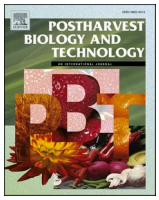
Postharvest Biology and Technology 199 (2023) 112281

Contents lists available at ScienceDirect

Postharvest Biology and Technology

journal homepage: www.elsevier.com/locate/postharvbio



Estimation of soluble solids content and fruit temperature in 'Rocha' pear using Vis-NIR spectroscopy and the SpectraNet–32 deep learning architecture

J.A. Martins [a,*], D. Rodrigues [a], A.M. Cavaco [a], M.D. Antunes [a], R. Guerra [a,b]

[a] Center for Electronic, Optoelectronic and Telecommunications (CEOT), Campus de Gambelas, University of Algarve, Faro, Portugal
[b] Physics Department, FCT, Campus de Gambelas, University of Algarve, Faro, Portugal

ARTICLE INFO

Keywords:
Deep learning
Residual network
Near-infrared
Spectroscopy
Pear fruit

ABSTRACT

Spectra-based methods are becoming increasingly important in Precision Agriculture as they offer non-destructive, quick tools for measuring the quality of produce. This study introduces a novel approach for estimating the soluble solids content (SSC) of 'Rocha' pears using the SpectraNet–32 deep learning architecture, which operates on 1D fruit spectra in the visible to near-infrared region (Vis-NIRS). This method was also able to estimate fruit temperatures, which improved the SSC prediction performance. The dataset consisted of 3300 spectra from 1650 'Rocha' pears collected from local markets over several weeks during the 2010 and 2011 seasons, which had varying edaphoclimatic conditions. Two types of partial least squares (PLS) feature selection methods, under various configurations, were applied to the input spectra to identify the most significant wavelengths for training SpectraNet–32. The model's robustness was also compared to a similar state-of-the-art deep learning architecture, DeepSpectra, as well as four other classical machine learning algorithms: PLS, multiple linear regression (MLR), support vector machine (SVM), and multi-layer perceptron (MLP). In total, 23 different experimental method configurations were assessed, with 150 neural networks each. SpectraNet–32 consistently outperformed other methods in several metrics. On average, it was 6.1% better than PLS in terms of the root mean square error of prediction (RMSEP, 1.08 *vs.* 1.15%), 7.7% better in prediction gain (PG, 1.67 *vs.* 1.55), 3.6% better in the coefficient of determination (R^2, 0.58 *vs.* 0.56) and 5.8% better in the coefficient of variation ($CV_{\%}$, 8.35 *vs.* 8.86).

1. Introduction

The current society is experiencing a significant shift in agriculture, with the principles of *Precision Agriculture* and *Agriculture 5.0* gaining increasing traction. These concepts are centered on the efficient management of natural resources at various interconnected levels, from production to consumption. When it comes to fresh fruit, quality control assessment tends to have strict tolerances, often regulated by government legislation that controls fruit quality at harvest (*e.g.*, regulating Protected Designation of Origin fruit) or logistics chain specificities (*e.g.*, cold storage requirements). The ultimate goal is to maximize the perceived quality of the fruit by consumers.

Research in this field often focuses on determining the Optimal Harvest Date (OHD) of specific fruit, which helps to reduce spoilage by minimizing the time between harvest and consumption and improving the perceived freshness of the fruit. OHDs are typically estimated by evaluating various internal quality attributes (IQAs) or properties of the fruit, and there is an increasing use of non-destructive assessment techniques for this purpose, either on the field or in grading machines. One of the most promising techniques for this purpose is Visible-Near diffuse Infrared Reflectance Spectroscopy (Vis-NIRS, 400–2500 nm), as machine learning methods based on this range of spectral data have been shown to be accurate and fast predictors of several IQAs related to OHD optimization.

Classical hyperspectral imaging (HI) illuminates the entire fruit and acquires an image at each wavelength. However, the information about the fruit's pulp that is obtained through this method is largely overshadowed by information about the fruit's skin, as superficial reflection

* Corresponding author.
E-mail addresses: jamartins@ualg.pt (J.A. Martins), dacr@uevora.pt (D. Rodrigues), acavaco@ualg.pt (A.M. Cavaco), mantunes@ualg.pt (M.D. Antunes), rguerra@ualg.pt (R. Guerra).

https://doi.org/10.1016/j.postharvbio.2023.112281
Received 12 October 2022; Received in revised form 18 January 2023; Accepted 30 January 2023
Available online 13 February 2023

from the skin is much stronger than diffuse reflection from the pulp. As a result, classical HI is primarily useful for detecting and classifying external defects and pathologies, but is not well-suited for predicting internal quality parameters of the fruit. Most models in the literature that attempt to use HI data to predict internal quality parameters rely on indirect correlations between the skin pigments and the internal quality of the pulp, and are generally developed using uniform batches of samples, which may not be robust when applied to new batches of fruit. In order to accurately predict internal quality parameters of fruit using optical methods, it is important to ensure that the light used to probe the fruit's pulp is collected by the system and that any photons reflected by the skin or other external parts of the fruit are not included. One way to do this is by using classical point spectroscopy with a reflection probe that has well-separated injection and collection fibers. This setup ensures that the light collected by the system consists primarily of photons that have passed through the pulp, and is therefore more reliable for predicting internal quality parameters than classical HI, which relies on superficial reflection from the skin. A detailed review on this subject can be read in Li et al. (2018), Walsh et al. (2020), Cavaco et al. (2022). Nevertheless, there is another type of hyperspectral imaging (HI) setup, which we can refer to as *spot-HI*, that was first described by Lu and Peng (2006). Spot-HI involves illuminating a small beam spot on the fruit's surface and imaging the diffuse scattering halo around the incidence point. This technique is mostly sensitive to photons that emerge from the pulp of the fruit and satisfies the requirement of detecting only pulp photons. In addition, spot-HI provides spatial information about the light intensity of the scattering halo, which makes it theoretically superior to point spectroscopy. However, the best results obtained using spot-HI are similar to the best results obtained using point spectroscopy, meaning that spot-HI has not been shown to be actually better than point spectroscopy in practice. The main issues with spot-HI are stray light in the beam spot and the non-spherical shape of the fruit surface, which can hinder the retrieval of spatial information. Additionally, the setup for spot-HI is more expensive than that of point spectroscopy, making it a less practical or theoretically compelling choice for predicting fruit internal quality.

Classical machine learning approaches, such as Partial Least Squares (PLS) regression, have traditionally achieved good results in predicting internal quality attributes (IQAs) using one-dimensional (1D) spectral data. However, they require a case-by-case, trial-and-error process for selecting the best data preprocessing methods for each specific dataset before the regression algorithm can be applied. In contrast, deep learning approaches have mainly focused on fruit detection and/or categorization in trees (*e.g.*, estimating fruit counts), which is a different problem. There is a lack of research on developing generalized regression architectures specifically for IQA prediction using 1D spectral data. Some recent studies have used hyperspectral imaging approaches Yu et al. (2018), Benelli et al. (2020), but these are expected to be complementary methodologies as most hyperspectral images only capture fruit peel data (Benelli et al., 2020). In contrast, Vis-NIR point spectroscopy is designed to capture internal information of fruit, as demonstrated in a recent study on predicting internal browning in 'Rocha' pear samples using semi-transmittance spectra (Cruz et al., 2021).

Mishra and Passos (2021) developed a deep multi-block regression method for evaluating dry matter in mango fruit by dividing the visible (Vis) and near-infrared (NIR) parts into separate blocks, outperforming a single-block CNN as well as sequential and orthogonalized partial least-squares (SO-PLS) regression.

Zhang et al. (2019) introduced DeepSpectra, a novel architecture for predicting internal quality attributes (IQAs) based on the Inception module developed for GoogLeNet (Szegedy et al., 2015). It was tested on 1D spectral data using four Vis-NIRS datasets: corn, pharmaceutical tablets, wheat, and soil patches. It outperformed three simpler CNN models and PLS results.

Martins et al. (2022) proposed the SpectraNet–53 architecture, which demonstrated promising results for determining the Soluble Solids Content (SSC) in a dataset of 'Newhall' oranges with a low sample count (616 spectra). The authors also described the various steps and considerations involved in creating the architecture. The findings of this research served as the foundation for the current study, as it outperformed Partial Least Squares (PLS) regression, a classical machine learning technique that has been successfully used for this type of problem.

Fruit temperature was also an important factor to consider. As demonstrated by Kaur et al. (2022), rising temperatures tend to broaden and shift absorbance peaks to shorter wavelengths (as molecules move faster) with an accompanying intensity shift that may be higher or lower depending on the spectral region. These changes in spectral shape and characteristics can complicate accurate SSC prediction. If a method is able to infer fruit temperature from spectra, some of this effect can be accounted for, leading to improved generalization.

It is essential to consider that certain wavelengths are more important than others when analyzing spectra for specific prediction models. The informativeness of spectra, the presence of inter-correlated features that repeat information, and the potential for hindering regression performance for specific internal quality attributes (IQAs) can all vary depending on the substance being analyzed. To address these issues, feature selection on the input data is critical for helping machine learning models identify the most informative features for a particular fruit attribute. Partial Least Squares (PLS) regression is a reliable technique for analyzing spectral data, and several feature selection methods based on this method have been developed in the literature (Mehmood et al., 2012). These methods can help improve the accuracy and efficiency of prediction models by focusing on the most relevant features in the data.

The SpectraNet–53 architecture, which was first introduced in our research on oranges (Martins et al., 2022), has not been widely evaluated as a reliable method in the literature. Prior to this study, it had only been applied to one specific use-case, and its effectiveness on other types of fruit, such as pears, was uncertain. Additionally, SpectraNet–32, which has 40% fewer layers than SpectraNet–53, is a distinct architecture that operates on feature-selected data, and had not been previously tested. To the best of our knowledge, SpectraNet represents the first residual architecture applied to Vis-NIR fruit spectra.

In summary, this article presents a comparison of the performance of DeepSpectra (Zhang et al., 2019) and a shallower configuration of SpectraNet (Martins et al., 2022) with 32 layers, referred to as SpectraNet–32, in conjunction with two PLS–based input feature selection methods on a dataset of 'Rocha' pears. The goal of this study was to reduce input and architecture complexity, decrease the training time of the network, and improve generalization performance by focusing on important input features. The main contributions of this article are thus:
.

1. Evaluating the performance of three deep learning approaches (DeepSpectra, SpectraNet–32, and SpectraNet–53) on a 'Rocha' pear dataset for predicting SSC and temperature.
2. Assessing the role of two PLS–based input feature selection methods on the three deep learning approaches for the same dataset.
3. Comparing the performance of SpectraNet–32 to previously published classical machine learning results on the same dataset (Passos et al., 2019).
4. Achieving state-of-the-art performance results for SSC prediction on an external validation dataset.

The organization of the article is as follows: In Section 2, the materials and methods used for the 'Rocha' pear dataset creation are described, including the spectra collection, processing, and measurement of fruit attributes. In Section 2.4.1, the DeepSpectra network configurations used in training and for performance comparisons are detailed. In Section 2.4.2, an overview of the SpectraNet architecture is

presented, and the feature selection methodologies and hyperparameter configurations are described in detail. In Section 3, all performance assessments are presented and discussed, with comparisons to other methods. In Section 4, the main achievements and contributions are highlighted and the article is concluded.

2. Materials and methods

The 'Rocha' pear is a variety of fruit that is grown in the western region of Portugal and has gained Protected Designation of Origin (PDO) status. It is a key commodity for both the national and export markets in Portugal, with an annual production of around 173,000 metric tons (ANP, 2022). In order to preserve its high quality and desirable organoleptic characteristics, such as flavor, smell, shape, texture, and color, 'Rocha' pears must be harvested at the optimal ripening stage, which typically falls between late July and mid-August. PDO legislation in Portugal sets specific requirements for the optimal harvest date (OHD) of 'Rocha' pears, including requirements for pulp firmness, soluble solids content (SSC), acidity, days past full flowering, and seed color. To meet these requirements, 'Rocha' pears must have a pulp firmness between 5.5 and 6.5 Kg/0.5 cm^2, SSC between 11 % and 13 %, acidity between 2 and 3 g/L of malic acid, at least 135–140 days past full flowering, and an appropriate brown seed color. These criteria are established in order to ensure that the 'Rocha' pear maintains its high quality and is suitable for consumption.

2.1. Fruit processing: evaluating temperature, spectra and soluble solids content (SSC)

For this dataset, 1650 pears were acquired from local markets, in batches with *circa* 50 pears each, over many weeks, as described in Passos et al. (2019). These were processed at three different temperature ranges: (a) after storage in a cooling chamber (885 fruit); (b) at room temperature (555 fruit); and (c) after heating (210 fruit). The goal was to replicate the typical real-world temperatures at which these fruit are measured: in cold storage, after harvest or while in the tree, respectively. Thus, for each fruit: .

1. The peel temperature was registered with an infrared thermometer (Fluke FoodPro Plus, Everett, WA, USA), at the equatorial zone (Passos et al., 2019);
2. Two spectra measurements were taken, 180° apart, along the equator. The hardware configuration used for this task is described below;
3. The two probed regions were cut out and the pulp squeezed. A few drops of juice were deposited on a digital refractometer (Atago Model PAL-1, Atago Co. Ltd., Tokyo, Japan), thus obtaining two for SSC measurements per fruit. Histograms are represented in Fig. 1(a) for SSC, and Fig. 1(b) for the three temperature ranges. The mean SSC value was 12.95% with a standard deviation of 1.75%.

In total the 1650 pears resulted in 3300 spectra, with 1024 wavelengths each, between 432.6 and 1146.74 nm.

We also measured the firmness of the pears, but our showed that it is difficult to predict with a high degree of accuracy. As a result, we have not included the firmness data in our analysis. Firmness assessment in fruit remains a significant challenge in postharvest research. The traditional method of assessment is destructive, involving the insertion of a penetrometer into the fruit. In recent years, there have been many claims for the non-invasive assessment of fruit firmness using Vis-NIRS. However, despite being a popular topic in the literature, with approximately 10% of all papers published in the period 2015–2020 involving the assessment of firmness, there is still no consensus on the robustness of using Vis-NIRS for direct firmness assessment. Firmness is mainly a physical parameter, and no obvious chemical bands may be assigned to it. While firmness changes are associated with biochemical processes, such as changes in pectin levels, this correlation is faint and easily masked by other processes occurring during ripening. It is generally believed that the scattering coefficient of the pulp tissue is a good measure of firmness, but this has not proven to be a robust relation. Additionally, firmness levels are correlated to a range of other attributes, such as water content, pigment level and starch-sugar conversion during ripening, providing fertile ground for secondary Vis-NIRS correlations. Thus, the non-invasive assessment of fruit firmness remains a 'holy grail' in postharvest research, and further investigations are needed to establish a reliable method for assessing firmness using Vis-NIRS (Walsh et al., 2020).

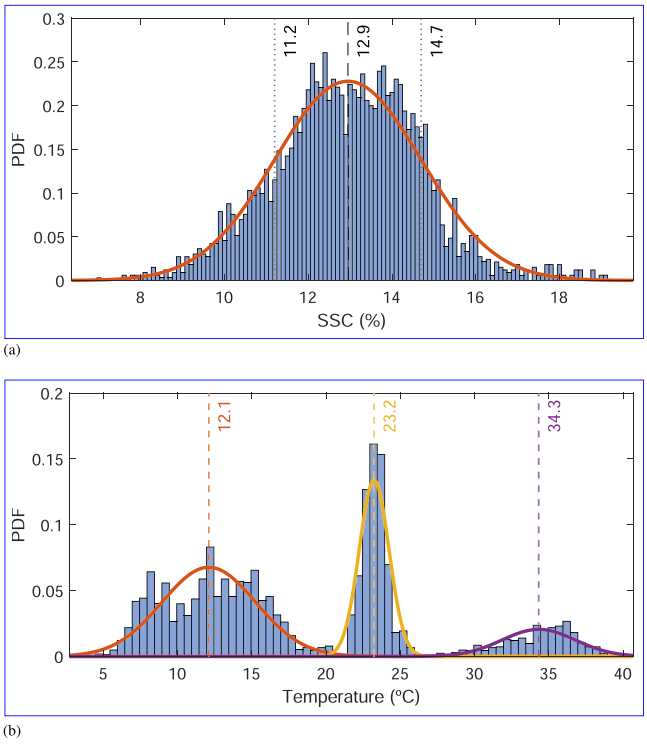


Fig. 1. Probability Density Function (PDF) histograms for SSC and temperature, of all the pear samples in the dataset. (a) SSC (%) histogram. The dashed line represents the mean value and the dotted lines the standard variation around the mean. (b) Temperature histogram of the 3 major clusters ($T < 19$°C, $19 \leq T \leq 27$°C, and $T > 27$°C), with 885, 555 and 210 fruit, respectively. The dashed lines represent cluster means.

Spectra were collected using two multimode optical fibers, with a 600 μm core and 0.22 NA, coupled to collimating optics, were used for illuminating and collecting light from the fruit. The illumination fiber was bifurcated, with the solo tip connected to a white halogen light (HL-2000-FHSA, Ocean Optics, USA), and the bifurcated tips were mounted 5 cm apart at a 5° vertical angle, so that the two collimated beams would converge. A collection fiber was positioned between the two, at a 0° vertical angle, aimed at the region between the illumination spots. All fibers were fixed at 170 mm above the base of the cup, where fruit were measured, as to replicate a real-world scenario typical of an industrial fruit sorting machine. Spectra along the equatorial region and Vis-NIR range of 432–1147 nm were recorded (at a 5 nm spectral resolution) with a Hamamatsu TG-9405CA spectrometer (Hamamatsu, Shizuoka, Japan), during 50 ms. For the absolute reference material, a disk of white Teflon was used. Before each acquisition, a dark spectrum was measured and used for correcting each sample's raw photon count. For more details, please refer to Passos et al. (2019).

Fig. 2 shows the reflection spectra (X_R) converted into an absorbance-like transform (X_A) values, which is done by: **(a)** truncating negative values to zero, *i.e.*, $X_R \in [0, \infty)$, and **(b)** computing $X_A = -\log_{10}(X_R + 0.1)$, so that $X_A \in (-\infty, 1]$. This differs from the usual absorbance formula by the addition of a 0.1 constant, resulting in

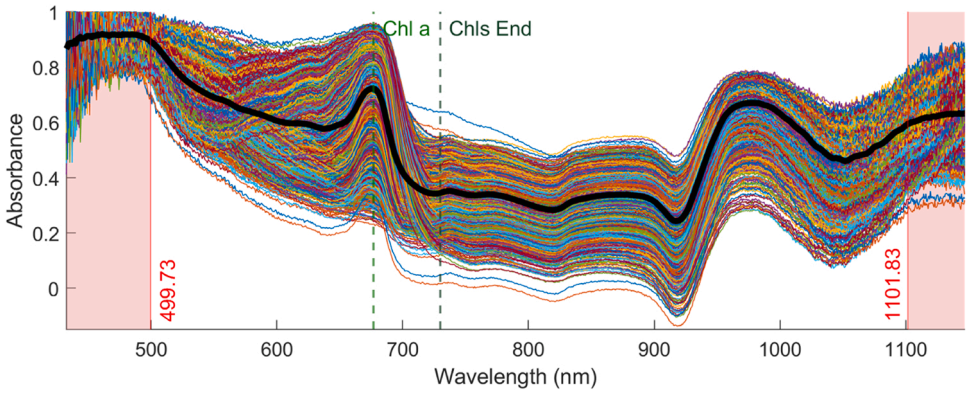


Fig. 2. Absorbance spectra of the 1650 pears (3300 samples). Median wavelength in black. Vertical solid red lines delimit the excluded wavelength ranges at both extremes, which are not used in the model (details in §2.1). The vertical dashed lines mark the Chlorophyll *a* absorption peak at 677 nm and the end of Chlorophylls absorption influence at 730 nm.

$\max(X_A) = -\log_{10}(0.1) = 1$. Also, at $X_R = 1$, $X_A = -\log_{10}(1.1) = -0.0414$. This transform limits the exponential scaling of low reflectance values into infinite absorbance, resulting in a more linear relation between adjacent low reflectance values (especially below 0.2), and helps to stabilize weights at the first convolutional layer during network training.

The first and last 75 points of each spectra were trimmed: below 499.73 nm and above 1101.83 nm. These correspond to zones with a very high noise floor, and thus lower SNR; they are represented as red shaded regions in Fig. 2. This also allows for a direct comparison with the different optimization models presented in Passos et al. (2019), which use the exact same wavelengths.

2.2. Dataset partition

The 3300 samples were divided into five sets, with sets A through C from the 2010 harvest season and D–E from the 2011 season, which is an identical split to Passos et al. (2019). These will serve as external validation (EV) sets between themselves, as each correspond to a non-overlapping time period, in five different combinations, as shown in Table 1. Of the two properties measured, the goal is to achieve the best possible SSC prediction, represented in Fig. 3, as it is a crucial attribute for both harvest date prediction and sweetness perception. Fig. 4 highlights the statistical differences between the five EV sets, which are expected due to the variation in fruit sources, seasons and inherent seasonal variability; the soil types are different across batches, as well as the local and general climatic conditions (*e.g.*, colder *vs.* warmer seasons/years). In the spring of 2011, the weather in the region of Bombarral (Portugal), where these fruit were sourced, had an average low temperature of 17˚C, average high temperature of 24˚C, and average relative humidity (RH) of 64%. In the spring of 2012, the average low temperature was 15˚C, average high temperature of 22˚C, and RH of 62%, resulting in a colder weather during the months when pears grow the most.

2.3. Feature selection

This work compares two PLS-based feature selection methodologies, for selecting the most informative wavelengths for network training: .

Table 1
Combinations of train/test pairs for external validation.

Number of training sets	Possible combinations	
5	$C_4^5 \times 1 = 5$	ABCD/E, ABCE/D, ABDE/C, ACDE/B, BCDE/A

1. The first is a PLS *wrapper* method (Passos et al., 2019), based on backward variable elimination (BVE-PLS) (Mehmood et al., 2012). A fixed number of 10 latent variables was chosen for all trials. At the start, one PLS model is built using all available features (*i.e.*, wavelengths) and the root mean square of 5-fold cross-validation (rmse_{cv}) is computed. Afterwards, and at every iteration, the wavelength with the lowest absolute PLS regression coefficient is removed, the model is recalculated and rmse_{cv} is assessed. The process continues until rmse_{cv} increases, which signals that all remaining features are important. Fig. 5 shows an example of the bands kept and discarded using this method when using as input a second-order derivative spectra of the present dataset (Savitzky-Golay, 2nd polynomial order and 51-pt window). As CNNs can explore spatial relations of the input data, this method was also tested under two different conditions, with the selected features organized in ascending order either by:
   (a) Their PLS regression coefficient amplitude (*i.e.*, importance order); or
   (b) Their physical wavelength value.
2. The second is a PLS *filtering* method, based on the variable importance in projection (VIP) scores (Abdi, 2010). This is a single-pass method, much quicker than the previous one, as PLS only needs to be computed once. VIP scores are assessed and only variables (*i.e.*, wavelengths) equal to, or higher, than a specific threshold are kept (Gosselin et al., 2010). This filtering method was tested with three different VIP thresholds, at 0.8, 1.0 and 1.2, which should account for the threshold-choosing issues discussed in Chong and Jun (2005). The features were kept in their physical wavelength order.

2.4. Neural network architectures

2.4.1. DeepSpectra

DeepSpectra (Zhang et al., 2019) was chosen as a reference for performance comparisons with SpectraNet–32. It is a similar deep learning architecture for quantitative spectral analysis, originating from the Inception model (Szegedy et al., 2015). For this work, the architecture was implemented and run with hyperparameters similar to the ones used by the authors on the "Wheat" dataset, used to estimate protein content: Zhang et al. (2019) trained DeepSpectra on 775 wheat samples from 7 crop years, and tested on 107 samples from a single crop year. The spectra ranged from 400 to 2498 nm, on a 2 nm resolution, for a total of 1050 wavelengths. This was a dataset with a close number of features to the one described in this article (which had 874 wavelengths after cropping), although the spectral range was much broader for the wheat samples (400–2498 nm). In summary, the hyperparameter configuration used is: **kernel size 1**: 7-pts, **kernel size 2**: 3-pts, **kernel size 3**: 5-pts, **stride 1**: 3-pts, **stride 2**: 2-pts, **hidden number**: 32

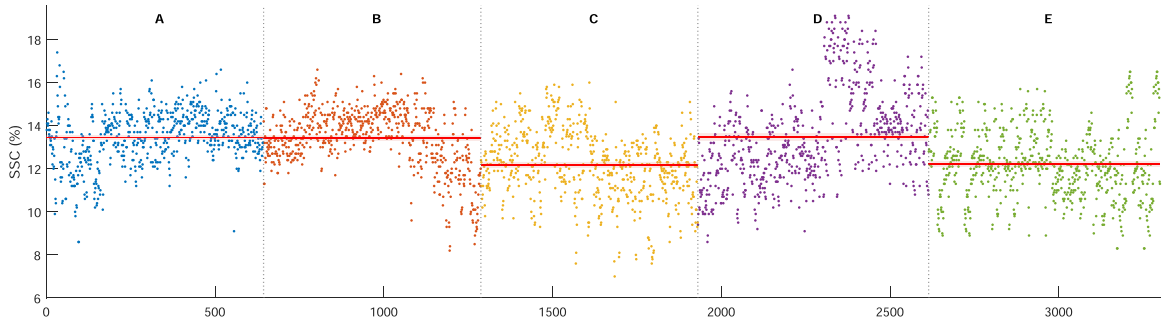


Fig. 3. SSC (%) for the 3300 samples in the dataset, divided into the five sets. Horizontal red lines represent the mean values of each set, with shaded 95% CI.

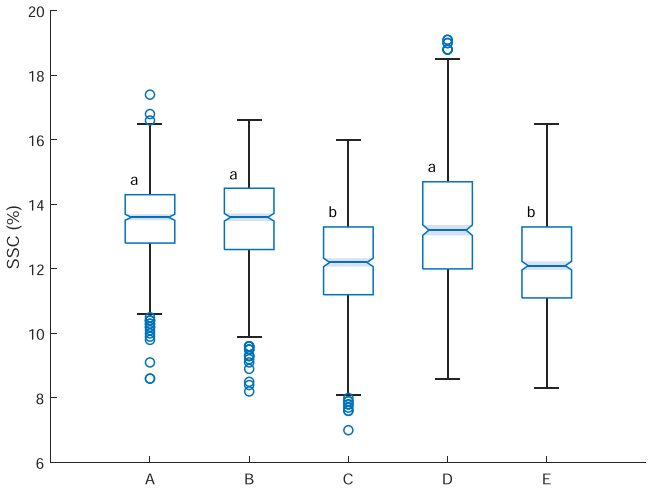


Fig. 4. SSC (%) for the 5 EV sets, with shaded 95% confidence intervals. Statistical significance of the differences was assessed with a one-way ANOVA [$F(4, 3295) = 115.61$, $p < 10^{-93}$]. A Tukey post-hoc test revealed significant pairwise differences between any element of the set {A,B,D} with any element of {C,E} ($p < 10^{-4}$), but not within-set.

neurons, **mini-batch size**: 128, **dropout rate**: 10%, **regularization coefficient**: $\lambda = 0.01$, **learning rate**: 1×10^{-2}, with a decay of 1×10^{-3} ($-$ 0.1%) at each epoch, and **epochs**: 10.

Two variations of the DeepSpectra architecture were used: one with a single output variable (SSC), referred below as *DeepSpectra*, and another with two outputs (SSC and fruit temperature), referred below as *DeepSpectra2*.

2.4.2. SpectraNet–32

SpectraNet is a robust deep regression residual network for spectra processing, designed for SSC estimation and able to be trained on a low amount of data. In a previous work by Martins et al. (2022), the SpectraNet architecture was used in a 53 layer configuration, with six Residual Unit (RU) blocks (He et al., 2016), which the current work reduces to just three, for a total of 32 layers. A performance comparison of both architecture configurations is presented in Section 3. This configuration can be trained much quicker than SpectraNet–53, as it has almost half as many optimizable parameters (0.7 M *vs* 1.3 M), which is also a bonus to inference performance.

The complete architecture for SpectraNet–32 is shown in Table 2, with a total of 702,276 trainable parameters. Each row is a consecutive layer that is connected to the one above and below, unless stated otherwise, as is the case for skip connections. Batch Normalization (BN) is directly applied to the input vector, at L1, effectively functioning as a learnable input standardization procedure and as a form of data augmentation (Simon et al., 2016). Each RU then starts with a convolutional layer, followed by BN and a Gaussian Error Linear Unit (GELU) activation (Hendrycks and Gimpel, 2016), which has several desirable properties as an activation function – the transfer function is the expected transformation of a stochastic regularizer, which facilitates network generalization by reducing overfit. The shortcut connections also use two types of identity mappings (He et al., 2016): (a) a direct path from L4 to L10 or (b) through 1×1 convolutional shortcuts on RUs two and three, as the network is designed to spatially downsample activation widths while upsampling channels.

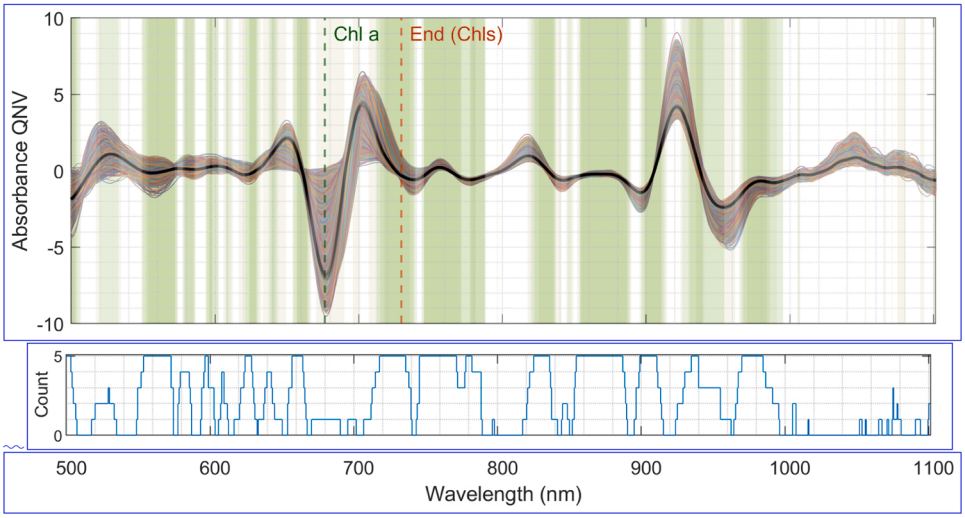


Fig. 5. *(top)* The *Savitzky-Golay* second-order derivative (2nd degree polynomial, 51-pt window) of the pear spectra, overlaid for the 5 EV sets. Selected features are shaded in green (white bands are discarded). Darker greens represent features selected in more than one set. Median wavelengths in black. The vertical dashed lines mark the Chlorophyll *a* absorption peak at 677 nm and the end of the Chlorophylls absorption influence at 730 nm. *(bottom)* The occurrence of selected features across the 5 EV sets, by wavelength.

Table 2
SpectraNet–32 Architecture.

Layer	Type	Activations	Learnable Parameters		Total Param.
	Input	1 × N × 1	—		0
1	Batch Normalization	1 × N × 1	Offset: 1 × 1	Scale: 1 × 1	2
2	Convolution	1 × N × 32	Weights: 1 × 21 × 1 × 32	Bias: 1 × 1 × 32	704
3	Batch Normalization	1 × N × 32	Offset: 1 × 1 × 32	Scale: 1 × 1 × 32	64
4	GELU	1 × N × 32	Mean: 1 × 1 × 32	Std: 1 × 1 × 32	64
5	S1U1 Convolution	1 × N × 32	Weights: 1 × 21 × 32 × 32	Bias: 1 × 1 × 32	21.536
6	S1U1 Batch Normalization	1 × N × 32	Offset: 1 × 1 × 32	Scale: 1 × 1 × 32	64
7	S1U1 GELU	1 × N × 32	Mean: 1 × 1 × 32	Std: 1 × 1 × 32	64
8	S1U1 Convolution	1 × N × 32	Weights: 1 × 21 × 32 × 32	Bias: 1 × 1 × 32	21.536
9	S1U1 Batch Normalization	1 × N × 32	Offset: 1 × 1 × 32	Scale: 1 × 1 × 32	64
10	Addition (L9 + L4)	1 × N × 32	—		0
11	GELU	1 × N × 32	Mean: 1 × 1 × 32	Std: 1 × 1 × 32	64
12	S2U1 Convolution (stride 2)	1 × (N/2) × 64	Weights: 1 × 21 × 32 × 64	Bias: 1 × 1 × 64	43072
13	S2U1 Batch Normalization	1 × (N/2) × 64	Offset: 1 × 1 × 64	Scale: 1 × 1 × 64	128
14	S2U1 GELU	1 × (N/2) × 64	Mean: 1 × 1 × 64	Std: 1 × 1 × 64	128
15	S2U1 Convolution	1 × (N/2) × 64	Weights: 1 × 21 × 64 × 64	Bias: 1 × 1 × 64	86.080
16	S2U1 Batch Normalization	1 × (N/2) × 64	Offset: 1 × 1 × 64	Scale: 1 × 1 × 64	128
17	*Skip Conn.* | Convolution (input L11, stride 2)	1 × (N/2) × 64	Weights: 1 × 1 × 32 × 64	Bias: 1 × 1 × 64	2.112
18	*Skip Conn.* | Batch Normalization	1 × (N/2) × 64	Offset: 1 × 1 × 64	Scale: 1 × 1 × 64	128
19	Addition (L16 + L18)	1 × (N/2) × 64	—		0
20	GELU	1 × (N/2) × 64	Mean: 1 × 1 × 64	Std: 1 × 1 × 64	128
21	S3U1 Convolution (stride 2)	1 × (N/4) × 128	Weights: 1 × 21 × 64 × 128	Bias: 1 × 1 × 128	172.160
22	S3U1 Batch Normalization	1 × (N/4) × 128	Offset: 1 × 1 × 128	Scale: 1 × 1 × 128	256
23	S3U1 GELU	1 × (N/4) × 128	Mean: 1 × 1 × 128	Std: 1 × 1 × 128	256
24	S3U1 Convolution	1 × (N/4) × 128	Weights: 1 × 21 × 128 × 128	Bias: 1 × 1 × 128	344.192
25	S3U1 Batch Normalization	1 × (N/4) × 128	Offset: 1 × 1 × 128	Scale: 1 × 1 × 128	256
26	*Skip Conn.* | Convolution (input L20, stride 2)	1 × (N/4) × 128	Weights: 1 × 1 × 64 × 128	Bias: 1 × 1 × 128	8.320
27	*Skip Conn.* | Batch Normalization	1 × (N/4) × 128	Offset: 1 × 1 × 128	Scale: 1 × 1 × 128	256
28	Addition (L25 + L27)	1 × (N/4) × 128	—		0
29	GELU	1 × (N/4) × 128	Mean: 1 × 1 × 128	Std: 1 × 1 × 128	256
30	Global Average Pooling	1 × 1 × 128	—		0
31	Dropout (25%)	1 × 1 × 128	—		0
32	Fully Connected	1 × 1 × 2	Weights: 2 × 128	Bias: 2 × 1	258
	MSE evaluation	—	—		0

Notes. N is the length of the input vector. Non-integer N/2 or N/4 values are rounded towards positive infinity (*i.e.*, the ceiling). More info in 2.5.

2.5. Overview

The next paragraphs briefly summarize all the data processing steps taken and neural network hyperparameters chosen for training. For preprocessing the raw spectra:

– As previously stated in 2.1, the raw reflectance spectra were converted into absorbance-like values, such that $x_A = log_{10}\left(\frac{1}{x_R} + 0.1\right)$.
– The absorbance-like data was normalized using a Quantile Normal Variate (QNV) transform (Martins et al., 2022) with 15 quantiles. This is an alternative to the Standard Normal Variate (SNV) transform, which can be helpful to reduce the skew effect of low-percentile and high-percentile noisy data, by using quantiles as representative samples of the underlying signal. QNV approaches SNV as the number of quantiles increases to "infinity", *i.e.*, a quantile for each data point.
– Savitzky-Golay smoothing was used on the normalized data, with similar parameters to Passos et al. (2019): 1st or 2nd order derivative, 2nd polynomial order, 51 point window (about 35.0 nm).
– The smoothed spectra are used as inputs for either the PLS wrapper method or PLS filter method, previously described in 2.3. The selected features for the first are shown in Fig. 5, and detailed below.

Either DeepSpectra or SpectraNet are then trained on the preprocessed data. For DeepSpectra, its hyperparameters were presented in 2.4.1, while regarding SpectraNet the following hyperparameters were used:

– Standard ADAM was used, with default parameters.
– 1×10^{-3}, with a drop multiplier of 0.90 (− 10%) at each epoch.
– All networks were trained during 15 epochs, which was enough for the training error to stabilize.
– The training mini-batch size was set at 128.
– A 21-pt filter window was used in all convolutional layers.
– The dropout layer at L32 had a drop probability value of 25%.
– With a coefficient of $\lambda = 0.05$. L2 Regularization (also designated as *weight decay*) adds a regularizing term to the weights of the loss function, reducing possible overfitting by a large gradient descent update. If $L(x)$ is the expected loss function, then $L_R(x) = L(x) + \lambda\Omega(W)$, with $\Omega(W) = \frac{1}{2}W^T W$.
– Another overfitting prevention measure. If the global L2-norm (L_2^g) of all the gradients of learnable hyperparamenters is higher than 0.8, then all are scaled by a factor of $0.8/L_2^g$. This avoids gradient explosion by stabilizing and allowing training at initial higher learning rates, while reducing the training vulnerability to outlier gradients.
– Convolutional layers are initialized with *He* weights (He et al., 2015), and the final fully connected layer (L33) is initialized with *Glorot* weights (Glorot and Bengio, 2010).
– The networks were trained on two output configurations, referred as either *SpectraNet* with only SSC as output, or *SpectraNet2* with two outputs, SSC and temperature.
The output data was normalized using a Quantile Normal Variate transform (Martins et al., 2022) with 15 quantiles.

Architectural changes between sets

Network layer activations vary between EV sets, as each set has a different number of input features, due to the feature selection procedure (Fig. 5). Regarding Table 2, for the PLS wrapper method, these are:

378 features selected (*N*) and 496 discarded. The activations are: $1 \times 378 \times 1$ (L1), $1 \times 378 \times 32$ (L2–L11), $1 \times 189 \times 64$ (L12–L20), $1 \times 95 \times 128$ (L21–L29), $1 \times 1 \times 128$ (L30–L31) and $1 \times 1 \times 2$ (L32);
548 features selected (*N*) and 326 discarded. Activations: $1 \times 548 \times 1$ (L1), $1 \times 548 \times 32$ (L2–L11), $1 \times 274 \times 64$ (L12–L20), $1 \times 95 \times 137$ (L21–L29), $1 \times 1 \times 128$ (L30–L31) and $1 \times 1 \times 2$ (L32);
319 features selected (*N*) and 555 discarded. Activations: $1 \times 319 \times 1$ (L1), $1 \times 319 \times 32$ (L2–L11), $1 \times 160 \times 64$ (L12–L20), $1 \times 80 \times 128$ (L21–L29), $1 \times 1 \times 128$ (L30–L31) and $1 \times 1 \times 2$ (L32);
229 features selected (*N*) and 645 discarded. Activations: $1 \times 229 \times 1$ (L1), $1 \times 229 \times 32$ (L2–L11), $1 \times 115 \times 64$ (L12–L20), $1 \times 58 \times 128$ (L21–L29), $1 \times 1 \times 128$ (L30–L31) and $1 \times 1 \times 2$ (L32);
449 features selected (*N*) and 425 discarded. Activations: $1 \times 449 \times 1$ (L1), $1 \times 449 \times 32$ (L2–L11), $1 \times 225 \times 64$ (L12–L20), $1 \times 113 \times 128$ (L21–L29), $1 \times 1 \times 128$ (L30–L31) and $1 \times 1 \times 2$ (L32);

The software used for implementing DeepSpectra and SpectraNet was MathWorks Matlab R2021a, with partial use of its Deep Learning Toolbox. The PLS wrapper method was implemented in Python 3.9 and the PLS filtering method in Matlab. Training on an AMD Ryzen 9 5900X CPU and Nvidia RTX 2080 Ti system took about 30 s per network (around 40% GPU usage at 1980 MHz and 1.8 GiB of VRAM). Each saved network occupies around 2.52MiB of disk space in the Matlab Version 7 binary file format. To assess inference time, a network assessment with 2657 train and 643 test spectra was run, resulting in an inference time for predicting 16500 spectra of 2.0594 s. This corresponds to 1.2381E-04 s per spectra, or 8012 spectra per second.

3. Results and discussion

The quality of predictions for citrus fruit in the literature is heavily influenced by several factors, including the quality of the spectrometer, the validation scheme used, the stability of the environmental conditions, and the homogeneity of the samples. In this study, we faced challenges in all of these areas: .

1. The spectrometer we used, the Hamamatsu TG-9405 CA, suffered from a significant etaloning effect in the IR region, which is a known limitation of back-thinned CCD spectrometers that is difficult to compensate for. This puts us at a disadvantage compared to the scientific-grade tabletop systems used in most other studies.
2. We used strict external validation, with a validation dataset that had different spectral characteristics from the calibration dataset (due to differences in origin, storage time, *etc.*). Most results in the literature are obtained using homogeneous datasets where the validation dataset has the same spectral characteristics as the calibration dataset, which leads to better performance figures for the final model.
3. The data was acquired at different temperatures, *i.e.*, thermal equilibration, used in most papers, was deliberately not followed. This makes the predictions even harder, but more representative of the conditions that are typically encountered in fruit packing houses.
4. Our fruit was obtained from a variety of sources (collected over several months) and stored under different conditions (in storage chambers with varying conditions), resulting in a dataset that was less homogeneous than those used in most studies, where the fruit is typically obtained from the same batch and producer.

Despite these challenges, we believe that the results we obtained are acceptable and demonstrate that it is possible to achieve adequate performance even under the most adverse conditions, highlighting the robustness of the chosen methods. Table 3 and Table 4 show key performance metrics for SSC and Temperature results (SN is an abbreviation of SpectraNet, with the synthax SN < #outputs > - < #layers >). Each row corresponds to a specific input data type – *i.e.*, raw reflectance data, absorbance-like converted data, and the three types of PLS feature selection applied on the absorbance-like data: wavelength ordered (λ-ord), importance ordered (*R*-ord) or wavelength-ordered VIP thresholded data – and method. For each row, 30 networks were trained in each of the five EV conditions (Table 1), resulting in 150 trained networks per row, and a total of $23 \times 150 = 3450$ networks. The table columns show the following performance metrics: **rmsec** – root mean squared (rms) error in the training (calibration) set; **rmsep** – rms error of prediction in the test (validation) set; **sdr** – standard deviation ratio $= \frac{\mathrm{std}(y)}{\mathrm{rmsep}}$, where y represents the reference test data; **pg** – prediction gain $= \frac{\mathrm{rmsep}'}{\mathrm{rmsep}}$, where rmsep' is calculated using the average y' of the calibration set as an universal predictor; $\mathbf{R}^2$ – squared correlation coefficient; **cv** – coefficient of variation (%) $= \frac{\mathrm{rmsep}}{\mathrm{mean}(y)} \times 100$; **bias** – the $\mathrm{mean}(\widehat{y}) - \mathrm{mean}(y)$, where $\widehat{y}$ represents the predicted test data; **slope** – the slope of the linear regression y *vs.* $\widehat{y}$. Each value represents a mean of means $\pm$ the standard deviation of the means, *i.e.*, for each set, the results of all 30 networks are averaged and then those averages are used to determine the mean and standard deviation across the 5 EV sets.

Looking at Table 3, all methods were able to meet or surpass a minimum standard of model performance, which was chosen as SDR > 1 (*i.e.*, predictions are better than a random guess around the test population mean), PG > 1 (*i.e.*, predictions are better than a random guess around the training population mean) and $R^2 > 0.16$ (*i.e.*, $|R| > 0.4$ meaning there is some linearity between predicted and true values). Fig. 6 details the SSC root-mean-squared error of prediction (rmsep) for the 5 external validation sets, by method.

Analyzing both Table 3 and Fig. 6: .

- For the raw spectra, the DeepSpectra methods were very similar in rmsep performance (averaging between [1.53, 1.56]), while SN2–32 results were either of comparable performance, or much better (on the second, third and fourth sets of Fig. 6), averaging between [1.34, 1.35];
- For the wavelength-ordered PLS feature selection method, FS(λ-ord), DeepSpectra had comparable performance to the raw spectra results, while SN2–32 networks showed marked improvements, especially the method using Savitzky-Golay second order filtering of the input data. The best overall performance was for the SN2–32 architecture, with Savitzky–Golay 2nd order smoothing; SN1 architectures were also less perform than SN2 architectures.
- For the PLS regression coefficient-ordered feature selection method, FS(*R*-ord), there is a similar improvement on applying it after Savitzky-Golay first or second order filtering. When compared to FS (λ-ord), the performance is very similar for SN2–32 with 1st or 2nd order Savitzky–Golay filtering, or even in the SN2–53 case – which suggests that the SN architecture can be robust to training overfit, by the fact that 21 additional layers only very slightly worsened the generalization capability of the method, when looking at performance scores.
- The PLS VIP-based feature selection methods were all worse than either FS(λ-ord) or FS(*R*-ord), although much better than using either reflectance or absorbance spectra for prediction, with the added benefit of requiring a trivial computation time. FS-VIP($\geq$0.8) and FS-VIP($\geq$1.0) had comparable metrics, while FS-VIP($\geq$1.2) may have filtered important features, resulting in a worse performance on several metrics.
- All SN2 networks, which also predicted fruit temperature along SSC, had better results than networks trained only on an SSC output (SN1). These were SN2–32 and SN2–53 in FS(λ-ord) and SN2–32 in FS(*R*-ord), which were always better than their counterpart SN1

Table 3
Summary of the SSC performance results for all trained networks and input types.

Input Type	Method	rmsec	rmsep	SDR	PG	R^2	CV (%)	Bias	Slope
Reflectance	DeepSpectra	1.44 ± 0.08	1.56 ± 0.33	1.04 ± 0.05	1.17 ± 0.29	0.27 ± 0.10	12.08 ± 2.66	− 0.03 ± 0.80	0.28 ± 0.11
Absorbance	DeepSpectra	1.43 ± 0.08	1.55 ± 0.35	1.04 ± 0.05	1.18 ± 0.31	0.29 ± 0.10	12.03 ± 2.77	− 0.03 ± 0.82	0.28 ± 0.11
	DeepSpectra2	1.44 ± 0.08	1.53 ± 0.35	1.06 ± 0.07	1.20 ± 0.33	0.29 ± 0.11	11.86 ± 2.86	0.05 ± 0.75	0.27 ± 0.12
	SpectraNet2–32	1.22 ± 0.06	1.34 ± 0.24	1.21 ± 0.18	1.34 ± 0.28	0.41 ± 0.12	10.35 ± 1.98	− 0.03 ± 0.38	0.47 ± 0.17
	SGol(1stOrd) + SN2–32	1.18 ± 0.07	1.35 ± 0.34	1.22 ± 0.23	1.37 ± 0.39	0.44 ± 0.13	10.48 ± 2.69	− 0.02 ± 0.56	0.49 ± 0.20
	SGol(2ndOrd) + SN2–32	1.26 ± 0.08	1.34 ± 0.36	1.23 ± 0.17	1.39 ± 0.41	0.39 ± 0.16	10.40 ± 2.97	0.00 ± 0.42	0.40 ± 0.17
FS(λ-ord)	DeepSpectra2	1.42 ± 0.06	1.53 ± 0.34	1.06 ± 0.09	1.19 ± 0.31	0.30 ± 0.12	11.86 ± 2.77	0.02 ± 0.75	0.29 ± 0.14
	SpectraNet2–32	1.01 ± 0.08	1.21 ± 0.17	1.34 ± 0.21	1.47 ± 0.28	0.50 ± 0.11	9.35 ± 1.36	− 0.12 ± 0.15	0.57 ± 0.15
	SGol(1stOrd) + SN2–32	1.04 ± 0.05	1.16 ± 0.26	1.40 ± 0.16	1.57 ± 0.37	0.55 ± 0.06	8.95 ± 1.96	− 0.04 ± 0.36	0.60 ± 0.15
	SGol(2ndOrd) + SN1–32	0.98 ± 0.05	1.12 ± 0.22	1.44 ± 0.02	1.62 ± 0.37	0.57 ± 0.05	8.63 ± 1.72	− 0.02 ± 0.32	0.60 ± 0.14
	SGol(2ndOrd) + SN1–53	1.04 ± 0.02	1.18 ± 0.27	1.37 ± 0.07	1.54 ± 0.35	0.53 ± 0.04	9.09 ± 1.99	− 0.07 ± 0.43	0.54 ± 0.08
	SGol(2ndOrd) + SN2–32	1.01 ± 0.04	**1.08 ± 0.22**	**1.49 ± 0.06**	**1.67 ± 0.37**	**0.58 ± 0.05**	**8.35 ± 1.67**	− 0.06 ± 0.14	0.57 ± 0.11
	SGol(2ndOrd) + SN2–53	1.00 ± 0.03	1.11 ± 0.20	1.45 ± 0.13	1.61 ± 0.31	0.57 ± 0.06	8.59 ± 1.47	0.01 ± 0.31	**0.61 ± 0.10**
FS(*R*-ord)	DeepSpectra	1.42 ± 0.07	1.53 ± 0.35	1.06 ± 0.08	1.20 ± 0.32	0.30 ± 0.12	11.83 ± 2.82	0.01 ± 0.75	0.28 ± 0.13
	DeepSpectra2	1.42 ± 0.07	1.52 ± 0.34	1.06 ± 0.08	1.20 ± 0.32	0.30 ± 0.12	11.80 ± 2.80	0.02 ± 0.74	0.28 ± 0.13
	SpectraNet2–32	1.02 ± 0.05	1.28 ± 0.24	1.27 ± 0.20	1.41 ± 0.31	0.45 ± 0.14	9.90 ± 1.95	− 0.02 ± 0.29	0.53 ± 0.19
	SGol(1stOrd) + SN2–32	0.96 ± 0.05	1.10 ± 0.22	1.46 ± 0.16	**1.64 ± 0.36**	0.58 ± 0.07	8.54 ± 1.73	0.01 ± 0.26	**0.62 ± 0.15**
	SGol(2ndOrd) + SN1–32	0.98 ± 0.05	1.11 ± 0.22	1.44 ± 0.02	1.62 ± 0.37	0.57 ± 0.04	8.63 ± 1.72	− 0.03 ± 0.33	0.60 ± 0.13
	SGol(2ndOrd) + SN2–32	0.96 ± 0.03	**1.08 ± 0.20**	**1.48 ± 0.13**	**1.66 ± 0.36**	**0.58 ± 0.06**	**8.40 ± 1.61**	0.04 ± 0.20	0.61 ± 0.13
	SGol(2ndOrd) + SN2–53	0.94 ± 0.03	**1.09 ± 0.18**	**1.47 ± 0.11**	1.63 ± 0.31	**0.58 ± 0.05**	**8.46 ± 1.40**	0.02 ± 0.32	**0.62 ± 0.09**
FS-VIP(≥0.8)	SGol(2ndOrd) + SN2–32	1.08 ± 0.05	1.17 ± 0.20	1.38 ± 0.10	1.54 ± 0.32	0.53 ± 0.08	9.03 ± 1.67	− 0.01 ± 0.30	0.54 ± 0.12
FS-VIP(≥1.0)		1.09 ± 0.05	1.17 ± 0.24	1.38 ± 0.10	1.56 ± 0.37	0.52 ± 0.07	9.03 ± 1.93	0.00 ± 0.26	0.55 ± 0.13
FS-VIP(≥1.2)		1.10 ± 0.04	1.21 ± 0.24	1.34 ± 0.15	1.50 ± 0.33	0.50 ± 0.10	9.34 ± 1.92	0.02 ± 0.35	0.54 ± 0.13

Notes. SSC (%) and SEL = 0.1, with SEL being the standard error of laboratory. All values represent the mean ± standard deviation of 150 networks (30 per EV set). All methods attained the minimum standard of performance, consisting of SDR ≥ 1, PG ≥ 1 and $R^2 > 0.16$ (meaning that $|R| > 0.4$). **Bold**: Top-3 performance methods for each column.

Table 4
Summary of the Temperature performance results for all trained networks and input types.

Input Type	Method	rmsec	rmsep	SDR	PG	R^2	CV (%)	Bias	Slope
Absorbance	DeepSpectra2	4.78 ± 0.16	5.53 ± 0.75	1.41 ± 0.31	1.53 ± 0.27	0.62 ± 0.08	30.83 ± 9.13	− 0.25 ± 3.01	0.52 ± 0.06
	SpectraNet2–32	1.85 ± 0.11	2.09 ± 0.19	3.72 ± 0.92	4.00 ± 0.33	**0.94 ± 0.02**	11.67 ± 3.45	0.12 ± 0.82	0.89 ± 0.06
	SN2–32 + SGol(1stOrd)	1.97 ± 0.12	2.14 ± 0.20	3.67 ± 1.03	3.93 ± 0.31	0.93 ± 0.03	11.86 ± 3.13	0.04 ± 0.82	0.91 ± 0.05
	SN2–32 + SGol(2ndOrd)	2.46 ± 0.12	2.83 ± 0.34	2.75 ± 0.72	2.96 ± 0.34	0.89 ± 0.05	15.64 ± 3.72	− 0.04 ± 1.36	0.84 ± 0.05
FS(λ-ord)	DeepSpectra2	4.32 ± 0.10	5.10 ± 0.84	1.55 ± 0.40	1.68 ± 0.36	0.67 ± 0.10	28.49 ± 9.28	− 0.16 ± 2.76	0.59 ± 0.07
	SpectraNet2–32	1.71 ± 0.08	**2.01 ± 0.15**	**3.85 ± 0.98**	**4.14 ± 0.24**	**0.94 ± 0.02**	**11.17 ± 2.78**	0.10 ± 0.86	0.92 ± 0.05
	SN2–32 + SGol(1stOrd)	1.67 ± 0.08	**1.83 ± 0.18**	**4.21 ± 1.01**	**4.55 ± 0.33**	**0.95 ± 0.02**	**10.09 ± 2.00**	− 0.23 ± 0.60	**0.92 ± 0.03**
	SN2–32 + SGol(2ndOrd)	1.77 ± 0.11	2.08 ± 0.30	3.80 ± 1.17	4.05 ± 0.50	**0.94 ± 0.02**	11.50 ± 3.06	0.00 ± 1.02	0.88 ± 0.07
FS(*R*-ord)	DeepSpectra	4.27 ± 0.19	5.01 ± 0.76	1.56 ± 0.39	1.70 ± 0.35	0.69 ± 0.07	27.87 ± 8.22	− 0.25 ± 2.72	0.60 ± 0.07
	DeepSpectra2	4.30 ± 0.25	5.07 ± 0.71	1.54 ± 0.37	1.68 ± 0.33	0.68 ± 0.07	28.19 ± 7.98	− 0.28 ± 2.80	0.60 ± 0.07
	SpectraNet2–32	1.70 ± 0.09	2.17 ± 0.43	3.65 ± 0.96	3.94 ± 0.66	**0.94 ± 0.02**	11.68 ± 1.25	− 0.33 ± 1.05	0.89 ± 0.07
	SN2–32 + SGol(1stOrd)	1.60 ± 0.12	**1.83 ± 0.27**	**4.25 ± 0.92**	**4.62 ± 0.63**	**0.95 ± 0.01**	**9.93 ± 1.34**	− 0.50 ± 0.66	**0.93 ± 0.03**
	SN2–32 + SGol(2ndOrd)	1.85 ± 0.15	2.17 ± 0.21	3.54 ± 0.84	3.82 ± 0.29	0.93 ± 0.03	11.97 ± 2.27	− 0.24 ± 0.94	0.90 ± 0.06
	SN2–53 + SGol(2ndOrd)	1.84 ± 0.13	2.15 ± 0.33	3.69 ± 1.19	3.92 ± 0.51	0.93 ± 0.03	12.03 ± 4.03	− 0.04 ± 1.04	**0.92 ± 0.04**
FS-VIP(≥0.8)	SN2–32 + SGol(2ndOrd)	2.30 ± 0.11	2.57 ± 0.18	2.96 ± 0.53	3.22 ± 0.28	0.90 ± 0.03	14.18 ± 2.88	− 0.10 ± 1.10	0.87 ± 0.05
FS-VIP(≥1.0)		3.03 ± 0.27	3.51 ± 0.52	2.19 ± 0.45	2.38 ± 0.36	0.83 ± 0.04	19.17 ± 3.19	0.02 ± 1.73	0.80 ± 0.06
FS-VIP(≥1.2)		3.36 ± 0.15	4.02 ± 0.68	1.95 ± 0.57	2.09 ± 0.34	0.77 ± 0.10	22.30 ± 6.38	0.01 ± 2.14	0.76 ± 0.09

Notes. Not all methods are shown. All values are the mean ± standard deviation of 150 networks (30 per EV set). All methods attained the minimum standard of performance, consisting of SDR ≥ 1, PG ≥ 1 and $R^2 > 0.16$ (meaning that $|R| > 0.4$). **Bold**: Top-3 performance methods for each column.

methods. This was expected, as CNN-based deep neural networks can create shared representations of information using internal correlations between targets as cues during prediction (Padarian et al., 2019; Ramsundar et al., 2015), as well as increasing prediction accuracy by reducing overfitting (Ruder, 2017).

Fig. 7 depicts the scatter plots of the best SpectraNet2–32 methods for predicting SSC and fruit temperature, which were *SN2–32 on SGol (2ndOrd) with FS(λ-ord)* and *SN2–32 on SGol(1stOrd) with FS(R-ord)*, respectively. For illustration of the results of each of the five sets, 5 representative networks per graph were selected, which were the ones (from the pool of 30 per set) that had the closest rmsep value to the median of each set. Aggregated statistics across all networks are also shown. Additional performance metrics are shown for the following parameters: **res** – resolution, defined as $\frac{\text{rmsep}}{\max(y)-\min(y)} \times 100$, which is a measure of how the prediction error resolves the range of y–variation; **N** – the total number of data points in each figure; %**OK** – the percentage of test samples which are correctly assigned below or above the population mean (shown as **@ y**); this value is further detailed in **%TN** and **%TP**, representing the percentages of correct assignments below (True Negative) or above (True Positive) the population mean.

Table 5 compares the SSC results obtained for the SN2–32 on SGol (2ndOrd) with FS(λ-ord) data to Passos et al. (2019). The authors reported that their full spectra results were very similar between evaluated methods, but removing the chlorophyll bands below 730 nm improved most results, in varying degrees, except for the multilayer perceptron (MLP) case. Their overall best results were for the support-vector machine (SVM) 'No Chls' case. Comparing it to SpectraNet2–32, the full spectra results are very similar (if only slightly better) to the SVM 'No Chls' case, which means that the network was able to achieve a good generalization performance, without being hindered by the chlorophyll bands. Overall, these are good results for SN–32 and highlight the robustness of the network in achieving a good generalization performance without depending on removing the most troublesome bands of

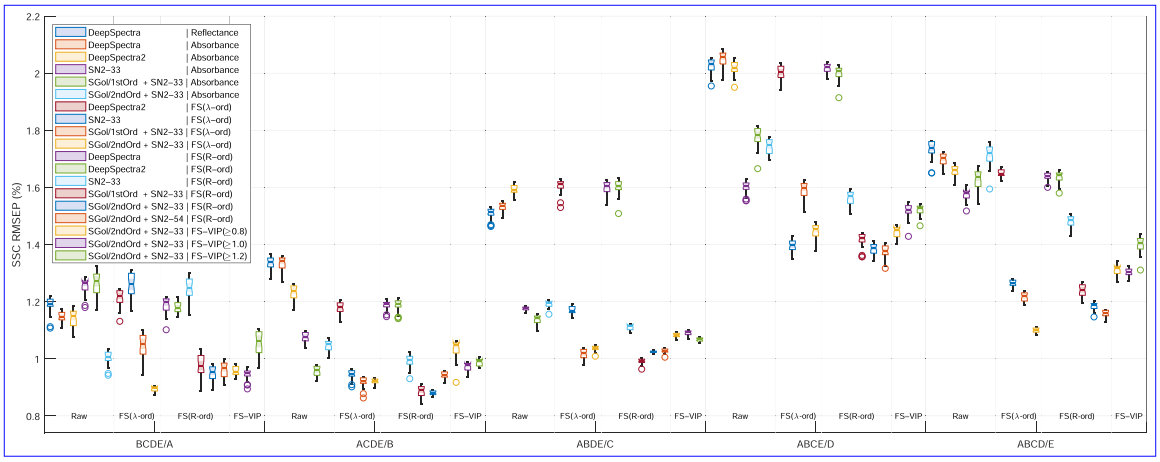


Fig. 6. Box plot of the Table 3 SSC results for the root-mean-squared error of prediction (rmsep). Each box represents 30 neural networks. These were trained using 23 different methods (of which 19 are shown here), across 5 training/validation population splits (from BCDE/A to ABCD/E). The shaded and notched regions of each bar represent the 95% confidence interval.

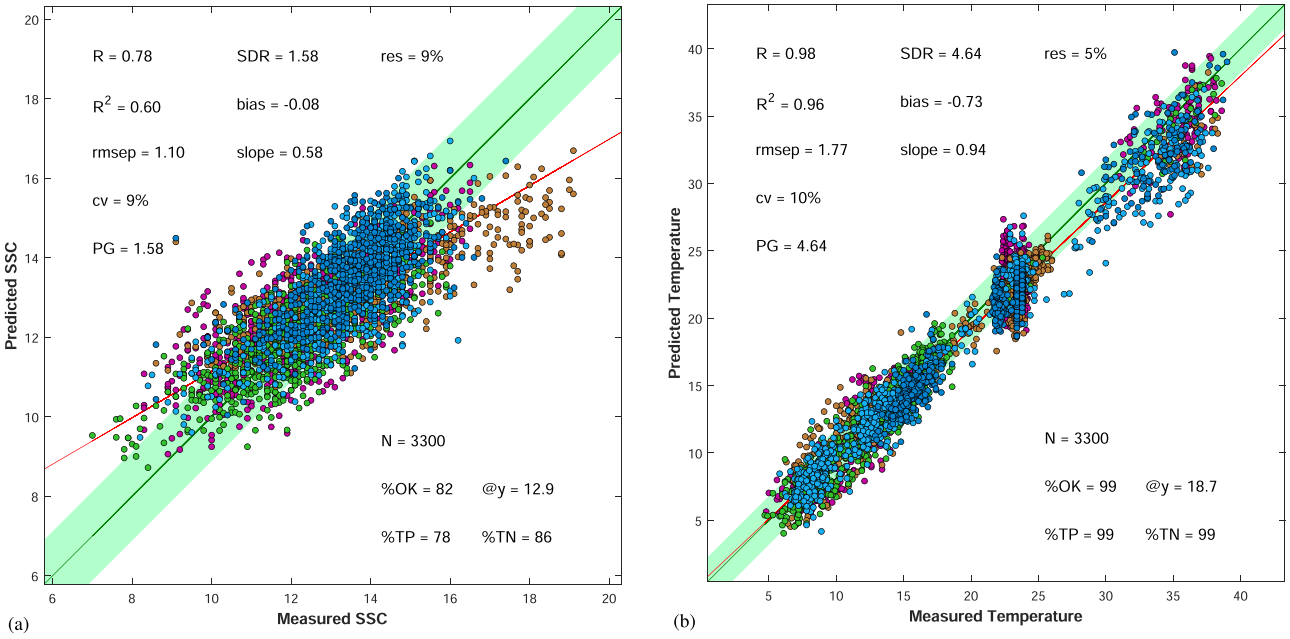


Fig. 7. Regression scatter plots of the best overall method for (a) SSC and (b) Temperature, using *representative networks*. In each graphic, these were the 5 networks with an rmsep value closest to the median of the 30 trained on each set, across the 5 EV sets. Statistics are computed across the 5 networks (one per set). Each marker colour corresponds to a different EV set.

Table 5
Comparison of SSC results with Passos et al. (2019).

	PLS		*MLR*		*SVM*		*MLP (NN)*		SN2–32
	Full	No Chls	Full	No Chls	Full	No Chls	Full	No Chls	SG (2nd)+FS (λ)
rmsec	0.96	0.97	0.89	0.91	0.93	0.92	0.6	0.49	1.01
rmsep	1.15	1.11	1.15	1.11	1.16	1.09	1.15	1.21	**1.08**
PG	1.55	1.6	1.54	1.59	1.56	1.63	1.55	1.46	**1.67**
R^2	0.56	0.58	*0.55*	0.57	0.57	**0.6**	0.57	0.51	0.58
$CV_{\%}$	8.86	8.57	8.9	8.63	8.94	8.45	8.89	9.64	**8.35**

Notes. Italic methods are results from Passos et al. (2019). The best prediction values are displayed in **bold**.

the input data.

4. Conclusion and final remarks

This study introduced a novel approach for predicting the soluble solids content (SSC) and temperature of 'Rocha' pears using the

SpectraNet deep learning architecture and 1D visible to near-infrared (Vis-NIRS) spectra. Feature selection was also applied to the input spectra, utilizing two types of partial least squares (PLS) methods, in order to identify the most significant wavelengths for training the model. For assessing the performance of SpectraNet–32, inference results were compared to a similar state-of-the-art deep learning architecture, DeepSpectra, as well as four classical machine learning algorithms: PLS, multiple linear regression (MLR), support vector machine (SVM), and multi-layer perceptron (MLP).

Our results have demonstrated that the modified SpectraNet–32 architecture, with a reduced number of layers and the incorporation of PLS-based input feature selection, is able to achieve superior performance for predicting both SSC and temperature of 'Rocha' pears. On average, it outperformed PLS, MLR and MLP in all metrics, including a 6.1% improvement in terms of the root mean square error of prediction (RMSEP) and a 7.7% improvement in prediction gain (PG) compared to PLS. It also outperformed SVM in three of the four measured metrics. Additionally, the SpectraNet–32 model exhibited a consistent low variability in performance for repeated training, highlighting its potential as a robust and reliable tool for predicting fruit quality attributes.

Additionally, simultaneously predicting temperature and SSC always resulted in better SSC estimates than using networks only trained to predict SSC. This highlights the importance of considering the influence of temperature on the spectral properties of fruit and the potential for improving prediction performance by accounting for it in machine-learning models.

The results presented in this article demonstrate the potential of the SpectraNet–32 model as a powerful tool for predicting fruit quality attributes, and highlight the importance of considering the relevance of individual wavelengths in the prediction process. The use of spectra-based methods in Precision Agriculture offers non-destructive, quick tools for measuring the quality of produce, making them a valuable asset in the field. The fast inference time of SpectraNet–32, at around 8000 spectra per second, makes it a practical tool for real-time quality assessment in precision agriculture. These findings provide valuable insights for future research, as the successful application of the SpectraNet–32 model to 'Rocha' pears paves the way for further exploration of its potential in predicting other fruit quality attributes, as well as in different commodities.

CRediT authorship contribution statement

J. A. Martins: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **D. Rodrigues**: Investigation, Data curation. **A. M. Cavaco**: Conceptualization, Methodology, Investigation, Resources, Data curation, Writing - review & editing, Supervision, Project administration, Funding acquisition. **M. D. Antunes**: Conceptualization, Resources, Project administration, Funding acquisition. **R. Guerra**: Conceptualization, Methodology, Investigation, Writing - review & editing, Visualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was developed within the Centre for Electronics, Optoelectronics and Telecommunications (CEOT), and supported by the Portuguese Science and Technology Foundation (FCT), within the national funds of DL57/2016/CP1361/CT0040 to J. A. Martins; *Ciência 2008* and post-doc fellowship SFRH/BPD/101634/2014 to A. M. Cavaco; and projects UIDB/00631/2020 CEOT BASE and UIDP/00631/2020 CEOT PROGRAMÁTICO. *Caixa Geral de Depósitos* supported Daniela Rodrigues (*Ideias em Caixa*, 2010). Part of the work was also supported by project OtiCalFrut (ALG-01–0247-FEDER-033652).

References

Abdi, H., 2010. Partial least squares regression and projection on latent structure regression (pls regression). Wiley Interdiscip. Rev.: Comput. Stat. 2, 97–106. https://doi.org/10.1002/wics.51.

ANP, Pêra rocha – a pera rocha distingue-se pelas suas características únicas, que fazem dela um sabor de portugal, 2022.⟨https://perarocha.pt/?lang=en⟩.

Benelli, A., Cevoli, C., Fabbri, A., 2020. In-field vis/nir Hyperspectral Imaging To Measure Soluble Solids Content Of Wine Grape Berries During Ripening. IEEE, pp. 99–103. ⟨https://ieeexplore.ieee.org/document/9277621/⟩, 10.1109/MetroAgriFor50201.2020.9277621.

Cavaco, A.M., Utkin, A.B., daSilva, J.M., Guerra, R., 2022. Making sense of light: The use of optical spectroscopy techniques in plant sciences and agriculture. Appl. Sci. (Switz.) 12, 997. https://doi.org/10.3390/app12030997. ⟨https://www.mdpi.com/2076-3417/12/3/997/htm⟩.

Chong, I.G., Jun, C.H., 2005. Performance of some variable selection methods when multicollinearity is present. Chemom. Intell. Lab. Syst. 78, 103–112. https://doi.org/10.1016/j.chemolab.2004.12.011.

Cruz, S., Guerra, R., Brazio, A., Cavaco, A.M., Antunes, D., Passos, D., 2021. Nondestructive simultaneous prediction of internal browning disorder and quality attributes in 'rocha' pear (Pyrus communis l.) using vis-nir spectroscopy. Postharvest Biol. Technol. 179, 111562 https://doi.org/10.1016/j.postharvbio.2021.111562.

X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, volume 9, 2010, pp.249–256.⟨http://proceedings.mlr.press/v9/glorot10a.html⟩.

Gosselin, R., Rodrigue, D., Duchesne, C., 2010. A bootstrap-vip approach for selecting wavelength intervals in spectral imaging applications. Chemom. Intell. Lab. Syst. 100, 12–21. https://doi.org/10.1016/j.chemolab.2009.09.005.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. http://arxiv.org/abs/1502.01852 IEEE Int. Conf. Comput. Vis. (ICCV) 9, 1026–1034. https://doi.org/10.1109/ICCV.2015.123. ⟨http://ieeexplore.ieee.org/document/7410480/⟩. http://arxiv.org/abs/1502.01852.

K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, 2016.10.1007/978–3-319–46493-0_38.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning For Image Recognition, Volume 2016-Decem. IEEE, pp. 770–778. ⟨http://image-net.org/challenges/LSVRC/2015/⟩. http://ieeexplore.ieee.org/document/7780459/, http://arxiv.org/abs/1512.03385, 10.1109/CVPR.2016.90.

D. Hendrycks, K. Gimpel, Gaussian error linear units (gelus)(2016). ⟨http://arxiv.org/abs/1606.08415⟩.

Kaur, H., Künnemeyer, R., McGlone, A., 2022. Correction of temperature variation with independent water samples to predict soluble solids content of kiwifruit juice using nir spectroscopy. Molecules 27, 504. https://doi.org/10.3390/molecules27020504. ⟨https://www.mdpi.com/1420-3049/27/2/504/htm⟩.

Li, B., Lecourt, J., Bishop, G., 2018. Advances in non-destructive early assessment of fruit ripeness towards defining optimal time of harvest and yield prediction—a review. Plants 7, 1–20. https://doi.org/10.3390/plants7010003.

Lu, R., Peng, Y., 2006. Hyperspectral scattering for assessing peach fruit firmness. Biosyst. Eng. 93, 161–171. https://doi.org/10.1016/j.biosystemseng.2005.11.004.

Martins, J., Guerra, R., Pires, R., Antunes, M., Panagopoulos, T., Brázio, A., Afonso, A., Silva, L., Lucas, M., Cavaco, A., 2022. Spectranet–53: a deep residual learning architecture for predicting soluble solids content with vis–nir spectroscopy. Comput. Electron. Agric. 197, 106945 https://doi.org/10.1016/j.compag.2022.106945. ⟨https://linkinghub.elsevier.com/retrieve/pii/S0168169922002629⟩.

T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in partial least squares regression, 2012.10.1016/j.chemolab.2012.07.010.

Mishra, P., Passos, D., 2021. Deep multiblock predictive modelling using parallel input convolutional neural networks. Anal. Chim. Acta 1163, 338520. https://doi.org/10.1016/j.aca.2021.338520. ⟨https://linkinghub.elsevier.com/retrieve/pii/S0003267021003469⟩.

Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning for digital soil mapping. Soil 5, 79–89. https://doi.org/10.5194/soil-5-79-2019. ⟨https://www.soil-journal.net/5/79/2019/⟩.

Passos, D., Rodrigues, D., Cavaco, A.M., Antunes, M.D., Guerra, R., 2019. Non-destructive soluble solids content determination for 'rocha' pear based on vis-swnir spectroscopy under 'real world' sorting facility conditions. Sens. (Switz.) 19, 5165. https://doi.org/10.3390/s19235165. ⟨https://www.mdpi.com/1424-8220/19/23/5165⟩.

B. Ramsundar, S. Kearnes, P. Riley, D. Webster, D. Konerding, V. Pande, Massively multitask networks for drug discovery(2015). ⟨http://arxiv.org/abs/1502.02072⟩.

S. Ruder, An overview of multi-task learning in deep neural networks, arXiv(2017). ⟨http://arxiv.org/abs/1706.05098⟩.

M. Simon, E. Rodner, J. Denzler, Imagenet pre-trained models with batch normalization (2016). ⟨http://www.inf-cv.uni-jena.de/Research/CNN.Models.htmlandhttps://github.com/cvjena/cnn-models⟩, ⟨http://arxiv.org/abs/1612.01452⟩.

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, volume 07–12-June, 2015, pp.1–9.10.1109/CVPR.2015.7298594.

Walsh, K.B., Blasco, J., Zude-Sasse, M., Sun, X., 2020. Visible-nir 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. Postharvest Biol. Technol. 168, 111246 https://doi.org/10.1016/j.postharvbio.2020.111246. ⟨https://linkinghub.elsevier.com/retrieve/pii/S0925521419303230⟩.

Yu, X., Lu, H., Wu, D., 2018. Development of deep learning method for predicting firmness and soluble solid content of postharvest korla fragrant pear using vis/nir hyperspectral reflectance imaging. Postharvest Biol. Technol. 141, 39–49. https://doi.org/10.1016/j.postharvbio.2018.02.013.

Zhang, X., Lin, T., Xu, J., Luo, X., Ying, Y., 2019. Deepspectra: an end-to-end deep learning approach for quantitative spectral analysis. Anal. Chim. Acta 1058, 48–57. https://doi.org/10.1016/j.aca.2019.01.002.