Non-Destructive Quality Estimation Using a Machine Learning-Based Spectroscopic Approach in Kiwifruits

Georgios Tziotzos 1, Xanthoula Eirini Pantazi 1, Charalampos Paraskevas 1, Christos Tsitsopoulos 1, Dimitrios Valasiadis 2, Elpida Nasiopoulou 2, Michail Michailidis 2,* and Athanassios Molassiotis 2

Abstract: The current study investigates the use of a non-destructive hyperspectral imaging approach for the evaluation of kiwifruit cv. “Hayward” internal quality, focusing on physiological traits such as soluble solid concentration (SSC), dry matter (DM), firmness, and tannins, widely used as quality attributes. Regression models, including partial least squares regression (PLSR), bagged trees (BTs), and three-layered neural network (TLNN), were employed for the estimation of the above-mentioned quality attributes. Experimental procedures involving the Specim IQ hyperspectral camera utilization and software were followed for data acquisition and analysis. The effectiveness of PLSR, bagged trees, and TLNN in predicting the firmness, SSC, DM, and tannins of kiwifruit was assessed via statistical metrics, including R squared (R²) values and the root mean square error (RMSE). The obtained results indicate varying degrees of efficiency for each model in predicting kiwifruit quality parameters. The study concludes that machine learning algorithms, especially neural networks, offer substantial accuracy, surpassing traditional methods for evaluating kiwifruit quality traits. Overall, the current study highlights the potential of such non-destructive techniques in revolutionizing quality assessment during postharvest by yielding rapid and reliable predictions regarding the critical quality attributes of fruits.

Keywords: hyperspectral imaging; artificial intelligence; regression; machine learning; soluble solids; firmness; VisNIR; spectroscopy

1. Introduction

Kiwifruit is harvested when it reaches a physiologically mature stage, but is not fully ripe, with a soluble solid content (SSC) higher than 6.25%. Consequently, it should be preserved to undergo the process of ripening before it is suitable for consumption [1]. One of the consistent quality traits of kiwifruit concerns its dry matter (DM), which remains constant during ripening and only sustains minor losses owing to transpiration or/and respiration [2]. Typically, DM is determined by employing ovens to remove the moisture content, and calculating the ratio (%) between the dry and fresh weights [3]. The “Hayward” kiwifruit places great importance on pericarp firmness as a key quality attribute after harvest. A minimum value of 20 N has been established for the transportation and wholesale of fruit, while for retail sale and direct consumption, the desired value of firmness is considered to be 10 N [4]. The consumption of high-firmness kiwifruit has been associated with greater astringency [5], which is mainly due to the elevated tannin content, and thus affects the aftertaste intensity of kiwifruit [6].

Assessing the qualitative characteristics of kiwifruit before it reaches the stage of ripeness for consumption is highly valuable [7]. An accurate estimation of internal prop-
erties, including dry matter (DM) and soluble solid concentration (SSC), under real-time conditions belongs to crucial factors often related to quality and customer choice [8,9]. The traditional measurement analysis of these properties can reflect the quality of the fruit. However, this type of analysis is labor-intensive and requires destruction of the fruit [10]. Spectral and hyperspectral imaging belong to advanced non-destructive technologies that have drawn a lot of interest in the last few decades for their potential to measure fruit quality attributes [10]. These methods are more efficient, user-friendly, and reliable in postharvest applications compared to conventional approaches [11]. Particularly, visible/near-infrared (Vis/NIR) spectroscopy is a nondestructive analytical technique that appears promising and does not require pre-sample preparation for quality assessment. The Vis/NIR spectroscopy technique has been utilized to evaluate the qualitative attributes of various fruits, such as apples, citrus fruits, and kiwifruits [12]. Non-destructive techniques like NIR, along with hyperspectral imaging, have been extensively employed to assess the qualitative characteristics of kiwifruit, such as firmness, pH, soluble solid content (SSC), and dry matter (DM) [2,13–22]. The integration of Vis/NIR into machine learning (ML) algorithms enhances the effectiveness of learning, estimating, predicting, and classifying crucial quality parameters [23]. Their effectiveness lies in their capability of extracting targeted information from the investigated dataset, securing fruit sustainability and resource consumption that comply with the fundamental principles of precision agriculture. Moreover, artificial intelligence (AI) takes advantage of its computational potential, enabling the successful modeling of quality properties in fruits and vegetables [24], providing constant qualitative and quantitative tools for the assessment of several fruit profiles. The integration of both machine learning and visible/near-infrared spectroscopy has been utilized in the analysis of various fruits, particularly those with elevated levels of glucose, fructose, vitamins, and other vital nutrients [25]. Partial least square regression (PLSR) was used to predict the internal quality of Actinidia chinensis var. deliciosa (A. Chev.) cv. “Hayward” using Vis/NIR spectroscopy [26]. On the other hand, PLSR is known for its exceptional reliability and durability when it comes to creating models for SSC and other internal quality traits [17].

The current study aims to predict the quality traits of kiwifruit of individual fruit cv. “Hayward”, including firmness, tannins, SSC, and DM, with the help of hyperspectral data obtained from the kiwifruit surfaces. Three different regression analysis methods were used for the assessment of the four investigated quality traits. The employed regression models aimed to extract possible correlations between the hyperspectral data and each of the four investigated quality parameters to enable predictions regarding the tested parameters. The overall results documented that the PLSR, bagged trees, and TLNN models may reflect the prediction results for firmness, SSC, DM, and tannins in individual kiwifruit during postharvest ripening.

2. Materials and Methods

2.1. Plant Material and Sampling Process

Kiwifruit (Actinidia chinensis var. deliciosa (A. Chev.) A. Chev. “Hayward”) was harvested in early November (1–3) 2020 from 20 commercial orchards (from the area of Imathia, Northern Greece), and was immediately transferred to the Pomology Laboratory of the Aristotle University of Thessaloniki (AUTh). Regarding the applied sampling procedure, kiwifruit samples that demonstrated uniform appearance (with no defects) and weight were selected from each orchard to form a representative sample dataset. The samples were placed into 16-slot wooden shipping crates (the fruit and crates were marked) and stored at 0 °C (RH 95%) for 40 days, accomplishing a high variation in the tested kiwifruit quality traits. Afterward, kiwifruit was delivered to the Agricultural Engineering Laboratory of AUTh and the acquisition of the hyperspectral data was carried out on the same day at 20 °C. Each spectral shot was taken by placing a crate under halogen lamps. Each spectral shot was acquired by manually placing the crate stationary under the halogen lamp at a distance of 60 cm from the camera. Each shot required 3 mins to be acquired.
due to lower artificial light intensity compared to natural sunlight. Following the camera shot, firmness and dry matter (DM) were destructively determined in individual fruit, and then each marked fruit was sampled with liquid nitrogen and stored at $-80^\circ$C for further analysis.

2.2. Hyperspectral Data Acquisition

For data acquisition, a Specim IQ (Specim Ltd., Oulu, Finland) hyperspectral camera was used. The camera was portable and easy to carry since it only weighed 1.3 kg. There was also the possibility mounting it on a tripod to take hyperspectral images of objects at a distance that exceeds 150 mm. In the current study, the camera was placed on a tripod at a static position, following the Specim IQ standards for spectral data acquisition and following the Specim’s calibration procedure. It also provided the ability to capture data in the visible and near-infrared (Vis-NIR) spectra for 204 spectral wavelengths between 400 and 1000 nm. The device had a touch-screen and could be connected to a computer via a USB cable or a wireless Wi-Fi network; therefore, it was possible to immediately evaluate the quality. Principal component analysis (PCA) was performed on the attained "Hayward" kiwifruit Vis/NIR spectral data so as to extract the most relevant features from the acquired dataset. The extracted features denoted linear combinations of the original spectral bands that were captured through Specim IQ hyperspectral camera, conveying the most important characteristics of the attained hyperspectral data. This data reduction technique, implemented under the general framework of the data preprocessing procedure, was a crucial step for forming a more manageable dataset to be fed as input to the three employed regression models for the prediction of four investigated qualities in kiwifruit.

Hyperspectral captures were acquired using the default recording mode (DRM) option, in which raw data and reflection data are stored without being subjected to further processing by the device software. The device captures the absolute black (dark reference) with the shutter closed, and then the data are downloaded. The white reference confirmation was performed using simultaneous shooting by placing the special white reference panel provided by the manufacturer with the camera [27].

Image extraction was carried out with the help of specialized software, Specim IQ Studio version 2019.05.29.2, provided by the manufacturer of the hyperspectral camera. The software provides the ability to display the relevant downloads and reflectance data, create data libraries, and create groups in classification models. It also provides the ability to transfer the images from the camera environment to the computer environment.

2.3. Kiwifruit Quality Trait Determination

The kiwifruit firmness of pericarp was determined in each marked fruit using the Texture Analyzer TA XT2i (Stable Microsystems, Godalming, Surrey, UK), as previously described [28]. Initially, the peel was removed on two opposite sides at a depth of 1 mm; then, a steel cylinder with a 7.9 mm diameter (flat probe end) that was fitted in the machine branch was inserted into the pericarp (1 cm) at a speed of 20 mm/s. The results were expressed in Newton (N). After drying each marked fruit to a constant weight at 67 $^\circ$C, the dry matter (DM) content (%) of a 5 mm-thick equatorial slice was found by dividing the fresh weights by the dried weights. The soluble solid concentration (SSC, %) was assessed using the juice of each marked fruit employing a digital refractometer (Atago PR-1, Atago Co., Ltd., Tokyo, Japan). The polyphenolic substances were extracted using 70% acetone and 0.5% acetic acid. Kiwifruit pericarp was ground up and added to the extraction solution at a ratio of 1:10 [29]. Tannins were determined using the Folin–Ciocalteu method, with some modifications [30]. Absorbance was measured at 760 nm on a microplate reader (Tecan infinite M200 PRO). The polyphenolic extract was incubated with polyvinyl polypyrrolidone (PVPP) to create a tannin–PVPP complex at the following ratio of extraction: PVPP, 30:1, v/w, and an absorbance at 760 nm of the residual phenolic solution was subtracted from that of total phenols. The results were expressed in equivalents of mg gallic acid kg$^{-1}$ fresh weight (FW) [31].
2.4. Hyperspectral Visualization Software

The spectral data were entered into the Scyven program, which is a high-spatial (HS) image processing tool. The program has the capability to import and modify HS-type files (.hsz, .hdr, .tif) such as the Specim IQ downloads [32]. The Recover Reflectance option was utilized in the program to provide the reflectance data. Subsequently, the image label tool (specifically, the polygon tool) was employed to identify regions that matched the surface of each kiwifruit (Figure 1a,b).

![Figure 1](image_url)

**Figure 1.** (a) Acquisition of RGB images. (b) Polygon selection attained from the surface of the 8 + kiwifruit.

2.5. Hyperspectral Data Preprocessing

The attained hyperspectral data were exported to a Comma Separated Values, CSV-type file and then imported into the Excel environment where they were consolidated into an xlsx type file. A total of 104,394 spectral signatures were exploited for 110 different examined kiwifruits. During data import into the Matlab environment, data with zero values were removed, while missing reflectance values were replaced with the nearest acceptable value.

To avoid any potential issues and abnormalities that may occur during image acquisition that are often responsible for difficulties in analysis and misleading results, it was essential to ensure that data availability was of high quality to be subjected to further analysis. For this reason, one of the very well-known smoothing methods, called the Savitzky–Golay [33], was applied to the attained hyperspectral data (Figure 2).

![Figure 2](image_url)

**Figure 2.** Raw reflectance values per kiwifruit at each spectral point between 400–1000 nm.

This technique belongs to a common technique that is primarily based on the selection of sub-windows around a specific point, and the subsequent estimation of the points of the sub-window projection onto a polynomial fitting. Missing reflectance values were replaced...
with the nearest acceptable value. Moreover, outliers were removed with the help of the outlier function which removes values that are more than three standard deviations from the data dimension to secure and enhance the reliability and integrity of the investigated dataset. A 10-fold cross-validation was performed by splitting the dataset into 10 subsets, with each subset being further divided into training–validation–testing sets at 70%, 20%, and 10% respectively.

2.6. Partial Least Square Regression (PLSR)

PLSR is performed with the plsregress function in Matlab in the general form given as follows:

\[
\text{[Xloadings,Yloadings,Xscores,Yscores,beta,pctVar,mse,stats] = plsregress(X,y,n)}
\]

where \(X\) represents the dependent variables matrix, \(y\) is the independent variable matrix, \(n\) denotes the number of components, \(X\) loadings denotes the matrix with coefficients that define linear combinations of the components and simulate the initial data \(X\), \(Y\) loadings is the table with coefficients that define linear combinations of the components and simulate the initial data \(y\), \(X\) scores denotes the table where each row corresponds to an observation of the \(X\) table and each column to a component, \(Y\) scores represents the table where each row corresponds to an observation of the \(y\) table and each column to a component, \(beta\) is the table with the PLS coefficients, \(pct\) Var denotes the table containing one row with the percentage of variation (variance) of \(X\) and one row with the percentage of variation of \(y\), \(mse\) is the table with the mean squared error (mean squared error) for the values of \(X\), \(y\), and \(stats\) represents the statistics of the model with values of the weights (weights) and the R^2 error for the values of \(X\) scores.

For carrying out the analysis, a high number of components (\(n > 40\)) were initially selected to find the percentage of the model that is successfully predicted according to the number of components used by employing the \(pct\) Var.

2.7. Bagged Trees Regression

The bagged trees algorithm [34], also known as bootstrap aggregating or bagging, is a powerful machine learning technique designed to improve the predictive performance and robustness of decision trees. Its function is characterized by the construction of multiple decision tree models, each being trained on a subset of training data derived by repeated random sampling from the original training data, the samples being randomly selected with replacement. By aggregating the predictions of these individual trees, typically through “majority voting” for classification or averaging for regression problems, bagged trees reduce the risk of overtraining, improve model generalization, and increase overall accuracy. Additionally, this ensemble method leverages the diversity of individual trees to capture various patterns and noise in the data, resulting in a more robust and stable predictive model that is less sensitive to variations in the training data.

2.8. Three-Layered Neural Network (TLNN)

A three-layer neural network is a machine learning model that consists of an input layer, an intermediate layer, and an output layer [35]. Each layer contains neurons that are connected to each other. At the input layer, the input data access the network, then the information is transferred to the intermediate layer, where many parallel processes are performed, such as calculating weights and applying activation functions. The output layer extracts the final prediction or output of the model. Training the three-layer neural network consists of optimizing connection weights based on the training data, allowing it to learn complex relationships and patterns in the data. The structure of this three-layer neural network allows for an efficient representation and extraction of features from data, making it ideal for many applications such as pattern recognition, classification, and prediction. Layers with 10 parallel neurons per layer and ReLu activation function [36] were used in the model.
3. Results

Kiwifruit quality attributes, such as firmness (determined in kilograms), soluble solid concentration (SSC, determined as a percentage), dry matter (DM, determined as a percentage), and tannins (determined in milligrams per gram), were determined for 10 marked fruits from 20 different orchards after 40 days of cold storage at 0°C (RH 95%). The kiwifruit quality feature measurements are demonstrated in Table 1, while the kiwifruit quality measurements of each orchard are provided in Supplementary Table S1. Concisely, the mean kiwifruit firmness was 37 N, with the lowest and highest values being 4 and 66 N, respectively. The mean SSC was 12.7%, ranging from a minimum of 9.7% to a maximum of 15.7%. The average DM value was 17.6%, the lowest value was 14.1%, and the highest value was 23.5%. The mean tannin content was determined to be 198 mg kg\(^{-1}\), ranging from a minimum of 0 to a maximum of 562 mg kg\(^{-1}\).

<table>
<thead>
<tr>
<th>Firmness (N)</th>
<th>SSC (%)</th>
<th>DM (%)</th>
<th>Tannins (mg/kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>36.9</td>
<td>12.68</td>
<td>17.59</td>
</tr>
<tr>
<td>Median</td>
<td>37.7</td>
<td>12.60</td>
<td>17.29</td>
</tr>
<tr>
<td>Min. Value</td>
<td>4.4</td>
<td>9.70</td>
<td>14.11</td>
</tr>
<tr>
<td>Max. Value</td>
<td>65.7</td>
<td>15.70</td>
<td>23.50</td>
</tr>
<tr>
<td>Range</td>
<td>61.3</td>
<td>6.00</td>
<td>9.39</td>
</tr>
<tr>
<td>SD</td>
<td>14</td>
<td>1.25</td>
<td>1.93</td>
</tr>
</tbody>
</table>

Table 1. Kiwifruit quality extracted features.

The regression analysis regarding kiwifruit firmness, SCC, DM, and tannins was performed by applying three different regression techniques, including partial least squares regression (PLSR), bagged trees, and TLNN (Table 2). The models were evaluated by using R\(^2\) (coefficient of determination) and the root mean square error (RMSE). R\(^2\) is generally used for determining the correlation of the variance in the dependent variable with its corresponding independent variables. It takes values between 0 and 1, where higher values indicate better performance. A value of 1 signifies the model’s optimal prediction capability.

Table 2. Regression analysis results for firmness, SSC, DM, and tannins using the PLSR, bagged trees, and TLNN algorithm models.

<table>
<thead>
<tr>
<th></th>
<th>PLSR</th>
<th>Bagged Trees</th>
<th>TLNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firmness (n^\ast = 28)</td>
<td>(R^2)</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.3419</td>
<td>0.3486</td>
</tr>
<tr>
<td>SSC (n^\ast = 30)</td>
<td>(R^2)</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.3479</td>
<td>0.3700</td>
</tr>
<tr>
<td>DM (n^\ast = 19)</td>
<td>(R^2)</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.3685</td>
<td>0.7244</td>
</tr>
<tr>
<td>Tannins (n^\ast = 32)</td>
<td>(R^2)</td>
<td>0.91</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.3418</td>
<td>0.4141</td>
</tr>
</tbody>
</table>

\(n^\ast\) indicates the number of PLS components.

On the other hand, the RMSE denotes the square root of the mean squared error (MSE), which represents the average squared deviation between the predicted and actual values of the data. An RMSE equal to 0 indicates a perfect prediction, so values closer to 0 indicate better performance. As seen in Table 2, the method showcased distinct performances based on the R squared \(R^2\) values and root mean square error (RMSE).

For each of the applied ML models, the relevant linear regression equation was extracted, demonstrating that the mathematical relationship between the predicted and
observed values for the four investigated quality traits was demonstrated (Table 3). The linear regression equations hold the typical form, given as follows (Equation (1)): 

\[ Y = m \times x + b \]  

(1)

where \( m \) denotes the slope, indicating the rate at which each of the predicted quality traits changes with respect to changes in the observed variable; while \( b \) denotes the intercept, representing the predicted value when the observed variable is zero.

Table 3. Linear regression equations between the predicted and observed values regarding firmness, SSC, dry matter, and tannin content prediction derived from the PLSR, bagged trees, and TLNN models.

<table>
<thead>
<tr>
<th></th>
<th>Firmness</th>
<th>SSC</th>
<th>Dry Matter</th>
<th>Tannin Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLSR</td>
<td>( y = 0.9005x + 0.3667 )</td>
<td>( y = 0.9034x + 1.225 )</td>
<td>( y = 0.9047x + 1.676 )</td>
<td>( y = 0.9071x + 0.01835 )</td>
</tr>
<tr>
<td>Bagged Trees</td>
<td>( y = 0.8884x + 0.4081 )</td>
<td>( y = 0.872x + 1.665 )</td>
<td>( y = 0.8012x + 3.511 )</td>
<td>( y = 0.7691x + 0.04888 )</td>
</tr>
<tr>
<td>TLNN</td>
<td>( y = 0.9706x + 0.07884 )</td>
<td>( y = 1.002x - 0.02039 )</td>
<td>( y = 1.001x - 0.009929 )</td>
<td>( y = 1.004x + 0.005643 )</td>
</tr>
</tbody>
</table>

Moreover, statistical tests on the intercept and slope were performed to evaluate the prediction adaptability and consistency of each of the employed MLs with the actual observed values (Table 4).

Table 4. Results of the comparison of each linear regression with the corresponding 1:1 line. The n.s. symbol denotes statistically non-significant differences, and the * and ** symbols represent statistically significant differences between the lines, at \( p < 0.05 \) and 0.01, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Firmness</th>
<th>SSC</th>
<th>Dry Matter</th>
<th>Tannin Content</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comparison with a 1:1 Line</td>
<td>Comparison with a 1:1 Line</td>
<td>Comparison with a 1:1 Line</td>
<td>Comparison with a 1:1 Line</td>
</tr>
<tr>
<td>PLSR</td>
<td>ns **</td>
<td>ns **</td>
<td>* **</td>
<td>ns **</td>
</tr>
<tr>
<td>Bagged Trees</td>
<td>* **</td>
<td>* **</td>
<td>ns **</td>
<td>ns **</td>
</tr>
<tr>
<td>TLNN</td>
<td>ns **</td>
<td>* **</td>
<td>* **</td>
<td>**</td>
</tr>
</tbody>
</table>

For each combination consisting of one of the investigated traits and one of the three employed ML models, the fitted regression line was compared with the 1:1 line in order to explore statistically significant differences with respect to the slope and intercept. At this point, it is worthy to note that a non-significant intercept or slope may suggest the model’s capacity to capture an essential aspect of the relationship between the observed and predicted variables. Among all three applied ML regression models, the PLSR model demonstrated statistically significant relationships (\( p < 0.05 \)) between firmness and all predictors, while the bagged trees model showed significant relationships for firmness, dry matter, and tannin content, and marginally significant for soluble solid content. The TLNN indicated significant associations only between firmness and tannin content, with all other predicted values being statistically non-significant (Table 4).

3.1. Kiwifruit Firmness Analysis

By analyzing the percentage of variance regarding firmness prediction, it is evident that the initial 10 components did not provide enough information to the model. These components only account for up to 50% of the data, as demonstrated in Figure 3a. Conversely, it was noted that 28 components were successfully predicted with a considerably high accuracy that exceeds 90% for all three utilized models, namely PLSR, bagged trees,
and TLNN (Table 2). It is also worthy noting that the model’s performance tends to stabilize for a larger number of components, while there is a slight increase in the prediction rate. This behavior can be attributed to the employed model’s ability to capture some additional variations in the hyperspectral data that have not been fully accounted for with fewer components (Figure 3a).

![Figure 3](image-url)

**Figure 3.** Firmness prediction results regarding variance. (a) Regression scatter plots comparing the predicted vs. observed values for firmness derived from the PLSR. (b) Bagged trees (c) and TNN (d) models. The black line in the graphs (b–d) denotes the theoretical 1:1 line, while the yellow line denotes the fitted regression line.

The PLSR model exhibited a strong performance with an $R^2$ of 0.90, signifying that approximately 90% of the variance in the dependent variable could be explained by the model (Figure 3b). Nevertheless, the root mean square error (RMSE) of 0.3419 signifies a substantial degree of residual error, implying a deviation between the expected outcomes and the actual observed values.

On the other hand, bagged trees demonstrated a higher $R^2$ of 0.93, indicating a better fit to the hyperspectral data than PLSR. Despite this, its RMSE equal to 0.3486 implies slightly higher error residuals compared to the PLSR, indicating a slightly less accurate predictive performance on average (Figure 3c and Table 2). However, the most notable performance was observed in the case of TLNN, which demonstrated an $R^2$ of 0.97. This high value indicates an extremely strong correlation between the predicted values and
the target value, capturing 97% of the variance (Table 2). Moreover, the remarkably low RMSE of 0.246 signifies highly accurate predictions, indicating minimal deviation between the predicted and observed values. The aforementioned results emphasize the enhanced performance of the TLNN model in accurately representing the intricate connections within the kiwifruit data (Figure 3d). The PLSR and bagged trees models demonstrated satisfactory performances, but overall, the ANN model stands out among the other models for its remarkable predictive capacity and minimum error. This makes it a highly suitable and robust choice for predicting and modeling kiwifruit-related properties, as illustrated in Figure 3 and Table 2.

3.2. Soluble Solid Concentration (SSC) Results

Based on the variance percentage explained versus the number of components (n) regarding the SSC, it was noted that the first 12 components were proven insufficient for accurately predicting the model as they explain up to 50% of the data. When “n” is more than 30, the accuracy of the prediction increases over 90%, as demonstrated in Figure 4a. It was also observed that the curve tends to stabilize for a higher number of components, which slightly increases the prediction performance.

![Figure 4](image-url)  
**Figure 4.** Soluble solid concentration (SSC) prediction results regarding variance. (a) Regression scatter plots comparing the predicted vs. observed values for SCC derived from the PLSR; (b) bagged trees algorithm; and (c) TLNN algorithm (d) models. The black line in the graphs (b–d) denotes the theoretical 1:1 line, while the yellow line denotes the fitted regression line.
The analysis of kiwifruit’s soluble solid concentration using different regression techniques revealed interesting findings, as demonstrated in Table 2. Each employed method—PLSR, bagged trees, and TLNN—showed varying performances based on their R squared (R²) values and RMSE. The PLSR displayed a strong performance with an R² of 0.90, indicating that approximately 90% of the variation in the dependent variable was accounted for by the model. However, it also showed a moderate RMSE of 0.3479, signifying some deviation between the predicted and actual values (Figure 4b and Table 2).

Both bagged trees and ANN models exhibited an R² value of 0.91, indicating a strong correlation between the predictors and the target variable. This implies that these models are able to explain 91% of the variance in the data. However, bagged trees had an RMSE of 0.3700, slightly higher than the PLSR, while TLNN showed an RMSE of 0.3651, indicating a marginally less accurate prediction on average compared to the PLSR (Table 2). In predicting the soluble solid concentration of kiwifruit, all three models—PLSR, bagged trees, and TLNN—performed well, explaining a significant portion of the variation. The PLSR seemed to have a slight edge in predictive accuracy, closely followed by bagged trees and ANN (Figure 4c,d and Table 2).

3.3. Dry Matter (DM) Results

Based on the DM variance percentage depicted in Figure 5a, it was noticed that the first five components predict up to 70% of the model. Moreover, the curve shows a gradual increase and for n > 19, the prediction is greater than 90%.

![Figure 5](image-url)  
**Figure 5.** Dry matter (DM) content regression results regarding variance. (a) Regression scatter plots comparing the predicted vs. observed values for DM, derived from the PLSR; (b) bagged trees; (c) and TNN (d) models. The black line in the graphs (b–d) denotes the theoretical 1:1 line, while the yellow line denotes the fitted regression line.
The current assessment of kiwifruit dry matter using diverse regression methods revealed some interesting information. Each of the three employed approaches demonstrated varying performances, indicated by their relative $R^2$ and RMSE values. The PLSR model showcased respectable performance with an $R^2$ of 0.90, signifying that approximately 90% of the variance in the dependent variable was captured by the model. However, it displayed an RMSE of 0.3685, indicating a moderate level of deviation between the predicted and actual values (Figure 5b and Table 2). The bagged trees model exhibited an $R^2$ value of 0.87, indicating a strong fit to the data. However, the higher RMSE, equal to 0.7244, in comparison to the PLSR model indicates a less precise predictive performance (Figure 5c and Table 2). On the other hand, TLNN exhibited the highest $R^2$, equal to 0.93, suggesting a robust correlation between the predicted and the target value, capturing 93% of the variance. However, its RMSE of 0.5181, which is lower than that of bagged trees, still indicates a slightly higher level of error compared to the PLSR (Figure 5d and Table 2). For predicting the DM of kiwifruits, each method—PLSR, bagged trees, and TLNN—displayed distinctive performances. In general, the PLSR showcased a strong predictive capacity, followed by TLNN and bagged trees (Figure 5 and Table 2).

3.4. Tannin Results

Upon examination of Figure 6a, it is evident that the PLSR model prediction was observed for values of $n$ less than 20 at a rate that is below 70%. When $n$ is more than 32, the model’s successful prediction rate exceeds 90%. The curve appears to reach a stable state as the number of components increases, resulting in a minor improvement in the prediction percentage.

Analysis of tannin levels in kiwifruit using different regression techniques yielded varying results. The effectiveness of each applied strategy was assessed based on their $R^2$ values and root mean square error (RMSE), as presented in Table 2. The PLSR model demonstrated a solid performance with an $R^2$ of 0.91, capturing about 91% of the variation. Its RMSE was 0.3418, indicating some deviation between the predicted and actual values (Figure 6b and Table 2). Regarding the bagged trees algorithm, it displayed an $R^2$ equal to 0.88 and an RMSE 0.4141, indicating slightly lower accuracy compared to the PLSR (Figure 6c and Table 2). On the other hand, TLNN showcased the highest $R^2$ of 0.94, suggesting a strong correlation between the predicted and target variables, capturing 94% of the variance. Its RMSE of 0.2819 indicates a slight deviation between the predicted and observed values (Figure 6d and Table 2). Overall, in the evaluation of tannin levels in kiwifruits, artificial neural networks (ANN) have shown strong predictive capability, with partial least squares regression (PLSR), and bagged trees following closely behind.

Taking all the above into consideration, it is evident that all the utilized regression models demonstrate a significant ability to accurately predict the four investigated quality characteristics. The accuracy of these predictions ranges from 87% to 97%. The diverse regression models exhibited varying performances across the prediction of kiwifruit characteristics. Notably, notable predictions were achieved with the help of the three employed regression models, yielding high and constant predictive performances. Regarding the employed models’ efficiency, the PLSR has consistently shown strong correlations, while, on the other hand, demonstrating moderate variations in terms of prediction accuracy. Both bagged trees and TLNN yielded competitive results for certain attributes; however, they presented different levels of accuracy and error rates, recommending that the selection of the appropriate model should consider the trade-offs between predictive power, accuracy, and computational complexity (Figures 3–6 and Table 2).
Thus, accurately assessing variations in firmness over time may assist in developing postharvest handlings; for example, a dramatic softening could be observed in wounded areas 

Effective storage and marketing approaches for kiwifruit [18] require selecting areas 

The firmness of kiwifruit 

Both firmness of kiwifruit can be influenced by various factors, including the mineral composition, particularly the calcium level, which has a significant impact on the inherent quality of the fruit and its ripening process [38]. Furthermore, firmness is strongly influenced by both harvest and postharvest handlings; for example, a dramatic softening could be observed in wounded or even damaged (with no visual symptoms) kiwifruit during postharvest life [28,39]. In the current study, despite the above-mentioned severe difficulties in predicting the firmness of kiwifruits, the applied PLS models achieved optimal accuracies in predicting kiwifruit firmness, ranging from 90 to 97%. It is also worthy mentioning that the light scattering effect caused by the kiwifruit surface and the cell wall structure did not affect

Figure 6. Tannin content regression results regarding variance. (a) Regression scatter plots regarding tannin prediction comparing the predicted vs. observed values derived from the (b) PLSR; (c) bagged trees algorithm; and (d) TLNN algorithm models. The black line in the graphs (b–d) denotes the theoretical 1:1 line, while the yellow line denotes the fitted regression line.

4. Discussion

After a short period of 40 days of cold storage, four “Hayward” kiwifruit quality traits were destructively tested, namely firmness, soluble solid concentration (SSC), dry matter (DM), and tannin content, from fruit harvested from different orchards using a Specim IQ hyperspectral camera. The firmness of kiwifruit is a critical qualitative characteristic that plays a key role in determining its storage life and commercial viability. This is because the firmness of the fruit affects how long it can be stored after being harvested [37]. Thus, accurately assessing variations in firmness over time may assist in developing effective storage and marketing approaches for kiwifruit [18]. The firmness of kiwifruit can be influenced by various factors, including the mineral composition, particularly the calcium level, which has a significant impact on the inherent quality of the fruit and its ripening process [38]. Furthermore, firmness is strongly influenced by both harvest and postharvest handlings; for example, a dramatic softening could be observed in wounded or even damaged (with no visual symptoms) kiwifruit during postharvest life [28,39].
the performance of the applied models, since the PCA approach was performed for the attained reflectance data.

Two further significant indexes of kiwifruit quality are the content of soluble solids (SSC) and the proportion of dry matter (DM). Typically, kiwifruits have an SSC greater than 6.2% at the time of harvest [40], while a high dry matter content (over 16%) has been suggested as a predictor of quality in kiwifruit [3]. Furthermore, Hanker et al. [41] have suggested that SSC in ready-to-eat kiwifruits should fall between 10 and 14%. In a recent study, Titeli et al. [6] established an association between elevated levels of DM (dry matter), SSC (soluble solid content), and acidity, and a greater intensity of taste in kiwifruit. They emphasized the importance of these factors in determining the overall eating quality of kiwifruit. Non-destructive methodologies have been widely used to determine the quality features of kiwifruit, including firmness, SSC, and DM [2,13–19]. Therefore, these quality features significantly contribute to the overall quality and postharvest performance of kiwifruit. In the present study, the slightly lower performance in predicting the SSC can be attributed to the spectral range, as the data were acquired covering a spectral area from 400 to 1000 nm. It would be expected that more useful information for SSC prediction would be acquired once the data selection has been conducted within a wider range, covering up to 1450 nm. Since wavelengths at 1198 nm are associated with C-H single bonds, while the wavelengths near 1450 nm are correlated with O-H bonds [22], this could possibly yield useful information, enhancing the employed models’ performance regarding SSC prediction. On the other hand, the employed models’ performances achieved for DM were slightly worse compared to the other investigated quality traits, which could be attributed to possible variability in ripeness during the 40 days of cold storage.

Kiwifruit and persimmon fruits have been identified as high sources of tannin [42], and in addition to enhancing the antioxidant capacity of fruit, non-polymerized tannins in both fruits are associated with astringency that affects taste [43,44]. It has been supported that the high aftertaste intensity of kiwifruit was linked to a lower taste intensity [6], indicating a negative quality trait for consumption. Moreover, tannin content in fruit wines has been correlated with bitterness and astringency [45]. Therefore, the level of tannins in kiwifruit may be associated with the strength of the aftertaste, which is an important aspect in determining the overall flavor quality of the fruit. Compared to the DM prediction, the performances of the employed models regarding tannin prediction were slightly better.

The current study offers a comprehensive insight into the application of non-destructive technologies, particularly spectral and hyperspectral imaging, for the assessment of the above-mentioned kiwifruit quality traits. Vis/NIR spectroscopy represents a significant advancement over traditional destructive techniques used to assess internal quality traits such as soluble solid concentration (SSC), dry matter (DM), firmness, and tannins. This transition to non-destructive techniques aligns with the growing need for efficient, reliable, and rapid quality assessment in the postharvest industry [13,14,46–50].

The obtained results underscore the potential of Vis/NIR spectroscopy in accurately predicting the internal quality parameters of kiwifruit, a finding consistent with earlier studies on other fruits [51]. The use of principal component analysis (PCA), partial least square regression (PLSR), and artificial neural networks (ANN) further enhanced the predictive accuracy, each demonstrating unique strengths and limitations [51]. The performance of PLSR, in particular, stood out, demonstrating a strong correlation to the quality parameters while maintaining a balance between predictive power and computational complexity [52]. The TLNN model revealed superior predictive ability, most notably in terms of the R² and RMSE values (Table 2), indicating its effectiveness in capturing complex patterns within the data. This highlights the increasing importance of machine learning (ML) techniques in agricultural research which offer a more comprehensive understanding of fruit quality assessment. Future studies should focus on optimizing the employed models for specific quality attributes, considering the trade-offs between accuracy, efficiency, and computational demands.
The bagged trees model has found wide application in various fields due to its ability to improve model performance and reduce the effects of data variability and noise [23]. A minimum number of eight “leafs” (leaf size) was used, i.e., the algorithm takes into account at least eight samples in each decision tree. The number of trees was set equal to 30 “learners” which feed the algorithm with their decisions individually (Table 2).

The partial least squares regression (PLSR) model is well recognized and scientifically validated as a methodology for effectively identifying as well as accurately estimating the internal quality characteristics of kiwifruit [15–17,53]. Moreover, this study serves as a compelling testament to the enhanced efficacy and acumen showcased by neural networks in delivering superior performance when juxtaposed against their traditional statistical counterparts. As such, it was concluded that machine learning algorithms provide sufficient accuracy in predicting kiwifruit characteristics.

The current study involved a thorough examination of internal quality parameters in kiwifruits with the help of three discrete regression models. During these analyses, the reflected data obtained from the surface of kiwifruits underwent extensive analysis (Figure 2). These reflections were then thoroughly compared to the obtained laboratory measurements of four distinct quality characteristics: fruit firmness, SSC, DM, and tannin content (Figures 3–6). The main goal of the regression model was to precisely find and describe the strong correlations between the spectral data in each of the quality attributes (Table 2).

This intricate and nuanced interconnection facilitated the subsequent predictive models for these attributes, thoroughly grounded in the established relationships forged. The incorporation of advanced regression models like PLSR, bagged trees, and TLNN was pivotal in this study. Each model exhibited varying degrees of effectiveness in predicting different quality parameters, with PLSR consistently showing strong correlations across all parameters (Table 2). The TLNN model exhibited outstanding prediction accuracy and minimum error, highlighting the promise of ML approaches in agricultural applications. This conclusion is particularly remarkable (Table 2). The focus on the above critical internal quality parameters has highlighted the capability of artificial intelligence approaches to provide rapid, reliable, and efficient quality assessment, surpassing traditional destructive methods, demonstrating performances ranging from 87% to 97% (Table 2). All this has been achieved by astutely employing a judicious and optimal array of components (max = 32) drawn from a vast gamut of 198 distinct wavelengths examined across the visible and near-infrared spectra. Moreover, the predictive model tailored for determining DW showcased an upsurge in performance, manifesting the remarkable ability to deliver comparable prediction percentages while mandating fewer components, hovering approximately at 2/3 of the previously required quantity (Table 2).

5. Conclusions

The results of the present study confirm that the effective combination of a non-destructive approach using Vis/NIR spectroscopy and machine learning algorithms offers a promising alternative to traditional methods, providing a comprehensive understanding regarding “Hayward” kiwifruit quality assessment, especially for the qualitative firmness, SSC, DM, and tannin content after postharvest cold storage. Among the three employed ML algorithms, namely PLSR, bagged trees, and TLNN, the latter demonstrated the highest prediction capability, demonstrating performances ranging from 87% to 97%. The high efficiency of the TNN model demonstrated the remarkable capabilities in transforming quality assessment processes by offering several advantages such as rapidity, reliability, and efficiency, often surpassing traditional destructive techniques. The ML techniques have not only been proven capable of improving the efficiency of quality evaluation, but also aid in more effective resource allocation and reduction in food waste through precise sorting and grading. The effective integration of spectral imaging and machine learning techniques can revolutionize quality assessment in the postharvest sector. The current approach has significant implications for the entire supply chain, from growers and
processors to retailers and consumers, ensuring the provision of high-quality kiwifruits. Subsequent studies should prioritize the enhancement of these models, investigating their suitability for a broader variety of fruits. In addition, it is important to focus on integrating these nondestructive approaches into real-time, on-field quality assessment systems. This integration has the potential to significantly enhance the efficiency and sustainability of postharvest operations.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/horticulturae10030251/s1, Table S1. Kiwifruit quality traits of the 20 orchards evaluated.


Funding: This work was supported by E.P. Competitiveness, Entrepreneurship & Innovation (EPANEK), within the framework of the national action “RESEARCH-CREATE-INNOVATE” (ESPA 2014–2020) (project code: TZEAK-03007; Premium Kiwi).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References
6. Titeli, V.S.; Michailidis, M.; Tanou, G.; Molassiotis, A. Physiological and Metabolic Traits Linked to Kiwifruit Quality. Horticulturae 2023, 9, 915. [CrossRef]


43. Kim, J.G.; Beppu, K.; Kataoka, I. Varietal differences in phenolic content and astringency in skin and flesh of hardy kiwifruit resources in Japan. Sci. Hortic. 2009, 120, 551–554. [CrossRef]
53. Afonso, A.M.; Antunes, M.D.; Cruz, S.; Cavaco, A.M.; Guerra, R. Non-destructive follow-up of ‘Jintao’ kiwifruit ripening through VIS-NIR spectroscopy—Individual vs. average calibration model’s predictions. Postharvest Biol. Technol. 2022, 188, 111895. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.