*Asian Journal of Research and Reviews in Physics*

***Volume 8, Issue 4, Page 53-65, 2024; Article no.AJR2P.124560***
***ISSN: 2582-5992***

# Predicting Hass Avocado Maturity with NIR Spectroscopy for Non-invasive Dry Matter Estimation in Hass Avocados

**Mercy Amondi [a*], Jared Ombiro [a], Zephania Birech [b] and Duke Oeba [c]**

*[a] Department of Mathematics and Physical Science, Maasai Mara University, Kenya.*
*[b] Department of Physics, University of Nairobi, Kenya.*
*[c] Egerton University, Kenya.*

***Authors' contributions***

*This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.*

***Article Information***

DOI: https://doi.org/10.9734/ajr2p/2024/v8i4175

**Open Peer Review History:**
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/124560

***Original Research Article***

***Received: 01/08/2024***
***Accepted: 03/10/2024***
***Published: 29/10/2024***

## ABSTRACT

**Aims:** To develop a rapid, non-invasive method for predicting Hass avocado maturity using near-infrared diffuse reflectance spectroscopy (NIR-DRS) combined with machine learning algorithms, and to identify the optimal NIR wavelength range for accurate dry matter content prediction.
**Study Design:** An experimental design involving spectral data collection from Hass avocados and the development of machine learning models for dry matter prediction.
**Methodology:** Spectral data from 200 Hass avocados were collected using near-infrared diffuse reflectance spectroscopy (900-2500 nm). To improve the quality of the spectral data and reduce noise, standard normal variate was used to correct for scattering and remove unwanted variability in

**Corresponding author: E-mail: mercyamondi@gmail.com, mercyamondi161@gmail.com;*

***Cite as:*** *Amondi, Mercy, Jared Ombiro, Zephania Birech, and Duke Oeba. 2024. "Predicting Hass Avocado Maturity With NIR Spectroscopy for Non-Invasive Dry Matter Estimation in Hass Avocados". Asian Journal of Research and Reviews in Physics 8 (4):53-65. https://doi.org/10.9734/ajr2p/2024/v8i4175.*

the spectral data. PCA was then performed to reduce the dimension of the spectral data while retaining the most significant variance. Following preprocessing, machine learning models, including Convolutional Neural Networks (CNN), were trained to predict dry matter content, and the optimal wavelength range was determined for accurate prediction.

**Results:** The CNN model demonstrated superior performance for dry matter prediction with $R^2$ of 0.91 in the testing set. The wavelength range of 1000-1500 nm was identified as optimal, offering accurate predictions while reducing computational complexity. This range shows potential for developing cost-effective NIR devices for real-time maturity assessment.

**Conclusion:** NIR spectroscopy combined with machine learning offers a non-invasive, accurate method for predicting avocado dry matter, with potential applications for quality control in the avocado industry. The findings demonstrate that focusing on specific wavelength ranges can lead to more affordable and efficient NIR solutions.

*Keywords: Avocado maturity; convolutional neural networks; dry matter content; machine learning; near-infrared spectroscopy; non-destructive assessment; spectral analysis.*

## 1. INTRODUCTION

Avocado, scientifically termed *Persea Americana* has emerged as a significant global crop, with an annual production of approximately 8.06 million tonnes. Kenya, ranked as the sixth-highest avocado producer globally, plays a crucial role in this market [1,2]. The main type of Avocado produced and exported by Kenya is the *Hass* Avocado, which is unique for its buttery and creamy texture and a sweet taste. The quality and market value of avocados, including *Hass* are largely determined by their maturity level at harvest, which significantly influences post-harvest quality, shelf life, and consumer satisfaction [3,4]. The maturity phase of avocados is also a critical factor in decision-making throughout the supply chain, from harvesting to storage. Immature fruits may wither during storage and have unpleasant taste and texture, while overly mature fruits have shorter shelf lives and are more susceptible to post-harvest issues, resulting in economic losses [5,6]. However, the identification of optimal maturity phases for various avocado varieties poses significant challenges due to the lack of observable external morphological changes during the maturation process [7]. Conventional methods for assessing avocado maturity include measuring dry matter content, oil content, and fruit firmness. These methods, while effective, are destructive, time-consuming, and labor-intensive [8]. For instance, the assessment of dry matter content involves extracting and drying fruit samples, a process that can take up to 72 hours [9]. These traditional approaches lead to significant post-harvest losses and inefficiencies, posing challenges for large-scale production and commercial operations [10,11].

Near-infrared diffuse reflectance spectroscopy (NIR-DRS) has emerged as a promising non-destructive technique for assessing fruit maturity [12]. The method analyzes the light reflected from the fruit to provide information on internal characteristics such as moisture and oil content without causing damage [13]. NIR-DRS has been successfully applied to various fruits for quality assessment, including apples, peaches, and mangoes [14,15,16]. However, the technique has various limitations related to variability in fruit surface properties, external light interference, and the need for precise calibration. The complex nature of the NIR spectra, influenced by factors such as fruit moisture, tissue differences, and scattering coefficients, thus necessitates sophisticated data analysis techniques [17]. Moreover, selecting appropriate wavelength ranges and preprocessing methods can significantly impact the accuracy of predictions [18].

Recent advancements in machine learning and artificial intelligence offer potential solutions to enhance the accuracy and reliability of NIR-DRS measurements [19]. These algorithms can analyze complex spectral data to identify patterns or correlations that may not be readily apparent through conventional methods. Various machine learning techniques, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests (RF), and Convolutional Neural Networks (CNN), have shown promise in spectral data analysis [20,21,22]. The preprocessing of spectral data plays a crucial role in the performance of predictive models. Common preprocessing techniques include SavitzkyGolay smoothing, first and second derivatives, Multiplicative Scatter Correction (MSC), and Standard Normal Variate

(SNV) transformations [23]. These methods aim to reduce noise, correct for baseline shifts, and enhance relevant spectral features [24].

This study aims to address the practical limitations of traditional destructive methods by developing and optimizing a machine learning-assisted NIR-DRS technique for rapid and non-invasive assessment of Hass avocado maturity. It also aims to identify the optimal NIR wavelength regions for accurate dry matter content prediction in Hass avocados. By integrating machine learning techniques with NIR spectroscopy, achieving, this research seeks to establish a robust framework for accurate maturity prediction in avocados. The successful implementation of this technique could lead to substantial economic benefits for avocado producers by reducing post-harvest losses, improving efficiency in maturity determination, and enhancing competitiveness in the global market.

## 2. MATERIALS AND METHODS

### 2.1 Fruit Sampling and Spectral Measurements

This study was conducted during the 2024 harvest season using two varieties of Hass avocados: Giant Hass and Golden Hass. Fruit samples were harvested from five different orchards across three distinct regions in Kenya: Gachie in Kiambu County (1°13'S, 36°45'E), Mecheo in Nyamira County (0°56'S, 35°45'E), and near Melelo market in Narok County (0°49'S, 34°59'E). A total of 200 fruits were harvested from five trees during the early (March) and mid (May) periods to incorporate location variability. After harvesting, the avocados were taken to the laboratory, where they were allowed to equilibrate at 24°C for 24 hours before spectral data acquisition. Spectral measurements were performed using a NIRQuest 512-2.5 spectrometer (Ocean Optics, USA) with a wavelength range of 900-2500 nm. The spectrometer was equipped with a Hamamatsu G9208-512w GaAs linear array detector, providing an optical resolution of 6.3 nm and a signal-to-noise ratio of 10000:1.

Samples were excited at a fixed 45° angle using an Ocean Optics TC-DR probe, which included a halogen tungsten light source and collection optics, as shown in Fig. 1. Spectral data were collected from three distinct spots on each fruit: the peduncle, the base, and the equator. Each spectrum was an average of 5 scans to obtain a more representative spectrum and reduce detector thermal noise [25]. A boxcar width of five scans was used for smoothing the spectra. Calibration was performed before each measurement using a Spectralon 99% white reflective reference standard. The integration time was automatically optimized using Ocean View software, with a typical value of 3 seconds. The spectrometer and light source were allowed 30 minutes of warm-up time before measurements to stabilize the detector response.


**Fig. 1. Set-up for diffuse reflectance of Hass avocado samples**

## 2.2 Destructive Analysis

After conducting spectral measurements, core samples were extracted from the same areas analyzed spectroscopically. These samples were weighed, processed using a blender, and then dried in a hot air oven at 75°C for 72 hours until they achieved a stable weight [26]. Dry matter content was calculated as a percentage of the initial weight as per equation 1.

$$DM\% = \frac{W_1}{W_2} X100 \quad (1)$$

Where $W1$ denotes the weight of the sample after it has been dried in the oven (measured in grams), while $W2$ is the weight before drying.

## 2.3 Development of Machine Learning Models

In this study, four machine learning models—Convolutional Neural Network (CNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF)—were developed to predict dry matter content in avocado fruits using near-infrared (NIR) spectral data. These models were selected due to their effectiveness in handling regression problems with high-dimensional data, such as spectra, and their wide use in similar applications. The spectral data were preprocessed to enhance model performance, and data were split into training (70%) and testing (30%) sets. While other machine learning models, such as XGBoost, could also be effective, the models chosen here cover a diverse range of approaches, from deep learning to classical machine learning. Future studies may explore alternative models for further comparison. All statistical analyses were performed using R version 4.4.1 (2024) with the caret package [27].

### 2.3.1 Data preprocessing

Before model development, the spectral data were preprocessed using several techniques to ensure robust performance. Principal Component Analysis (PCA) was first applied to reduce dimensionality and identify outliers. Baseline shifts and noise were addressed using Savitzky-Golay smoothing with a five-point window and second-degree polynomial fitting. Finally, normalization was applied to standardize the data, ensuring equal weighting of all spectral features. These preprocessing steps were critical for reducing the impact of noise and scattering, which could otherwise degrade model performance.

### 2.3.2 Convolutional neural network (CNN)

A 1D Convolutional Neural Network (CNN) was implemented using the Keras library to handle the high dimensional nature of spectral data. CNNs are well-suited to this task because they can effectively extract local features from the spectral input, which is key for predicting dry matter content. The model architecture consisted of two convolutional layers with 64 and 128 filters, each using a kernel size of 3 and ReLU activation functions. Max pooling layers followed each convolutional layer to reduce dimensionality, while dropout layers (dropout rate = 0.5) were added to prevent overfitting. The model was trained over 1500 epochs with a batch size of 64. Despite the small dataset size (200 fruits), techniques like dropout regularization and data augmentation (by taking multiple spectral readings from different locations on each fruit) were employed to mitigate overfitting. Cross-validation with three folds was used to optimize hyperparameters.

### 2.3.3 Artificial neural network (ANN)

A feedforward Artificial Neural Network (ANN) was constructed with three hidden layers containing 50, 30, and 10 neurons, respectively. Logistic activation functions were applied at each layer. This model was trained using resilient backpropagation, which adjusts weights based on error feedback, with a threshold of 0.01 and a maximum of 10 million steps. ANN was selected due to its flexibility and ability to model complex relationships, serving as a benchmark for deep learning approaches.

### 2.3.4 Support vector machine (SVM)

A Support Vector Machine (SVM) with a radial basis function (RBF) kernel was developed for predicting dry matter content. SVMs are known for their ability to handle non-linear regression problems and perform well with smaller datasets. The model's hyperparameters, including the cost and sigma values, were optimized using three-fold cross-validation.

### 2.3.5 Random forest (RF)

A Random Forest (RF) model was developed with 100 trees and 5 randomly selected predictor variables at each split. Random Forest models

are ensemble learning techniques that are robust to overfitting, particularly in small datasets. They also provide feature importance, offering insights into the most relevant spectral regions for predicting dry matter content. Hyperparameter tuning was performed using cross-validation to optimize the model for the given dataset.

Due to the relatively small size of the dataset, the risk of overfitting, especially with deep learning models like CNNs and ANNs, was a concern. To mitigate this risk, dropout layers were incorporated into the CNN architecture, and data augmentation was used to artificially expand the dataset by capturing multiple spectral readings from each fruit.

### 2.4 Optimization of NIR Wavelength Range

To identify the optimal NIR wavelength range for dry matter prediction, the CNN model (which showed the best performance in initial tests) was trained and evaluated on different spectral regions: (a) Full spectrum: 1000-2350 nm, containing all overtone and combination bands, (b) Second overtone region: 1000-1500 nm, dominated by absorption features from water (H O), hydroxyl (ROH), and various hydrocarbon (CH) groups, (c) First overtone region: 1500-1900 nm, where strong absorption of O-H, CH, and N-H bonds occurs, critical for measuring moisture and organic compounds, and (d) Combination band region: 1900-2350 nm, containing more complex molecular interactions like C-H, C=O, and N-H, which provide detailed information on the structural composition of organic materials.

These specific wavelength regions were chosen because they capture distinct molecular vibrations. The second overtone region focuses on weaker but characteristic bonds, such as O-H and C-H, crucial for moisture detection. The first overtone region provides stronger signals for similar bonds, while the combination band region captures interactions that help to detect the overall composition of dry matter. Model performance for each spectral region was evaluated using Root Mean Square Error of Prediction (RMSEP), Mean Absolute Error (MAE), and coefficient of determination ($R^2$).

### 2.5 Model Evaluation

The performance of all models was assessed using MAE, RMSEP, and $R^2$. These metrics were calculated for both the training and testing datasets to evaluate model accuracy and generalizability. MAE is calculated as the average of the absolute differences between the predicted and actual dry matter percentages, as per equation 2 and : RMSE, a variation of MAE, is calculated by squaring the individual errors before averaging, which is given by equation 3:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \quad (2)$$

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2\right)} \quad (3)$$

where $y_i$ represents the predicted dry matter percentage, $x_i$ is the observed dry matter percentage and n the total sample number. $R^2$ was utilized to assess the overall correlation between the spectral data and the actual concentrations, calculated as per equation 4:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x_i})^2} \quad (4)$$

where $\bar{x}$ denotes the mean of the observed dry matter percentages.

## 3. RESULTS AND DISCUSSION

### 3.1 Development of Machine Learning Models for Dry Matter Prediction

Four machine learning models - Convolutional Neural Network (CNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF) - were developed to predict dry matter content in Hass avocados using NIR-DRS data. The performance of these models was evaluated using Root Mean Square Error of Prediction (RMSEP), Mean Absolute Error (MAE), and coefficient of determination ($R^2$) for both training and testing datasets. Table 1 summarizes the performance metrics for all the four models. The CNN model demonstrated superior performance, achieving the lowest RMSEP (1.9) and MAE (1.4) values, along with the highest $R^2$ (0.91) in the testing set. This indicates that the CNN model had the best generalization ability and predictive accuracy among all tested models. The superior performance of the CNN can be attributed to its ability to automatically learn hierarchical features from the spectral data [28]. They are particularly effective in handling high-dimensional input data, such as NIR spectra, by leveraging their

convolutional layers to capture local patterns and spatial hierarchies [29]. CNNs can preserve the spatial relationships within the spectra that other models might overlook. In contrast, ANN models, despite their flexibility, lack the convolutional structure that helps CNNs isolate key features. Similarly, SVM models excel in finding the decision boundary between classes, but they may struggle with highly complex, non-linear relationships in the data. The Random Forest (RF) model**,** despite its high training performance, struggled in the testing phase, which indicates overfitting. This was likely due to the model's tendency to create overly specific decision trees based on the training data, which failed to generalize to new, unseen data.

The ANN and SVM models showed comparable performance, with RMSEP values of 2.0 and 2.1, respectively, and $R^2$ values of 0.89 for both in the testing set. These results suggest that both models were able to capture the non-linear relationships between NIR spectra and dry matter content effectively, ANN models, despite their flexibility, lack the convolutional structure that helps CNNs isolate key features. Similarly, SVM models excel in finding the decision boundary between classes, but they may struggle with highly complex, non-linear relationships in the data. Interestingly, the RF model showed the best performance on the training set ($R^2$ = 0.98, MAE = 0.8), but its performance dropped considerably in the testing set ($R^2$ = 0.87, RMSEP = 2.5). This discrepancy between training and testing performance suggests that the RF model may have overfitted to the training data, limiting its generalization to new, unseen samples. Fig. 2 provides a visual comparison of the performance metrics for all four modules. The graph clearly illustrates the superior performance of the CNN model, as depicted by lower error rates (RMSEP and MAE) and higher $R^2$ value in the testing set. Fig. 3 shows the regression plots comparing the actual versus predicted dry percentages for all four models. The plots are based on the PCA training and testing datasets in the 1000-2350nm range. The plots also confirm the superior performance of the CNN model, as evidenced by the tighter clustering of points around the 1:1 line.

**Table 1. Performance metrics for the four machine learning models.**

| | T raining S et | | | Testing Set | | |
|---|---|---|---|---|---|---|
| **Modelrm** | **SEC** | **$R^2$** | **MAE** | **RMSEP** | **$R^2$** | **MAE** |
| **CNN** | 1.3 | 0.96 | 0.9 | 1.9 | 0.91 | 1.4 |
| **ANN** | 1.6 | 0.93 | 1.2 | 2.0 | 0.89 | 1.5 |
| **SVM** | 1.7 | 0.92 | 1.1 | 2.1 | 0.89 | 1.5 |
| **RF** | 1.0 | 0.98 | 0.8 | 2.5 | 0.87 | 1.8 |

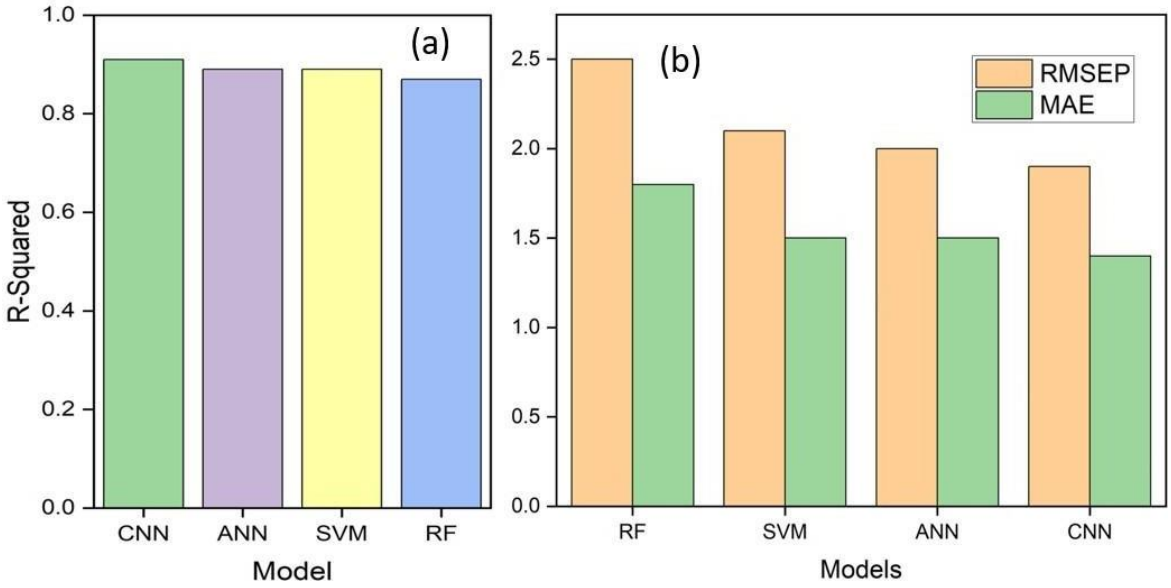


**Fig. 2. Comparison of CNN, SVM, DNN and RF models' evaluation metrics**

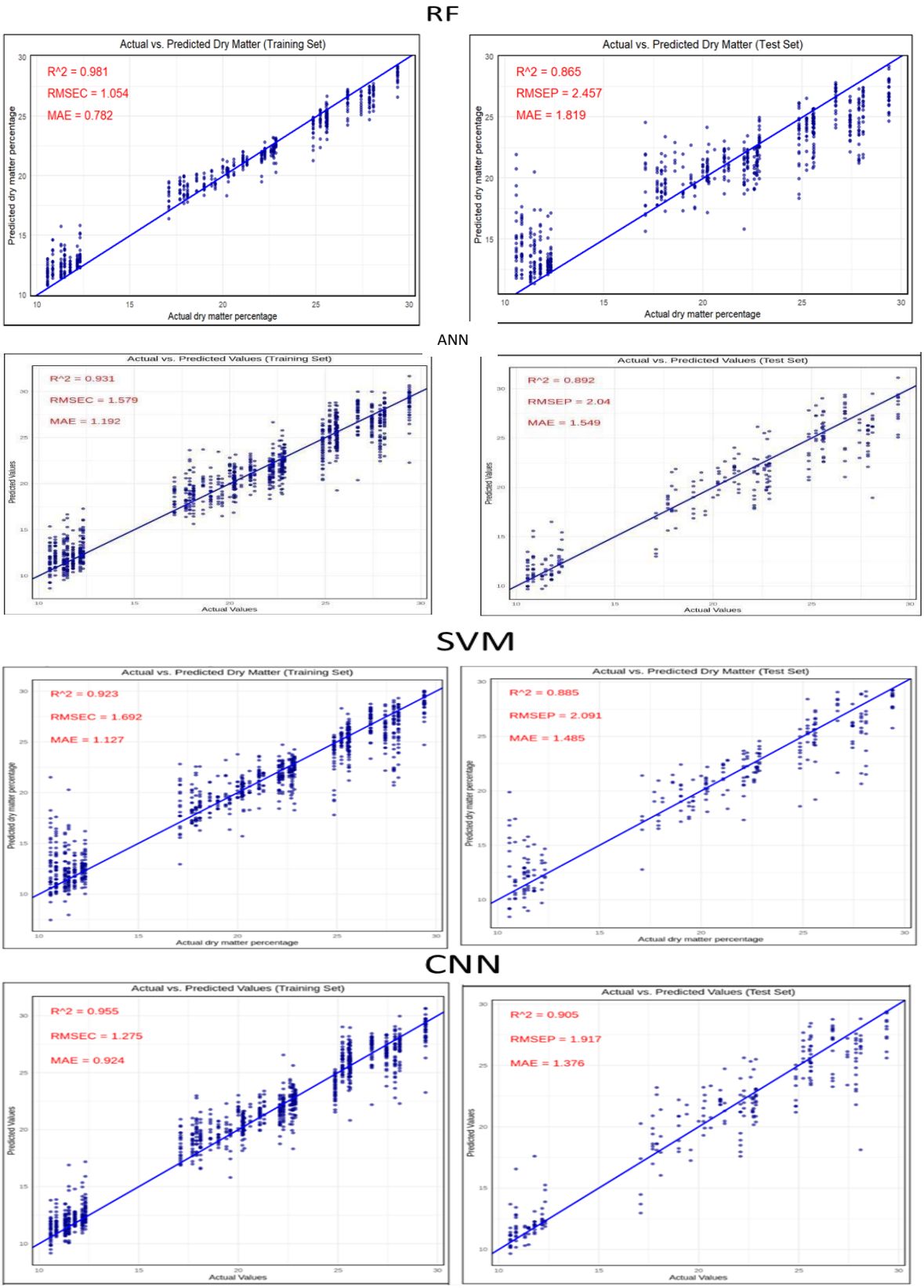


**Fig. 3. Regression plots showing the comparison of actual versus predicted dry matter percentages for CNN, SVM, RF, and ANN models**

**Table 2. Evaluation metrics of the machine learning models at different wavelength regions**

| | Training Set | | | Testing Set | | |
|---|---|---|---|---|---|---|
| **Model** | **RMSEC** | **MAE** | **$R^2$** | **RMSEP** | **MAE** | **$R^2$** |
| **1000-2350nm** | 1.1 | 0.83 | 0.97 | 1.52 | 1.17 | 0.94 |
| **1000-1500nm** | 1.11 | 0.82 | 0.97 | 1.58 | 1.19 | 0.94 |
| **1500-1900nm** | 1.44 | 1.1 | 0.94 | 2.02 | 1.51 | 0.89 |
| **1900nm-2350nm** | 1.15 | 0.86 | 0.96 | 1.65 | 1.28 | 0.93 |

The success of the CNN model in this study aligns with recent findings in spectral analysis. For instance, Acquarelli et al. [30] demonstrated that convolutional neural networks (CNNs) can efficiently classify vibrational spectroscopic data, even outperforming traditional methods like partial least squares (PLS). Their study showed that CNNs achieved an accuracy of 96% on preprocessed spectroscopic data, compared to 89% for PLS, and also performed better on non-preprocessed data. This capability mirrors the strength of CNNs observed in our study, where the model effectively learned hierarchical features from spectral data and achieved high predictive accuracy in assessing avocado dry matter content.

## 3.2 Optimization of NIR Wavelength Range for Dry Matter Prediction

To identify the optimal NIR wavelength range for dry matter prediction in Hass avocados, we evaluated the performance of the CNN model across different spectral regions. Table 2 presents the performance metrics for each spectral region tested. The full spectrum (1000-2350 nm) showed the best overall performance, with an $R^2$ of 0.94, RMSEP of 1.52, and MAE of 1.17 in the testing set. This broad range captures a comprehensive set of molecular vibrations and combinations, providing detailed information for accurate predictions. Notably, the 1000-1500 nm region (second overtone region) showed comparable performance to the full spectrum, with an $R^2$ of 0.94, RMSEP of 1.58, and MAE of 1.19. This region is particularly important due to its association with C-H, C-$H_2$, and C-$H_3$ bonds, which are related to the second overtone of C-H stretching vibrations [31]. These bonds are indicative of the organic content, particularly the lipid and carbohydrate structures in avocados, which are closely related to their dry matter content. This region offers strong absorption features that directly correlate with the lipid and moisture composition in the fruit, making it an important range for prediction.

The 1500-1900 nm region (first overtone region) showed the lowest performance among all tested ranges, with an $R^2$ of 0.89, RMSEP of 2.02, and MAE of 1.51. This suggests that, although it contains important spectral information, it may also include noise or less informative features for dry matter prediction. It corresponds to water content, which is an important component but might not fully represent the variance needed for precise dry matter prediction. The 1900-2350 nm region (combination band region) showed good performance ($R^2$ = 0.93, RMSEP = 1.65, MAE = 1.28), but not as strong as the full spectrum or the 1000-1500 nm region. This region is typically associated with combination bands of fundamental vibrations and can provide valuable information about molecular structure [32]. But it did not perform as well as the 1000-1500 nm range due to the potential inclusion of more complex molecular interactions that introduce noise, reducing its predictive power.

Fig. 4 illustrates the comparison between evaluation metrics and computational time for various wavelength regions. This visualization highlights an important practical consideration: while the full spectrum (1000-2350 nm) provides the best performance, the 1000-1500 nm region offers nearly equivalent predictive power with significantly reduced computational time (16 minutes versus 47 minutes for model training). The strong performance of the 1000-1500 nm region is particularly noteworthy from an applied perspective. This narrower range could potentially allow for the development of simpler and more cost-effective NIR instruments for avocado maturity assessment, without significantly compromising prediction accuracy. This finding aligns with the growing interest in developing targeted, economically feasible NIRS applications in the food industry [33]. The regression plots showing predicted versus actual dry matter content for different NIR regions, as shown in Fig. 5, visually confirm the strong performance of both the full spectrum and the 1000-1500 nm region, as evidenced by the tight clustering of points around the 1:1 line. Thus, the 1000-1500 nm region offers a promising balance between prediction accuracy and practical considerations such as instrument complexity and computational efficiency.

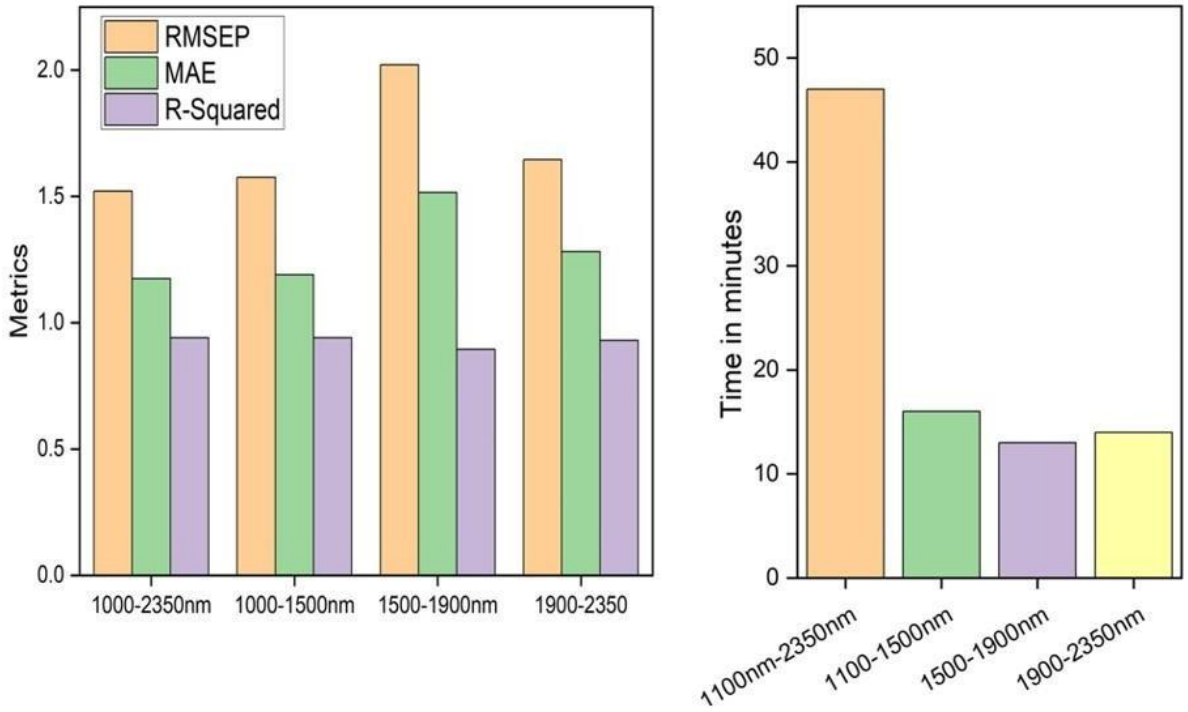


**Fig. 4. Comparison between evaluation metrics and computational time for various wavelength regions**

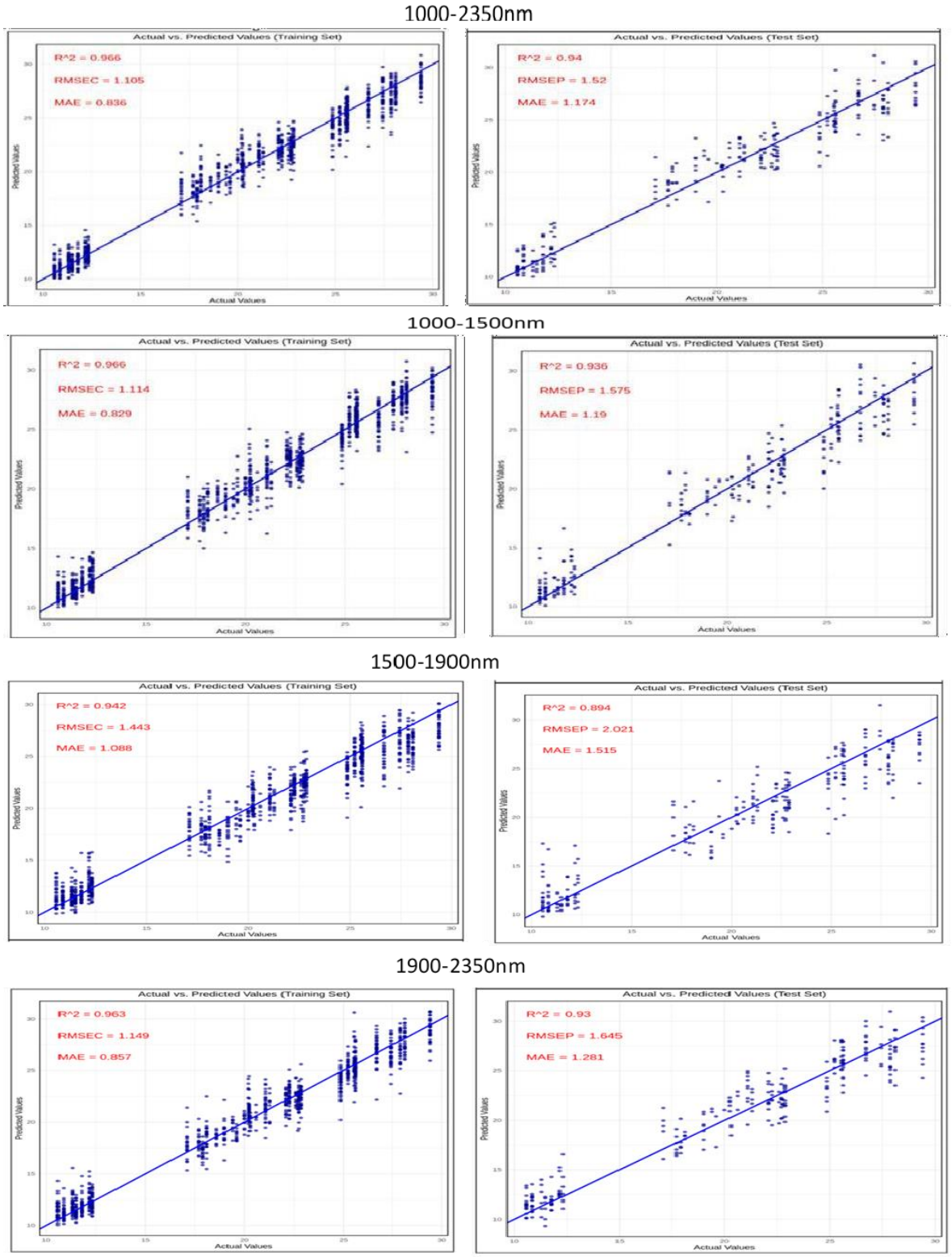


**Fig. 5. Regression plots showing the predicted dry matter compared to actual dry matter for different NIR regions**

The findings from this study have significant implications for the development of rapid, non-destructive methods for assessing avocado maturity. By demonstrating the effectiveness of CNN models and identifying optimal spectral ranges, this research paves the way for more efficient and accurate quality control processes in the avocado industry.

## 4. CONCLUSION

This study has successfully demonstrated the feasibility of using near-infrared diffuse reflectance spectroscopy (NIR-DRS) in combination with machine learning models for the rapid and non-destructive assessment of Hass avocado maturity. The primary objective was to develop accurate predictive models for dry matter content and identify optimal NIR wavelength ranges to ensure efficient and reliable performance. Among the models tested, the Convolutional Neural Network (CNN) emerged as the most effective, achieving a high predictive accuracy with an $R^2$ of 0.91, a root mean square error of prediction (RMSEP) of 1.9, and a mean absolute error (MAE) of 1.4 on the test set. The superior performance of the CNN model underscores its capacity to learn hierarchical features from spectral data, which makes it particularly advantageous for analyzing complex NIR spectra.

Despite its success, this study acknowledges certain limitations. The dataset used was relatively small, which may affect the generalizability of the model to different avocado varieties or environmental conditions, such as changes in humidity or temperature. Future research should focus on expanding the dataset and validating the model's robustness across diverse growing conditions. Additionally, while the full NIR spectrum (1000–2350 nm) provided the most comprehensive information for predicting dry matter content, we found that the narrower range of 1000–1500 nm offered comparable predictive accuracy with reduced computational demands, opening possibilities for more efficient, cost-effective NIR devices. The findings of this study have important practical implications for the avocado industry. By enabling rapid, non-destructive assessments of avocado maturity, this technology can help optimize harvest timing, reduce post-harvest losses, and enhance the quality of fruit reaching consumers. Moreover, the strong performance of the CNN model demonstrates the potential of deep learning techniques in the spectroscopic analysis of agricultural products, laying the groundwork for broader applications. This approach could be extended to the assessment of other fruits and vegetables, advancing the development of non-destructive quality assessment tools in the food industry. Future research could explore the use of advanced deep learning models, such as transformers or deep belief networks, which may offer even greater predictive power and efficiency.

## DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of this manuscript.

## COMPETING INTERESTS

Authors have declared that no competing interests exist.

## REFERENCES

1. Food and Agriculture Organization. FAOSTAT Statistical Database. Rome: FAO; 2021.
2. Rop IK, Karoney EM, Towett DK. Identification of income portfolios among smallholder avocado farmers in Funyula, Kenya. J Agric Crop Res. 2019;7(3):58-64.
3. Magwaza LS, Tesfay SZ. A review of destructive and non-destructive methods for determining avocado fruit maturity. Food Bioprocess Technol. 2015;8(10): 1995-2011.
4. Carvalho CP, Velásquez MA, Van Rooyen Z. Determination of the minimum dry matter index for the optimum harvest of 'Hass' avocado fruits in Colombia. Agron Colomb. 2014;32(3):399-406.
5. Ochoa-Ascencio S, Hertog MLATM, Nicolaï BM. Modelling the transient effect of 1-MCP on 'Hass' avocado softening: A Mexican comparative study. Postharvest Biol Technol. 2009;51(1):62-72.
6. Woolf AB, White A, Arpaia ML, Gross KC. Avocado. In: Gross KC, Wang CY, Saltveit M, editors. The commercial storage of fruits, vegetables, and florist and nursery stocks. Agriculture Handbook 66. Washington, DC: USDA-ARS. 2016;226-229.
7. Osuna-García JA, Doyon G, Salazar-García S, Goenaga R, González-Durán

IJL. Effect of harvest date and ripening degree on quality and shelf life of Hass avocado in Mexico. Fruits. 2010;65(6):367375.
8. Kassim A, Workneh TS, Bezuidenhout CN. A review on postharvest handling of avocado fruit. Food Bioprocess Technol. 2013;6(10):2465-2482.
9. Lee SK, Young RE, Schiffman PM, Coggins CW. Maturity studies of avocado fruit based on picking dates and dry weight. J Am Soc Hortic Sci. 1983;108(3):390-394.
10. Bertling I, Bijzet Z. Effect of pruning on avocado fruit size, yield and alternate bearing - A review. S Afr Avocado Grow Assoc Yearb. 2018;41:70-75.
11. Bill M, Sivakumar D, Thompson AK, Korsten L. Avocado fruit quality management during the postharvest supply chain. Food Rev Int. 2014;30(3): 169-202.
12. Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biol Technol. 2007;46(2):99-118.
13. Subedi PP, Walsh KB, Owens G. Prediction of mango eating quality at harvest using short-wave near infrared spectrometry. Postharvest Biol Technol. 2007;43(3):326-334.
14. Bobelyn E, Serban AS, Nicu M, Lammertyn J, Nicolai BM, Saeys W. Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance. Postharvest Biol Technol. 2010;55(3):133-143.
15. Shao Y, He Y. Nondestructive measurement of acidity of strawberry using Vis/NIR spectroscopy. Int J Food Prop. 2008;11(1):102-111.
16. Jha SN, Narsaiah K, Sharma AD, Singh M, Bansal S, Kumar R. Quality parameters of mango and potential of non-destructive techniques for their measurement - A review. J Food Sci Technol. 2010;47(1):1-14.
17. Cozzolino D, Cynkar WU, Shah N, Smith P. Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. Food Res Int. 2011;44(7):1888-1896.
18. Pasquini C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. Anal Chim Acta. 2018;1026:8-36.
19. Liu Y, Sun X, Ouyang A. Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN. LWT - Food Sci Technol. 2010;43(4):602-607.
20. Guo Z, Huang W, Chen L, Wang X, Peng Y. Nondestructive evaluation of soluble solid content in strawberry by near infrared spectroscopy. In: Li D, Liu Y, Chen Y, editors. Computer and Computing Technologies in Agriculture IV. CCTA 2010. IFIP Advances in Information and Communication Technology, vol 344. Berlin, Heidelberg: Springer. 2011;587-596.
21. Balabin RM, Lomakina EI. Support vector machine regression (SVR/LS-SVM) - an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. Analyst. 2011;136(8): 1703-1712.
22. Acquarelli J, van Laarhoven T, Gerretzen J, Tran TN, Buydens LMC, Marchiori E. Convolutional neural networks for vibrational spectroscopic data analysis. Anal Chim Acta. 2017;954:22-31.
23. Rinnan Å, van den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem. 2009;28(10):1201-1222.
24. Naes T, Isaksson T, Fearn T, Davies T. A user-friendly guide to multivariate calibration and classification. Chichester, UK: NIR Publications; 2002.
25. Cozzolino D, Cynkar WU, Shah N, Smith P. Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. Food Res Int. 2011;44(7): 1888-1896.
26. Lee SK, Young RE, Schiffman PM, Coggins CW. Maturity studies of avocado fruit based on picking dates and dry weight. J Am Soc Hortic Sci. 1983;108(3): 390-394.
27. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28(5):1-26.
28. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-444.
29. Acquarelli J, van Laarhoven T, Gerretzen J, Tran TN, Buydens LMC, Marchiori E. Convolutional neural networks for

vibrational spectroscopic data analysis. Anal Chim Acta. 2017;954:22-31.
30. Acquarelli J, van Laarhoven T, Gerretzen J, Tran TN, Buydens LMC, Marchiori E. Convolutional neural networks for vibrational spectroscopic data analysis. Anal Chim Acta. 2017;954:22-31.
31. Osborne BG, Fearn T, Hindle PH. Practical NIR spectroscopy with applications in food and beverage analysis. Harlow, UK: Longman Scientific & Technical; 1993.
32. Williams P, Norris K. Near-infrared technology in the agricultural and food industries. St. Paul, MN: American Association of Cereal Chemists; 2001.
33. Pasquini C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. Anal Chim Acta. 2018;1026:8-36.



*Peer-review history:*
*The peer review history for this paper can be accessed here:*
*https://www.sdiarticle5.com/review-history/124560*