Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 59-23 (2025) 102–107

3D Convolutional Networks Outperform Traditional Methods for Strawberry Analysis with Spectral Imaging

Salvador Castillo-Girones*, Jos Ruizendaal, Xiomara Salas-Valderrama**, Sandra Munera***, Jose Blasco*, Gerrit Polder****

*Agroengineering center, Instituto Valenciano de Investigaciones Agrarias. Carretera CV-315, Km 10.7, 46113, Moncada, Valencia, Spain (castillo_salgirb@gva.es; blasco_josiva@gva.es)

**Greenhouse Horticulture Business Unit, Wageningen University & Research. Droevendaalsesteeg 4, 6708 PB, Wageningen, Netherlands (jos.ruizendaal@wur.nl; xiomasav@gmail.com; gerrit.polder@wur.nl)

*** Department of graphic engineering, Universitat Politècnica de València. Camino de Vera, s/n. 46022, Valencia, Spain (sanmupi@upv.es)

Abstract: Spectral imaging combined with machine learning offers a powerful approach to predicting quality parameters and classifying strawberry cultivars. This study compares the performance of Partial Least Squares (PLS) models and 3D Convolutional Neural Networks (3D-CNNs) for Total Soluble Solids (TSS, °Brix) prediction and cultivar discrimination using a dataset of 17 strawberry cultivars from two origins. For TSS prediction, the 3D-CNN model achieved superior accuracy with an R^2 of 0.82 compared to 0.71 for the PLS model. For cultivar classification, the 3D-CNN model outperformed traditional approaches with an F1 score of 0.87, compared to 0.75 for the PLS model. The CNN's ability to utilise both spatial and spectral features allowed it to capture subtle morphological differences among cultivars, which traditional models struggled to identify effectively. These findings demonstrate the superiority of deep learning models over traditional spectral methods in handling complex datasets and highlight the potential of spectral imaging and CNNs for robust quality assessment and classification in agricultural applications. Although previous studies have classified strawberry cultivars and predicted TSS, none have included as many varieties or samples (3,564) as this study. Additionally, the proposed model's simple and replicable structure makes it especially useful for cultivar identification and easy for other researchers to adopt.

Keywords: strawberry; quality; consumer acceptance; spectral imaging; machine learning.

1. INTRODUCTION

Neural networks are increasingly being recognised for their ability to improve predictive modelling in various fields, often outperforming traditional machine learning models. Their capacity to capture non-linear and complex patterns within data has made them particularly valuable in solving challenging problems. Convolutional Neural Networks (CNNs), in particular 3D-CNNs, are gaining attention for their effectiveness in processing image-based data. Traditionally, data analysis has relied on machine learning algorithms trained on large labelled datasets to perform classification or regression tasks. While methods like Partial Least Squares (PLS) regression have been effective in certain contexts, they often struggle to capture the intricate relationships inherent in complex datasets. In contrast, CNNs excel at identifying hidden patterns and non-linear relationships, making them a promising alternative for hyperspectral imaging applications. Specifically, 3D-CNNs are designed to process n-dimensional data, enabling them to leverage the full potential of hyperspectral images (Chollet, 2021).

Spectral imaging and spectroscopy are widely used for monitoring crops and fruits. While spectroscopy gathers information from specific parts of a sample, spectral imaging captures both spatial and spectral data from the entire sample, providing a more comprehensive dataset. This makes spectral imaging especially suitable for agricultural applications, where understanding both the internal and external properties of products is crucial (Polder et al., 2024). Previous studies have demonstrated the potential of spectral imaging combined with machine learning techniques for evaluating the internal properties of fruits. For example, Pullanagari & Li. (2021) used PLS regression to predict Total Soluble Solids (TSS) and firmness in gooseberries, achieving R2 values of 0.88 and 0.60, respectively. While these results are promising, there is a growing need to explore advanced neural network models like CNNs to further enhance prediction accuracy.

This study aims to evaluate the use of spectral imaging for predicting the TSS content of seventeen strawberry cultivars while also discriminating between these cultivars. A key objective was to develop and compare predictive models using a traditional machine learning algorithm (PLS) and CNNs by employing both complete image data and mean spectra. By highlighting the advantages of 3D-CNNs in processing hyperspectral data compared to traditional methods such as PLS regression and mean spectra analysis, this study aims to enhance the use of spectral imaging for evaluating fruit quality.

While previous studies have predicted total soluble solids (TSS) in strawberries and performed cultivar classification, none have examined as many different varieties (17) or included as large a sample size (3,564 samples) as this study. Moreover, the proposed model stands out for its simple and replicable structure, making it particularly accessible for other researchers. Its straightforward design and demonstrated

10.1016/j.ifacol.2025.11.771

effectiveness make it especially suitable for cultivar identification, encouraging broader adoption in future studies.

2. MATERIALS AND METHODS

2.1 Fruit samples and experimental design

The experiments were conducted in a laboratory located in the Greenhouse Horticulture & Flower Bulbs department at Wageningen University & Research in Bleiswijk, The Netherlands. A total of 3,564 strawberries from 17 different cultivars were analysed, with eight cultivars grown in the Netherlands and nine in Spain. For each cultivar, nine boxes of strawberries were provided, with each box containing 27 strawberries arranged in trays of five rows and either five or six columns.

The strawberries cultivated in the Netherlands were harvested three days before the experiment and transported to the laboratory in a refrigerated van at 5°C. Similarly, the strawberries from Spain were collected three days before the experiment and shipped refrigerated at 5°C to the laboratory. In both cases, the fruit was stored in the laboratory at 5°C and 90% relative humidity (RH) to prevent degradation before the experiments began.

Once all samples were collected, the boxes containing the strawberries were imaged. From these images, the mean spectra for each box, as well as the individual hyperspectral images of each strawberry, were extracted. Following the image acquisition, the strawberries were separated, blended, and their TSS content was measured using a refractometer, as shown in Figure 1.

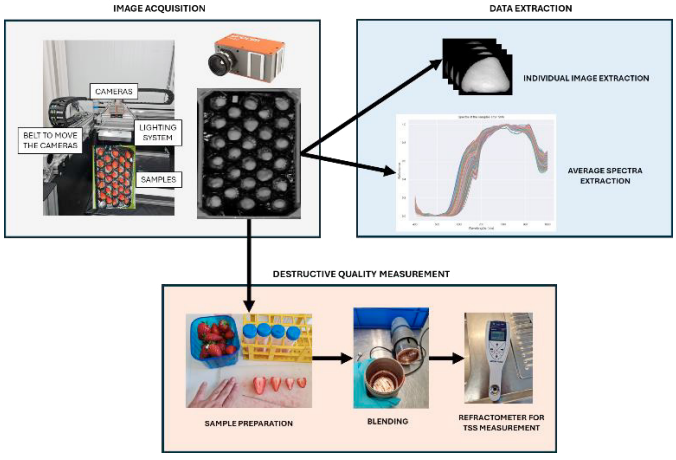

Figure 1. Acquisition of images and collection of spectral data and total suspended solids (TSS) content flow.

2.2 Spectral image acquisition

Prior to image acquisition, the strawberries were allowed to reach room temperature (20 °C) for one hour. A total of 3564 strawberry samples were imaged in a laboratory setting under controlled conditions at 20 °C using a line-scanning camera (Specim FX10, Specim, Oulu, Finland) with sensitivity in the spectral range of 400 to 1000 nm. The images were obtained with spectral intervals of 2.7 nm, a spatial resolution of 0.28 mm per scan, and an exposure time of 10 ms. The scene was illuminated using a halogen lamp (125 V, 3.120 A, 390 W, Phillips, Amsterdam, The Netherlands) for the NIR range and LED lights (Effi-Flex-HSI-300-910-970-TR-$, Effilux, Les Ules, France) for the Vis/NIR range. After imaging, the samples were returned to storage at 5 °C and 90% relative humidity. On the following day, quality assessments and consumer sensory evaluations were conducted.

Once images were acquired, the reflectance of the images was corrected using a white polytetrafluorethylene (PTFE) reference (99 % reflectance) and dark references (image with the camera shutter closed). Equation X shows the image correction.

Image reflectance correction is described in equation (1), where $\rho^{Reference}(\lambda)$ is the reflectance value of the standard surface at a specific wavelength, $Im_{Dark(\lambda)}$ and $Im_{White(\lambda)}$ are the dark and white refeences at that specific wavelength, and $Im_{Raw(\lambda)}$ is the raw image at that wavelength.

$$Im_{Refl} = \rho^{Reference}(\lambda)\frac{Im_{Raw(\lambda)} - Im_{Dark(\lambda)}}{Im_{White(\lambda)} - Im_{Dark(\lambda)}} \quad (1)$$

2.3 Destructive quality measurements

For the quality assessment, 174 samples out of the total 3564 strawberries were selected for destructive quality measurements, TSS in this case, which involved destroying the sample to analyse internal properties. These samples were randomly chosen to ensure equal representation across all cultivars and boxes, providing a fair and comprehensive evaluation of the fruit's quality.

The selected fruit was blended into a uniform mixture to measure TSS, an indicator of sugar content, expressed in degrees Brix (°Brix). The measurement was conducted using a refractometer (Mettler Toledo, 4004 JK Tiel, Netherlands), an instrument that determines the refractive index of the juice. This method is based on the principle of total refraction, where the refractometer calculates how light bends as it passes through the sample, correlating this to sugar concentration.

2.4 Image processing and data extraction

The reflectance of the images was corrected using a white polytetrafluorethylene (PTFE) reference and dark references. To isolate individual strawberries, a deep learning model was employed. Specifically, a Detectron 2 fast-RCNN network, trained on FX10 false-colour images derived from the spectral data, was used to segment single strawberries from images of entire boxes.

Once segmented, the spectral images of each strawberry were extracted and saved, forming a spectral image dataset. Additionally, the mean spectra of each strawberry were calculated and stored to create a separate spectral dataset later used to train a 3D Convolutional Neural Network model (3D-CNN). This dataset was used to analyse spectral differences among samples and to develop a machine Learning (ML) model, Partial Least Squares Discriminant Analysis (PLS-DA).

2.5 Data analysis and model creation

Using the spectra dataset, Partial Least Squares Regression (PLSR) was employed to predict TSS, while Partial Least Squares Discriminant Analysis (PLS-DA) was used to classify the strawberry varieties. Partial Least Squares (PLS) reduce

the dimensionality of high-dimensional data by identifying latent variables (components) that capture the most relevant variance for predicting a target variable. PLSR is specifically used for continuous variables, such as TSS, while PLS-DA extends this approach for categorical classification tasks, such as identifying strawberry varieties. These methods are particularly effective for analysing spectral data, where high dimensionality and multicollinearity are common challenges (Müller & Guido, 2016).

Using the image dataset, a 3D-CNN model was developed for each parameter under study. To address computational limitations, the images were resized to 100 x 100 pixels on the spatial dimension but not in the spectral dimension, maintaining all 224 wavelengths before training. Additionally, data augmentation techniques, such as image rotation and flipping, were applied to prevent overfitting during the training process.

First, a 3D-CNN model was developed to classify 17 different strawberry cultivars using the spectral dataset of 3,564 samples. After training the model for cultivar classification, transfer learning was applied to adapt the model for TSS prediction. Specifically, the final dense layers after flattening (global average pooling) were replaced to suit the regression task, and then these layers were fine-tuned using the TSS dataset (174 samples), while the earlier layers were kept frozen. This approach leveraged the feature representations learned from the larger classification dataset, reducing the risk of overfitting when training on the smaller TSS dataset.

The model structures are shown in Figure 2. Both models consisted of three convolutional layers with 16, 32, and 64 neurons, respectively, and kernel sizes of 3x3x3. Each convolutional layer was followed by a max-pooling layer with a pooling size of 2 and batch normalisation to stabilise training. After these layers, a global average max pooling 3D layer was included, followed by two dense layers: the first with 128 neurons and the second varying depending on the task. For the regression model predicting TSS, the final dense layer had 1 neuron, while the cultivar classification model had 17 neurons.

A dropout rate of 0.2 was applied between the first and last dense layers to reduce overfitting. All layers used ReLU activation functions except for the final dense layers. In the cultivar classification model, a softmax activation function was applied in the last layer to output probabilities for each class. A softplus activation function was used in the final layer for the TSS regression model to ensure non-negative predictions.

Both models were compiled using the Adam optimiser. The loss function varied depending on the task: Categorical Crossentropy was used for cultivar classification, while Mean Squared Error (MSE) was employed for TSS prediction. This architecture ensured flexibility and performance tailored to both regression and classification tasks.

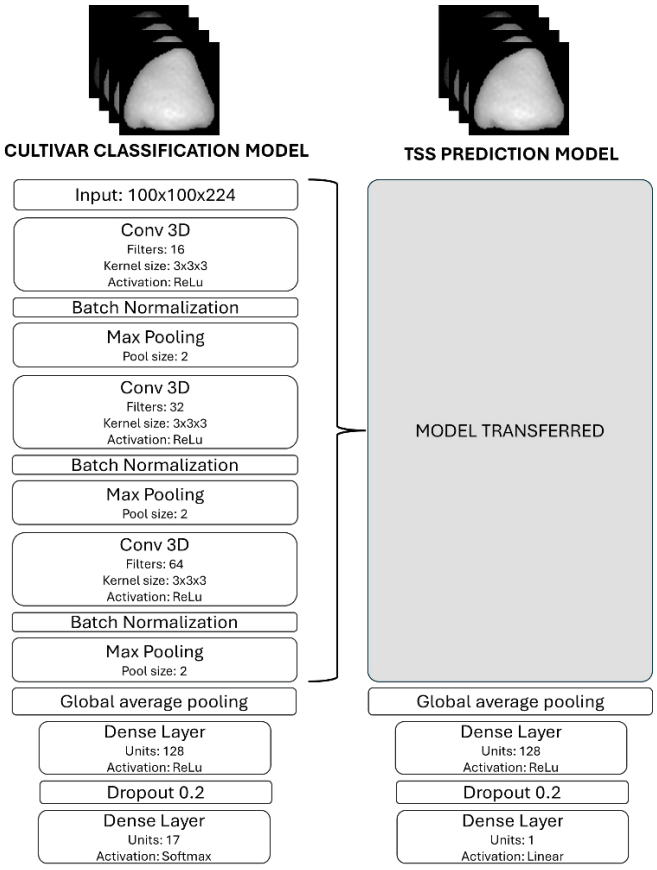

Figure 2. The 3D-CNN Model structures for cultivar (left) and TSS prediction (right).

70 % of the data was used for training, 15 % for validation during training, and 15 % for testing the created models. This was accomplished using scikit-learn's train-test split to separate between the training, validation, and test sets.

For classification models, the F1 score was used, and the confusion matrices were obtained on the test set prediction. The F1 score assesses accuracy by combining precision and recall. Precision reflects the reliability of the model's positive predictions, whereas recall (or sensitivity) measures the model's ability to identify all positive instances correctly.

Equation (2) shows how the F1 score was calculated, where TP means true positives, FP means false positives, and FN means false negatives.

$$F1 \quad \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{TP}{TP + \frac{1}{2} * (FP + FN)} \qquad (2)$$

A confusion matrix is a table that shows how well a classification model performs by displaying the number of correct and incorrect predictions for each class. It helps visualise the model's accuracy and where it makes mistakes.

The performance of the regression models was evaluated using three metrics: R^2, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). R^2 measures the proportion of variance in the dependent variable explained by the independent variable in the regression model. Measures how well the model explains the variance in the data. MAE quantifies the average absolute difference between the

predicted and actual values, while RMSE calculates the square root of the average of the squared differences between the predicted and observed values.

Model training and data analysis were conducted using Python 3.9 with the sci-kit-learn library on a 14-core processor (Intel i9 12th generation; Intel Inc., Santa Clara, CA, USA) equipped with 64 GB of DDR5 RAM and an 8 GB GPU (NVIDIA GeForce RTX 3070 Ti; Nvidia Inc., Santa Clara, CA, USA).

3. RESULTS AND DISCUSSION

3.1 Cultivar discrimination

For cultivar discrimination, high F1 scores were obtained across all models, as shown in Figure 3. The 3D-CNN model achieved the highest F1 score (0.87), followed by the PLS model with an F1 score of 0.75. The superior performance of the CNN can be attributed to its ability to extract both spatial and spectral features from image data, enabling it to capture subtle morphological differences between cultivars that other models could not effectively identify (Ketkar & Moolayil, 2021).

The PLS model correctly classified a significant number of samples for some cultivars (e.g., 50 for cultivar 1 and 58 for cultivar 3). However, it showed notable misclassifications for others, such as cultivars 6, 7, and 17, where predictions were more dispersed across incorrect classes. In the case of the 3D-CNN, most samples were correctly classified for all cultivars. For example, cultivars such as 9, 10, and 13 show strong classification performance with minimal misclassifications.

These findings are consistent with earlier research using spectroscopy for cultivar classification but demonstrate improvements due to deep learning and image data integration. For example, Sánchez et al. (2012) used a PLS model to classify five strawberry cultivars based on spectral data, achieving accuracies ranging from 57% to 78%, which corresponds to an F1 score similar to that of the PLS model in this study (F1 = 0.75). In contrast, the CNN's F1 score of 0.87 highlights its ability to outperform traditional spectral approaches by leveraging additional spatial information.

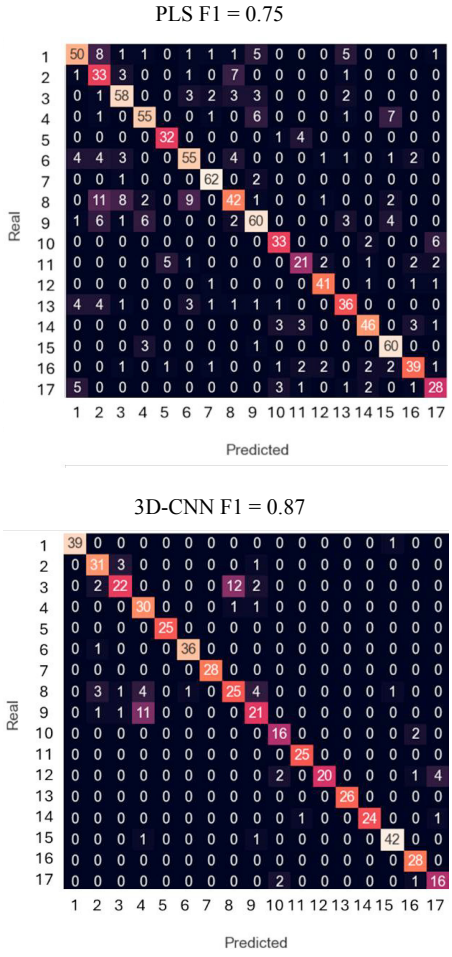

Figure 3. Confusion matrices of the cultivar discrimination on the test set using PLS (up) and the proposed 3D-CNN (down).

3.2 Prediction of TSS

Table 1 summarises the performance of the models for predicting TSS (°Brix) in strawberries, evaluated using the test set. The metrics include the coefficient of determination (R2), MAE and RMSE. The results indicate that the 3D-CNN model outperformed the PLS model in terms of R^2 (0.82 vs. 0.71), demonstrating superior predictive accuracy. Also, the 3D-CNN model achieved slightly lower error values for MAE (0.65 vs. 0.67) and RMSE (0.81 vs. 0.85), suggesting more consistent absolute predictions. Thus, the CNN model excelled at capturing complex spatial and spectral features across a highly variable dataset, underscoring its potential for applications requiring robust predictive accuracy under diverse conditions.

Table 1. Classification metrics of the TSS prediction using PLS and the 3D-CNN

Model	TSS (°Brix)		
	R^2	MAE	RMSE
PLS	0.71	0.67	0.85
3D-CNN	0.82	0.65	0.81

R^2 = coefficient of determination; MAE = mean absolute error; RMSE = root mean squared error; TSS = total soluble solids

3D-CNNs can exploit both spatial and spectral information, capturing complex patterns and local variations within the fruit that are lost when averaging spectra or using only spectral data.

This richer feature extraction enables more accurate modelling of TSS. In contrast, PLS with mean spectra ignores spatial context and thus cannot leverage the full potential of hyperspectral imaging (Audebert et al., 2019).

These findings align with previous studies that have demonstrated the effectiveness of spectral imaging and machine learning models for predicting quality parameters in strawberries, though differences in datasets and methodologies must be considered. For instance, ElMasry et al. (2007) reported a correlation coefficient (R^2) of 0.80 for TSS prediction using a PLS model with mean spectra from a single cultivar, which is comparable to the PLS model's performance in this study. Su et al. (2021) also used image data with a 3D-CNN model and achieved a lower R^2 value of 0.56 for TSS prediction on fruit from one cultivar at four ripeness levels. In contrast, the 3D-CNN model in this study achieved significantly higher accuracy (R^2 = 0.82), likely due to improvements in architecture and its application to a broader dataset encompassing 17 strawberry cultivars from two different origins.

4. CONCLUSIONS

This study highlights the superiority of 3D-CNNs over other traditional ML models, such as PLS, for predicting TSS and classifying strawberry cultivars using spectral imaging data. The results demonstrate that the 3D-CNN model achieved higher predictive accuracy for TSS, with an R2 of 0.82 compared to 0.71 for the PLS model. This underscores CNN's ability to capture complex spatial and spectral patterns in a diverse dataset comprising 17 strawberry cultivars from two origins. For cultivar discrimination, the 3D-CNN model significantly outperformed the PLS model, achieving an F1 score of 0.87 compared to 0.75 for the PLS model. The confusion matrices revealed that the CNN correctly classified most samples across all cultivars with minimal misclassifications. In contrast, the PLS model struggled with misclassifications in certain cultivars, reflecting its limitations in leveraging spatial information. These findings align with previous studies that demonstrated the potential of spectral imaging for quality prediction and cultivar classification but highlight significant improvements due to deep learning integration. This study supports the potential of combining spectral imaging with advanced machine learning techniques, such as 3D-CNNs, for effective quality assessment and cultivar classification in strawberries.

Unlike similar studies, this study examines a much larger set of strawberry varieties and samples. The simple and replicable model presented is effective for cultivar identification and easy for other researchers to use.

ACKNOWLEDGEMENTS

This work is part of the project PPS Smaakborging Groente en Fruit (Taste assurance of fruits and vegetables), partly funded by the Ministry of Economic Affairs, Secretariat Top Sector Horticulture and Starting Materials. The authors thank Fresh Forward, Brookberries, Onethird and Bakker Barendrecht for providing the fruit, support and technical supervision. Also, of MICIU AEI PID2023-150192OR-C31, C32 and C-33 with the support of FEDER funds and GVA-PROMETEO CIPROM/2021/014. Salvador Castillo Gironés thanks INIA for the FPI-INIA grant number PRE2020-094491, co-funded by EU EFS funds. Sandra Munera thanks the postdoctoral contract Juan de la Cierva-Formación (FJC2021-047786-I) co-funded by MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR.

REFERENCES

Audebert, N., Saux, B., & Lefèvre, S. (2019). Deep Learning for Classification of Hyperspectral Data: A Comparative Review. IEEE Geoscience and Remote Sensing Magazine, 7(2), 159–173. https://doi.org/10.1109/MGRS.2019.2912563

Chollet, F. (2021). Deep Learning with Python, Second Edition. Deep Learning with Python. https://www.oreilly.com/library/view/deep-learning-with/9781617296864/

ElMasry, G., Wang, N., ElSayed, A., & Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. Journal of Food Engineering, 81(1), 98–107. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2006.10.016

Ketkar, N., & Moolayil, J. (2021). Convolutional Neural Networks. Deep Learning with Python, 197–242. https://doi.org/10.1007/978-1-4842-5364-9_6

Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python : a guide for data scientists. 378. https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/

Polder, G., Dieleman, J. A., Hageraats, S., & Meinen, E. (2024). Imaging spectroscopy for monitoring the crop status of tomato plants. Computers and Electronics in Agriculture, 216, 108504. https://doi.org/https://doi.org/10.1016/j.compag.2023.108504

Pullanagari, R. R., & Li, M. (2021). Uncertainty assessment for firmness and total soluble solids of sweet cherries using hyperspectral imaging and multivariate statistics. Journal of Food Engineering, 289, 110177. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2020.110177

Sánchez, M.-T., De la Haba, M. J., Benítez-López, M., Fernández-Novales, J., Garrido-Varo, A., & Pérez-Marín, D. (2012). Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. Journal of Food Engineering, 110(1), 102–108. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2011.12.003

Su, Z., Zhang, C., Yan, T., Zhu, J., Zeng, Y., Lu, X., Gao, P., Feng, L., He, L., & Fan, L. (2021). Application of Hyperspectral Imaging for Maturity and Soluble Solids Content Determination of Strawberry With Deep

Learning Approaches. Frontiers in Plant Science, 12, 736334.
https://doi.org/10.3389/FPLS.2021.736334/BIBTEX