2008). Multiplicative scatter correction reduced baseline variability and eliminated scatter effects. Additionally, it also reduces the number of latent variables compared to approaches that do not involve preprocessing (Khuriyati et al., 2023). The spectra of the first pre-treatment are presented in Figure 1b. The application of MSC significantly improved the spectral data, producing smoother and more uniform results compared to the raw spectra.

Additionally, the first derivative Savitzky-Golay method was employed to reduce noise while preserving the integrity of the data, effectively minimising the impact of noise on the signal (Candan & Inan, 2014). The first derivative Savitzky-Golay is performed to remove baselines that are thrown off in the spectral data and small spectral differences are amplified (Zaid et al., 2020). The first derivative Savitzky-Golay algorithm fits a polynomial to each successive segment of the curve, replacing the original values with smoother, more consistent variations (Khuriyati et al., 2023). The analysis of the first derivative revealed noticeable changes in the original data, as shown in Figure 1c, which demonstrates significant differences when compared to the raw spectra.

Detrending is a straightforward pre-treatment method used to remove baseline variations by fitting the spectrum to a polynomial of a specified degree and replacing it with the residuals of the fit. This process involves subtracting a linear or polynomial baseline fit from the original spectrum to eliminate tilted baseline variations (Roger et al., 2022).

These results demonstrate significant differences when compared to the raw spectra. Variation between valley shapes and peaks attributed to chilli powder spectra was observed. The peak and depletion spectra of the detrended spectral data showed strong and weak absorption features of chilli powder. This is likely due to the chemicals in the chilli powder. Figure 1d shows the detrended spectral results.

Colour Analysis

L*, a*, and b* values of the chilli colour group can be seen in Figure 2. The brightness value (L*) produced in the chilli colour group tends to decrease in value. The brightness value (L*) in the green group is 51.49, the dark green group is 49.07, the orange group is 48.07, and the red group is 47.24. The a* value shows that the green group has a negative value of -1.8729, the dark green, orange, and red groups have positive a* values of 2.28, 13.09, and 21.41 respectively. The a* value has an upward trend. This is in accordance with (Becker, 2016) where a positive a* value indicates a red colour tendency and a negative a* indicates a green colour. The b* value shows that all chilli colour groups have a positive b* value and have a tendency to increase in value. The b* value in the green group is 21.29, the dark green group is 22.43, the orange group is 24.65, and the red group is 25.20. Positive b* values indicate yellow colour, while negative b* values indicate blue colour (Becker, 2016). Compounds related to the colour of chilli are carotenoids and chlorophyll compounds. One of the dominant carotenoids in red chilli is β-carotene, which is an important intermediate in synthesising red xanthophyll. The concentration of these pigments usually increases when the chilli reaches maturity (Deli et al., 2001). Meanwhile, chlorophyll is known to decrease during the maturation that occurs in chilli peppers. At the beginning of chilli ripening, chlorophyll decreases significantly, while carotenoids increase because chloroplasts turn into chromoplasts which cause the appearance of red colour (Sun et al., 2018).

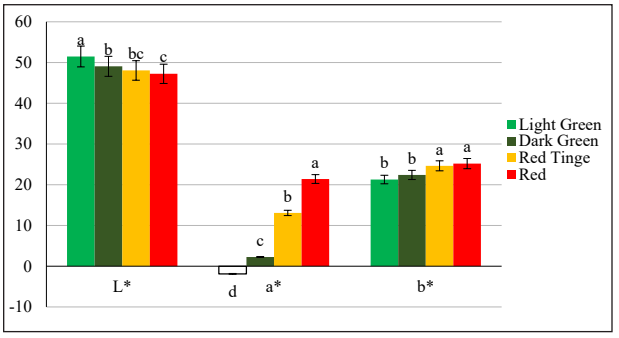

Figure 2. Colour index comparison of chilli powder

Random Forest Model

Dataset Overview

In this paper, the feature data variables (input) used are datasets of NIR spectra measurement results (raw spectra) and pre-treatment results value of raw spectra (detrending, first derivative Savitzky-Golay, and MSC). The target data variable (output) used for the Random Forest Classifier is the pre-classified colour (light green, dark green, red tinge, and red), while for the Random Forest Regressor is the colour index from the Chroma Meter measurement (L*, a*, and b* values). Furthermore, the dataset is divided into two parts, namely 75% training data and 25% testing data. In addition, data division is carried out using the same random seed to ensure consistency in data separation. The results of the dataset division are presented in Table 1.

Table 1
Data division of sample

Colour	Data Division of Samples for Each Colour		Number of Data
	Training	Testing	
Light Green	15	6	21
Dark Green	17	4	21
Red Tinge	16	6	22
Red	15	5	20
Total	63	21	84

Colour Classification Using Random Forest.

Random Forest Classifier is a machine learning algorithm used in this research and proven to be able to classify the colour of chilli powder using data on each treatment (raw spectra, de-trending, first derivative Savitzky-Golay, and MSC). The ability of each Random Forest algorithm to perform classification is interpreted in the accuracy score on evaluation using testing data. The higher accuracy score, better classification into the positive class. The evaluation results for each treatment are presented in Figure 3.

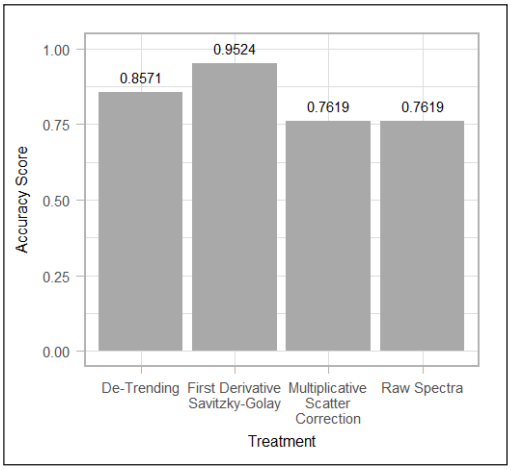

Figure 3. Accuracy scores for each treatment classification

Based on Figure 3, the accuracy score on raw spectra is 0.76, which means that the success rate of colour classification reaches 76.19%, where 16 samples are correctly identified. After applying de-trending pre-treatment on raw spectra, there was improvement in the score accuracy to 0.8571.

This indicates that the use of de-trending was able to increase the classification success to 85.71%, with the number of correctly classified samples being 18 data, although there were colour groups that experienced a decrease in the number of correct classifications. Furthermore, the use of the first derivative Savitzky-Golay pre-treatment also showed an increase in accuracy score to 0.9524, indicating that the first derivative Savitzky-Golay pre-treatment provided a classification success rate of 95.24%, with 20 properly classified samples. On the other hand, using the MSC pre-treatment did not improve the accuracy score, remaining at 0.7619, with 16 correctly classified samples. Although MSC provides a similar success rate to raw spectra, the number of correctly classified samples for each

colour group appears to be different. The details of the colour classification in each treatment are presented in Figure 4.

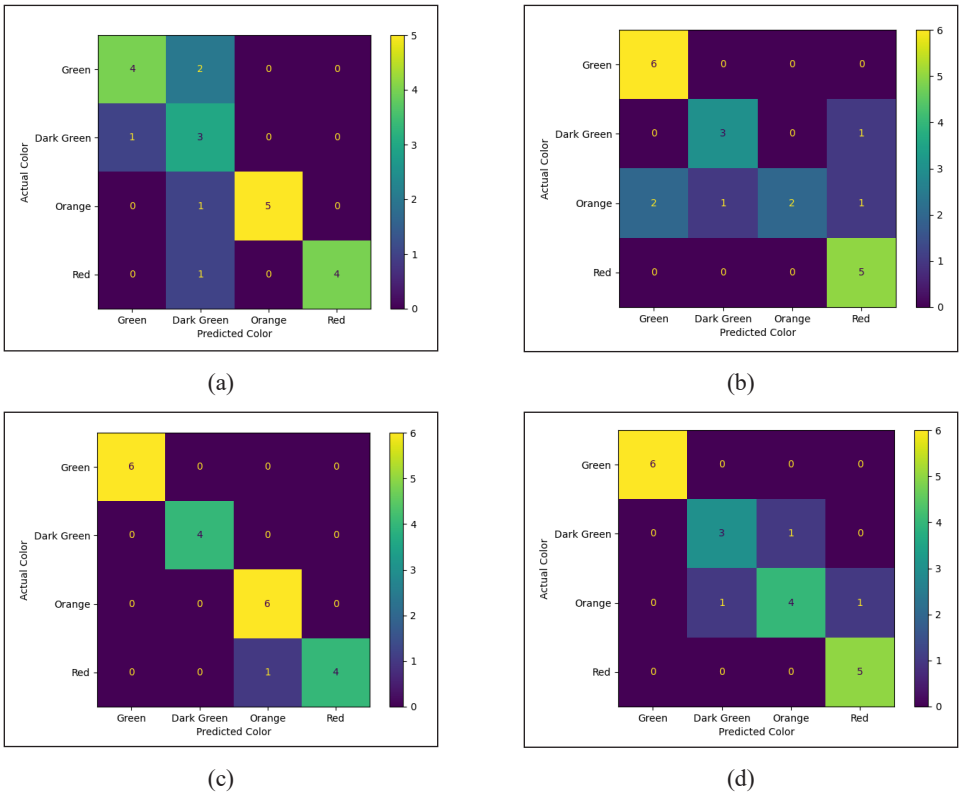

Figure 4. Confusion matrix (a) raw spectra; (b) MSC; (c) first derivative Savitzky-Golay; and (d) detrending

Colour Prediction Using Random Forest.

The need for fast and non-destructive colour measurement in the chilli industry arises because existing methods remain destructive, labour-intensive, and prone to inconsistency, while NIR spectroscopy offers a promising alternative for real-time and non-invasive colour prediction (Theanjumpol et al., 2023). Research on chilli powder colour prediction using NIR spectroscopy is still limited. The present study addresses this gap by combining spectral pre-treatments with the random forest algorithm.

Random Forest Regressor is a machine learning algorithm used in this research and proven to be suitable for predicting the L*, a*, and b* values of chilli powder colour based on the data of each treatment (raw spectra, detrending, first-derivative Savitzky-Golay, and MSC). The ability of each Random Forest algorithm to make predictions is interpreted by the RMSE value on evaluation using testing data. The smaller the RMSE value, the

better the prediction model. This is based on measuring the level of compatibility between the prediction results and the actual data. The evaluation results for each colour index are presented in Figure 5.

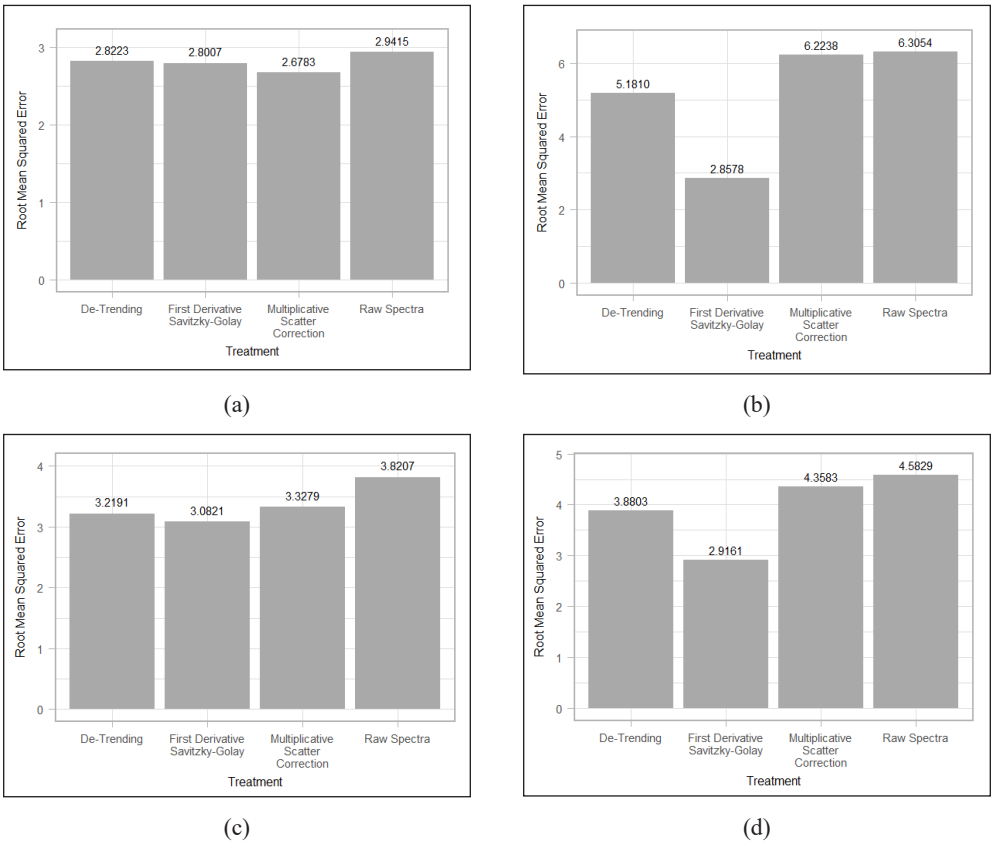

Figure 5. RMSE as an evaluation of the prediction results of the index values of (a) L*; (b) a*; (c) b*; and (d) overall

Based on Figure 5, the ability of the model to predict values of L*, a*, and b* based on NIR data with different treatments is known. Overall, the use of pre-treatment has been shown to improve the performance of the model to predict the values of the L*, a*, and b* indices, which is interpreted through the lower RMSE value. In predicting the L* index values, the best treatment was identified as Multiplicative Scatter Correction, with an RMSE of 2.68, which performed better than the first-derivative Savitzky-Golay, with an RMSE of 2.80. These results indicate that there is a better option for developing and applying a model to predict the L* index value, which is the use of the MSC pre-treatment.

First-derivative Savitzky-Golay pre-treatment achieves the best performance in predicting the a* index value, with the lowest RMSE of 2.86. The results of predicting

the a* index value using the first derivative Savitzky-Golay pre-treatment show better performance than using other treatments, especially when compared to raw spectra, which has the highest RMSE of 6.3054. In addition, first derivative Savitzky-Golay also provides the best results in predicting the b* index value with an RMSE of 3.0821, better than raw spectra, which has an RMSE of 3.8207. A comparison of the predicted and actual L*, a*, and b* index values for each treatment is presented in Figure 6.

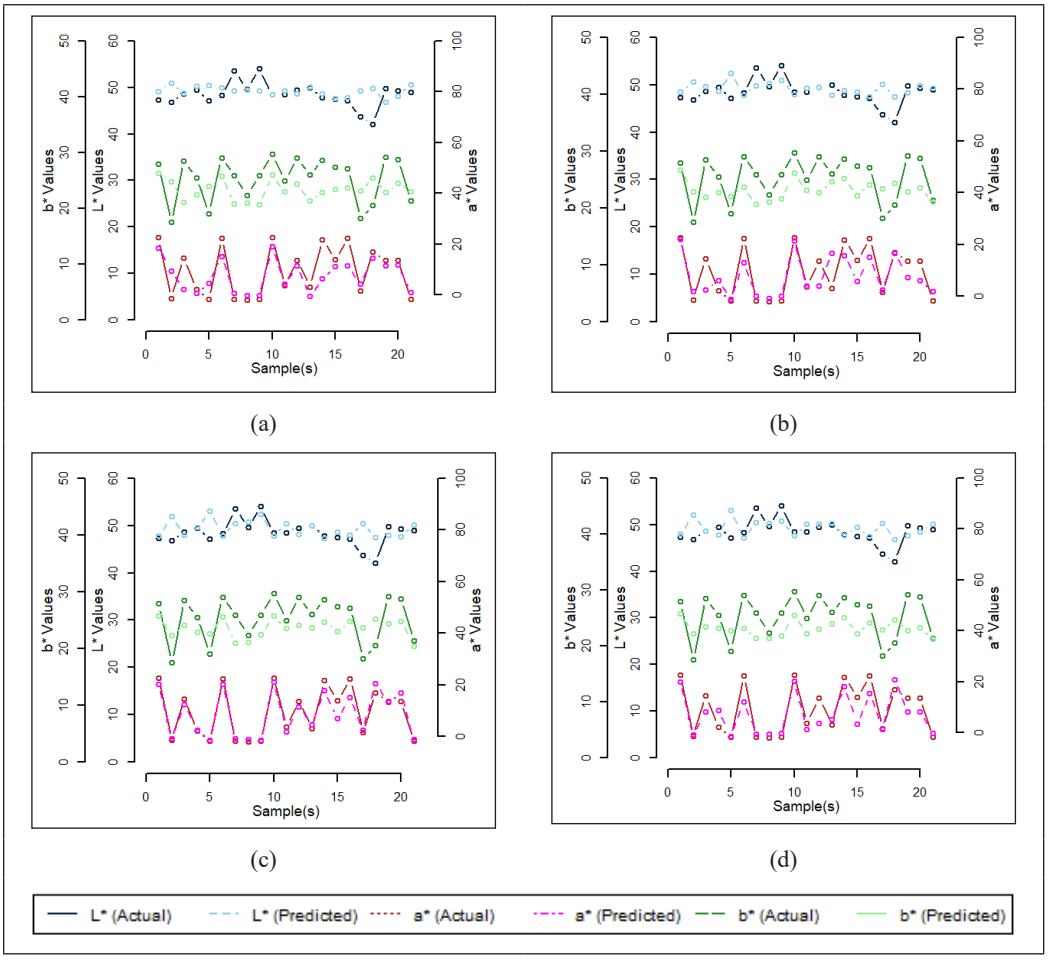

Figure 6. Colour prediction from (a) raw spectra; (b) MSC; (c) first derivative Savitzky-Golay; and (d) detrending

The difference in the best pre-treatment between L* and the other two indices can be explained by the nature of each parameter. The L* value reflects brightness, which is strongly influenced by light scattering on the powder surface, so the correction of multiplicative scatter effects by MSC is more helpful (Zhao et al., 2013).

In contrast, the a* and b* values are closely related to pigment absorption by carotenoids and chlorophyll, so baseline removal and amplification of subtle differences by first-derivative Savitzky-Golay are more effective (Zimmermann & Kohler, 2013). Accordingly, the best pre-treatment for classifying and predicting chilli powder colour depends on the parameter being estimated: first-derivative Savitzky-Golay for classification, a*and b*, and MSC for L*.

First-derivative Savitzky-Golay pre-treatment has been proven to be the best pre-treatment for colour classification and prediction of L*, a*, and b* index values in chilli powder using a Random Forest Classifier and Regressor. This selection is based on the pre-treatment's ability to classify the colour of chilli powder with a 95% success rate. In addition, overall, the first-derivative Savitzky-Golay pre-treatment predicted L*, a*, and b* index values very well, showing improved performance compared to raw spectra. This is interpreted through lower RMSE values compared to raw spectra. According to (Zimmermann & Kohler, 2013) Savitzky-Golay is able to attenuate high-frequency signals, such as noise, while at the same time tending to preserve important features of the chemical analyte signal. Therefore, first-derivative Savitzky-Golay is the most widely used smoothing algorithm in spectroscopy. According to previous research, Savitzky-Golay has also been recognised as the most effective pre-treatment method for predicting antioxidant activity, water content, total carotene, and vitamin C (Khuriyati et al., 2022, 2023).

Previous studies that applied support vector machines and neural networks for product quality assessment reported good performance, but both have limitations with NIR spectral data. Neural networks require large training datasets and are susceptible to overfitting, while support vector machines require careful parameter selection (Crone & Kourentzes, 2010; Ghasemi & Tavakoli, 2013). Random forest, as an ensemble method, is more robust to noise and does not demand complex parameter tuning, making it more suitable for NIR spectral data with a limited number of samples (Ghasemi & Tavakoli, 2013; Manneh et al., 2024).

CONCLUSION

The random forest machine learning method has been successfully applied to classify and predict the colour of chilli powder using near-infrared spectroscopy. Classification of four colour groups achieved the best accuracy of 0.9524 with the first derivative Savitzky-Golay, while raw spectra and MSC reached 0.7619. L* was predicted most accurately with MSC (RMSE 2.68), meanwhile a* and b* were predicted most accurately with first derivative Savitzky-Golay (RMSE 2.86 and 3.0821). The findings could hold significant value for industries that utilise chilli powder as an ingredient, as they can more accurately and reliably classify and predict colour using near-infrared spectroscopy and random forests. However, several limitations remain, including the relatively small sample size and the inclusion of

samples from only one region. Future research should therefore involve a larger sample size and include samples from multiple regions to improve the generalisability of the findings.

ACKNOWLEDGEMENT

The authors are grateful to the Department of Agro-industrial Technology, Faculty of Agricultural Technology, Universitas Gadjah Mada, for providing laboratory facilities to support this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest associated with the publication of this article. The authors confirm that the data and manuscript are free from plagiarism.

REFERENCES

Alfiani, R. N., Ushada, M., Ainuri, M., & Falah, M. A. F. (2024). Modeling consumer overall acceptance for traditional spice-based ready-to-drink using artificial neural network and Kansei engineering. *International Journal on Advanced Science, Engineering and Information Technology*, *14*(4), 1178-1184. https://doi.org/10.18517/ijaseit.14.4.19314

Aşkın, B., & Yazici, S. O. (2021). Bioactive compounds, antioxidant potential and colour properties of dried red pepper (*Capsicum annuum* L.). *International Journal of Innovative Approaches in Agricultural Research*, *5*(2), 158-171. https://doi.org/10.29329/ijiaar.2021.358.1

Batiha, G. E. -S., Alqahtani, A., Ojo, O. A., Shaheen, H. M., Wasef, L., Elzeiny, M., Ismail, M., Shalaby, M., Murata, T., Zaragoza-Bastida, A., Rivero-Perez, N., Beshbishy, A. M., Kasozi, K. I., Jeandet, P., & Hetta, H. F. (2020). Biological properties, bioactive constituents, and pharmacokinetics of some *Capsicum* spp. And capsaicinoids. *International Journal of Molecular Sciences*, *21*(15), Article 5179. https://doi.org/10.3390/ijms21155179

Becker, D. (2016). Colour measurement. In *Colour Trends and Selection for Product Design* (pp. 179-182). William Andrew Publishing. https://doi.org/10.1016/B978-0-323-39395-9.00037-2

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32. https://doi.org/10.1007/978-3-030-62008-0_35

Candan, Ç., & Inan, H. (2014). A unified framework for derivation and implementation of Savitzky-Golay filters. *Signal Processing*, *104*, 203-211. https://doi.org/10.1016/j.sigpro.2014.04.016

Chen, G., Zhang, X., Wu, Z., Su, J., & Cai, G. (2021). An efficient tea quality classification algorithm based on near infrared spectroscopy and random forest. *Journal of Food Process Engineering*, *44*(1), Article e13604. https://doi.org/10.1111/jfpe.13604

Chen, X., Wu, J., Zhou, S., Yang, Y., Ni, X., Yang, J., Zhu, Z., & Shi, C. (2009). Application of near-infrared reflectance spectroscopy to evaluate the lutein and β-carotene in Chinese kale. *Journal of Food Composition and Analysis*, *22*(2), 148-153. https://doi.org/10.1016/j.jfca.2008.10.007

Crone, S. F., & Kourentzes, N. (2010). Naive support vector regression and multilayer perceptron benchmarks for the 2010 Neural Network Grand Competition (NNGC) on time series prediction. In *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN.2010.5596636

Deli, J., Molnár, P., Matus, Z., & Tóth, G. (2001). Carotenoid composition in the fruits of red paprika (*Capsicum annuum* var. *lycopersiciforme rubrum*) during ripening; Biosynthesis of carotenoids in red paprika. *Journal of Agricultural and Food Chemistry*, *49*(3), 1517-1523. https://doi.org/10.1021/jf000958d

Elmasry, G., Sun, D. -W., & Allen, P. (2012). Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *Journal of Food Engineering*, *110*(1), 127-140. https://doi.org/10.1016/j.jfoodeng.2011.11.028

Fei, S. -W., Miao, Y. -B., & Liu, C. -L. (2009). Chinese grain production forecasting method based on particle swarm optimisation-based support vector machine. *Recent Patents on Engineering*, *3*(1), 8-12. https://doi.org/10.2174/187221209787259947

Foroni, F., Pergola, G., & Rumiati, R. I. (2016). Food colour is in the eye of the beholder: The role of human trichromatic vision in food evaluation. *Scientific Reports*, *6*, Article 37034. https://doi.org/10.1038/srep37034

Gabriëls, S. H. E. J., Mishra, P., Mensink, M. G. J., Spoelstra, P., & Woltering, E. J. (2020). Non-destructive measurement of internal browning in mangoes using visible and near-infrared spectroscopy supported by artificial neural network analysis. *Postharvest Biology and Technology*, *166*, Article 111206. https://doi.org/10.1016/j.postharvbio.2020.111206

Ghasemi, J. B., & Tavakoli, H. (2013). Application of random forest regression to spectral multivariate calibration. *Analytical Methods*, *5*(7), 1863-1871. https://doi.org/10.1039/c3ay26338j

Hagos, M., Redi-Abshiro, M., Chandravanshi, B. S., & Yaya, E. E. (2022). Development of analytical methods for determination of β-carotene in pumpkin (*Cucurbita maxima*) flesh, peel, and seed powder samples. *International Journal of Analytical Chemistry*, *2022*(1), Article 9363692. https://doi.org/10.1155/2022/9363692

Hong, F. W., & Chia, K. S. (2021). A review on recent near infrared spectroscopic measurement setups and their challenges. *Measurement*, *171*, Article 108732. https://doi.org/10.1016/j.measurement.2020.108732

Jiang, J., Cen, H., Zhang, C., Lyu, X., Weng, H., Xu, H., & He, Y. (2018). Nondestructive quality assessment of chili peppers using near-infrared hyperspectral imaging combined with multivariate analysis. *Postharvest Biology and Technology*, *146*, 147-154. https://doi.org/10.1016/j.postharvbio.2018.09.003

Jiménez, D., Vardanega, R., Salinas, F., Espinosa-Álvarez, C., Bugueño-Muñoz, W., Palma, J., Meireles, M. A. A., Ruíz-Domínguez, M. C., & Cerezal-Mezquita, P. (2021). Effect of drying methods on biorefinery process to obtain capsanthin and phenolic compounds from *Capsicum annuum* L. *Journal of Supercritical Fluids*, *174*, Article 105241. https://doi.org/10.1016/j.supflu.2021.105241

Khuriyati, N., Sukartiko, A. C., & Alfiani, R. N. (2022). Non-destructive measurement of antioxidant activity and water content in chili powder (*Capsicum annuum* L.) using near-infrared spectroscopy. *International Food Research Journal*, *29*(2), 320-327. https://doi.org/10.47836/ifrj.29.2.10

Khuriyati, N., Sukartiko, A. C., Falah, M. A. F., & Alfiani, R. N. (2023). Determination of total carotene and vitamin C in chili powder (*Capsicum annuum* L.) non-destructively using near-infrared spectroscopy.

International Journal on Advanced Science, Engineering and Information Technology, *13*(1), 69-76. https://doi.org/10.18517/ijaseit.13.1.16800

Kim, N., Ha, K. -J., Park, N. -W., Cho, J., Hong, S., & Lee, Y. -W. (2019). A comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern United States, 2006-2015. *ISPRS International Journal of Geo-Information*, *8*(5), Article 240. https://doi.org/10.3390/ijgi8050240

Konishi, A., Furutani, N., Minamiyama, Y., & Ohyama, A. (2019). Detection of quantitative trait loci for capsanthin content in pepper (*Capsicum annuum* L.) at different fruit ripening stages. *Breeding Science*, *69*(1), 30-39. https://doi.org/10.1270/jsbbs.18070

Li, C., Guo, H., Zong, B., He, P., Fan, F., & Gong, S. (2019). Rapid and non-destructive discrimination of special-grade flat green tea using near-infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *206*, 254-262. https://doi.org/10.1016/j.saa.2018.07.085

Liu, L., Wang, Y., Zhao, D., An, K., Ding, S., & Wang, Z. (2014). of carbonic maceration pre-treatment on drying kinetics of chili (*Capsicum annuum* L.) flesh and quality of dried product. *Food and Bioprocess Technology*, *7*(9), 2516-2527. https://doi.org/10.1007/s11947-014-1253-6

Manneh, M., Ansah, P., Tetarave, S. K., Mishra, M. R., & Kalaimannan, E. (2024). A comparative analysis of random forest and support vector machine techniques on the UNSW-NB15 dataset. In K. Daimi & A. Al Sadoon (Eds.), *Proceedings of the Third International Conference on Innovations in Computing Research (ICR'24)*, (Vol. 1058, pp. 194-203). Springer. https://doi.org/10.1007/978-3-031-65522-7_18

Nowacka, M., & Wedzik, M. (2016). Effect of ultrasound treatment on microstructure, colour and carotenoid content in fresh and dried carrot tissue. *Applied Acoustics*, *103*(Part B), 163-171. https://doi.org/10.1016/j.apacoust.2015.06.011

Roger, J. -M., Mallet, A., & Marini, F. (2022). Preprocessing NIR Spectra for Aquaphotomics. *Molecules*, *27*(20), Article 6795. https://doi.org/10.3390/molecules27206795

Sun, T., Yuan, H., Cao, H., Yazdani, M., Tadmor, Y., & Li, L. (2018). Carotenoid metabolism in plants: The role of plastids. *Molecular Plant*, *11*(1), 58-74. https://doi.org/10.1016/j.molp.2017.09.010

Swammy, K. R. M. (2023). Origin, distribution, taxonomy, botanical description, genetic diversity and breeding of capsicum (*Capsicum annuum* L.). *International Journal of Development Research*, *13*(3), 61956-61977. https://www.journalijdr.com/origin-distribution-taxonomy-botanical-description-genetic-diversity-and-breeding-capsicum-capsicum

Theanjumpol, P., Kaur, A., Muenmanee, N., Chanbang, Y., & Maniwara, P. (2023). Prediction of quality traits in dry pepper powder using visible and near-infrared spectroscopy. *International Food Research Journal*, *30*(1), 193-204. https://doi.org/10.47836/ifrj.30.1.16

Vidyarthi, S. K., Tiwari, R., Singh, S. K., & Xiao, H. -W. (2020). Prediction of size and mass of pistachio kernels using random forest machine learning. *Journal of Food Process Engineering*, *43*(9), Article e13473. https://doi.org/10.1111/jfpe.13473

Wang, D., Dowell, F. E., & Lacey, R. E. (1999). Single wheat kernel colour classification using neural networks. *Transactions of the American Society of Agricultural Engineers*, *42*(1), 233-240. https://doi.org/10.13031/2013.13200

Windig, W., Shaver, J., & Bro, R. (2008). Loopy MSC: A simple way to improve multiplicative scatter correction. *Applied Spectroscopy*, *62*(10), 1153-1159. https://doi.org/10.1366/000370208786049097

Younis, K., Ahmad, S., Osama, K., & Malik, M. A. (2019). Optimization of de-bittering process of mosambi (*Citrus limetta*) peel: Artificial neural network, Gaussian process regression and support vector machine modeling approach. *Journal of Food Process Engineering*, *42*(6), Article e13185. https://doi.org/10.1111/jfpe.13185

Zaid, A., Abu-Khalaf, N., Mudalal, S., & Petracci, M. (2020). Differentiation between normal and white striped turkey breasts by visible/near infrared spectroscopy and multivariate data analysis. *Food Science of Animal Resources*, *40*(1), 96-105. https://doi.org/10.5851/kosfa.2019.e88

Zhao, J., He, J. M., Chen, X., Zhu, X. Q., Wu, J. J., & He, G. T. (2013). A preprocessing algorithm of corn infrared spectrum based on MSC and DOSC. *Advanced Materials Research*, *734-737*, 2893-2897. https://doi.org/10.4028/www.scientific.net/AMR.734-737.2893

Zimmermann, B., & Kohler, A. (2013). Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy. *Applied Spectroscopy*, *67*(8), 892-902. https://doi.org/10.1366/12-06723