

Fedőlap

Magyar Agrár- és Élettudományi Egyetem

Budai Campus

Képesítő fordítás

Bognár Kitti Annamária

Agrár- és természettudományi szakfordító szakirányú továbbképzés

Vidékfejlesztés és Fenntartható Gazdaság Intézet

Idegen Nyelvi Tanszék

Budapest

2024

Belső címlap

Magyar Agrár- és Élettudományi Egyetem

Budai Campus

1. Feature Reduction for the Classification of Bruise Damage to Apple Fruit Using a Contactless FT-NIR Spectroscopy with Machine Learning

Az almatermésű gyümölcsök sérüléseinek osztályozására szolgáló tulajdonságok csökkentése az érintésmentes FT-NIR spektroszkópia és a gépi tanulás segítségével

2. Recent advancements in NIR spectroscopy for assessing the quality and safety of horticultural products: A comprehensive review

A NIR-spektroszkópia legfrissebb eredményei a kertészeti termények minőségének és biztonságának értékelésében: Átfogó áttekintés

Képesítőfordítás

Konzulens: Kissné Békássy Enikő
nyelvtanár

Hallgató: Bognár Kitti Annamária
G25PIG

Képzés: Agrár- és természettudományi szakfordító szakirányú továbbképzés

Vidékfejlesztés és Fenntartható Gazdaság Intézet

Idegen Nyelvi Tanszék

Budapest

2024

Köszönetnyilvánítás

Szeretném megköszönni témavezető tanáromnak, Kissné Békássy Enikőnek segítőkészségét, iránymutatását. Továbbá szeretném megköszönni Családomnak, Férjemnek és Kisfiamnak, hogy mindvégig mellettem álltak és segítettek.







Tartalomjegyzék

Forrásnyelvi szöveg 1	5
Az almatermésű gyümölcsök sérüléseinek osztályozására szolgáló tulajdonságok csökkentése az érintésmentes FT-NIR spektroszkópia és a gépi tanulás segítségével	23
1. Bevezetés	23
2. Anyagok és módszerek	25
2.1. Gyümölcsfajták	25
2.2. A kísérlet	25
2.3. Mintavétel	26
2.4. Adatok előzetes feldolgozása	27
2.5. Analitikai munkafolyamat	28
3. Eredmények	31
4. Vita	37
5. Következtetések	40
„A” függelék	41
Hivatkozások	43
Forrásnyelvi szöveg 2	49
A NIR-spektroszkópia legfrissebb eredményei a kertészeti termények minőségének és biztonságának értékelésében: Átfogó áttekintés	73
Bevezetés	74
A spektroszkópiai technikák alkalmazása	77
Következtetés	113
Hivatkozások	115
Magyar nyelvű összefoglaló	128
Magyar nyelvű összefoglaló idegen nyelvű fordítása	129

Forrásnyelvi szöveg 1

Article

Feature Reduction for the Classification of Bruise Damage to Apple Fruit Using a Contactless FT-NIR Spectroscopy with Machine Learning

Jean Frederic Isingizwe Nturambirwe ^{1,*}, Eslam A. Hussein ², Mattia Vaccari ^{1,2,3}, Christopher Thron ⁴, Willem Jacobus Perold ⁵ and Umezuruike Linus Opara ^{6,7}

- ¹ Eresearch Office, DVC Research and Innovation, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa
 - ² Inter-University Institute for Data Intensive Astronomy, Department of Physics and Astronomy, University of the Western Cape, Bellville 7535, South Africa
 - ³ Inter-University Institute for Data Intensive Astronomy, Department of Astronomy, University of Cape Town, Rondebosch 7701, South Africa
 - ⁴ Department of Science and Mathematics, Texas A&M University-Central Texas, Killeen, TX 76549, USA
 - ⁵ Department of Electrical and Electronic Engineering, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa
 - ⁶ SARCHI Postharvest Technology Research Laboratory, Africa Institute for Postharvest Technology, Faculty of AgriSciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa
 - ⁷ UNESCO International Centre for Biotechnology, Nsukka 410001, Nigeria
- * Correspondence: fisingizwe@uwc.ac.za



Citation: Nturambirwe, J.F.I.; Hussein, E.A.; Vaccari, M.; Thron, C.; Perold, W.J.; Opara, U.L. Feature Reduction for the Classification of Bruise Damage to Apple Fruit Using a Contactless FT-NIR Spectroscopy with Machine Learning. *Foods* **2023**, *12*, 210. <https://doi.org/10.3390/foods12010210>

Academic Editors: Jordi Riu and Barbara Giussani

Received: 4 November 2022

Revised: 19 December 2022

Accepted: 24 December 2022

Published: 3 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Spectroscopy data are useful for modelling biological systems such as predicting quality parameters of horticultural products. However, using the wide spectrum of wavelengths is not practical in a production setting. Such data are of high dimensional nature and they tend to result in complex models that are not easily understood. Furthermore, collinearity between different wavelengths dictates that some of the data variables are redundant and may even contribute noise. The use of variable selection methods is one efficient way to obtain an optimal model, and this was the aim of this work. Taking advantage of a non-contact spectrometer, near infrared spectral data in the range of 800–2500 nm were used to classify bruise damage in three apple cultivars, namely ‘Golden Delicious’, ‘Granny Smith’ and ‘Royal Gala’. Six prominent machine learning classification algorithms were employed, and two variable selection methods were used to determine the most relevant wavelengths for the problem of distinguishing between bruised and non-bruised fruit. The selected wavelengths clustered around 900 nm, 1300 nm, 1500 nm and 1900 nm. The best results were achieved using linear regression and support vector machine based on up to 40 wavelengths: these methods reached precision values in the range of 0.79–0.86, which were all comparable (within error bars) to a classifier based on the entire range of frequencies. The results also provided an open-source based framework that is useful towards the development of multi-spectral applications such as rapid grading of apples based on mechanical damage, and it can also be emulated and applied for other types of defects on fresh produce.

Keywords: variable selection; model optimisation; defect classification; machine learning; baseline; uncertainty quantification; feature reduction; quality control; bruise damage; apples

1. Introduction

Apple fruit are highly susceptible to mechanical damage resulting from handling practices during and after harvest. Such damage are characterized by tissue and cell deterioration and facilitate infections by microorganisms and disease spread, leading to fruit spoilage and thus postharvest loss. Damage prevention measures that are applicable to handling can help reduce bruise occurrence [1–3]. However, such measures are limited

by the requirement for personnel with expert training to do the handling, which is not always feasible, especially in developing countries [4]. Grading and sorting of produce based on the presence and degree of defects can help in repurposing them for appropriate uses, such as animal feed or processing when their consumer acceptability is not ideal for market display, and thus reduce the likelihood for further disease spread if fruit skin is broken [5,6]. This can also be an alternative and/or complementary solution for further reduction of losses and ensuring quality and safety of fresh produce.

Non-destructive techniques (NDT) for evaluating the presence of damage on fruit have seen improvements over the years [7], whereby techniques such as optical coherence tomography [8], multispectral imaging [9] and thermal imaging [10] among others, show promise for effective sorting and grading. Multispectral imaging relies on few, fixed wave bands that are most descriptive of a target defect or quality parameter and has the advantage that it can enable fast detection at industrial sorting speeds [11]. However, for each application the determination of specific wavebands that are the best determinant of the properties that are relevant to the application is required [12]. Vibrational spectroscopy is also a prominent option for defect detection, but produces high dimensional spectral data where many variables may contain information that is irrelevant to the problem at hand. The use of full spectra results in models that are complex and of which performance may be impaired by the inclusion of less informative variables. An efficient way of optimizing models in terms of simplicity and performance aims at selecting and including only variables that are most informative to the model [13].

There are various variable selection methods, but none is fit for all purposes [14]. However, studies on detecting various defects have reported successful model improvements using variable selection [15,16]. Using weighing coefficients of the best PC images, Huang et al. (2015) proposed 780, 850 and 960 nm as effective wavelengths within the range of 325–1100 nm, for detecting bruises on apples as an attempt to develop a multispectral imaging (MSI) system for online use. Nturambirwe et al. (2018) found that GA-PLS was consistently improving full spectra-based bruise classification models by a margin from 10% to 30% in terms of classification accuracy, when using a contact mode and 22 mm spot sample size scanning [17].

The enabling factors for industry applications include instrumental designs that are suitable for industrial systems such as capability for large sample exposure and fast scans for data acquisition, robust detection models and the ease of calibration transfers [18], as well as open platforms for collaborative development efforts. Other uses of NDT for food quality and material identification by various users, such as consumers and researchers have also known a growing interest [19] and favors user-friendly handheld and mobile spectrometer designs. The effectiveness of these applications rely on the proper identification of application specific and relevant wavelengths [20,21].

In this work, the aim was to determine the significant wavelengths for bruise discrimination in three apple cultivars, using a Fourier Transform Near-Infrared (FT-NIR) spectrometer that simulates online sample presentation (contactless exposure of large sample size up to 100 mm in diameter). The objectives were to first develop an open source software-based machine learning pipeline for modelling FT-NIR spectral data, secondly, establish the importance of wavelengths as it relates to bruise damage in apples, and lastly, provide a context of application prospects.

2. Materials and Methods

2.1. Fruit Material

Three apple cultivars, namely 'Golden Delicious' (GD) (yellowish green), 'Granny Smith' (GS) (green) and 'Royal Gala' (RG) (predominantly red), were acquired in two installments (in two consecutive months) from two different local retail shops from Stellenbosch, Western Cape, South Africa, in 2019. A batch of 100 apples were sourced first (source S1) and 114 apples were acquired in the second instance (source S2) with nearly

equal proportions of cultivars. Fruits that were free from visible defects were selected and used in the bruising experiment.

2.2. Experimental

The apples were kept in cold storage (5 °C, 85% RH) pending a bruising experiment and Fourier transform (FT)-NIR spectral measurements thereafter. They were left at room temperature for three hours prior to each bruising experiment, in order to carry out measurements at ambient, laboratory conditions (25 °C, 65% RH). Bruise damage was created by dropping a stainless steel ball from different heights (20, 35 and 65 cm) on two opposite sides of each apple, thus creating bruises with three degrees of severity. Experimental setup was done according to [6,22]. Two areas on opposite sides (bruised and non-bruised) around the equatorial plane of every apple were scanned under the non-contact emission head (EH) of the Matrix-F spectrometer (Matrix-F duplex from Bruker Optics, Ettlingen, Germany). For each single measurement, the spectrum was averaged over 64 scans. The NIR scanning range was between 12,500–4000 cm^{-1} , in intervals of 4 cm^{-1} [23]. The MATRIX-F FT-NIR spectrometer is equipped with a fiber optic NIR illumination and detection head (185 mm height and 230 mm diameter for sample sizes up to 100 mm in diameter) and allowed for measurement on half of the whole apple surface per single exposure. The fiber optic illumination head contains 4 air cooled tungsten NIR light sources (Tungsten halogen, 12 V, 20 W). The diffusely reflected light from the sample was collected and guided via a fiber optic cable to the spectrometer detector (a highly sensitive, thermoelectric cooled and temperature controlled InGaAs diode detector) [24].

2.3. Sampling

From the three apple cultivars, two main sample categories were created, namely bruised (B) and non-bruised (S) fruit. From the B category, three subcategories were created by representing the different levels of bruise severity (L1, L2, and L3), thus contributing to more variability in the data set.

The ratio between the two classes indicates that the data are balanced. Figure 1 shows the 3 apple data sets with 50 randomly selected infrared samples.

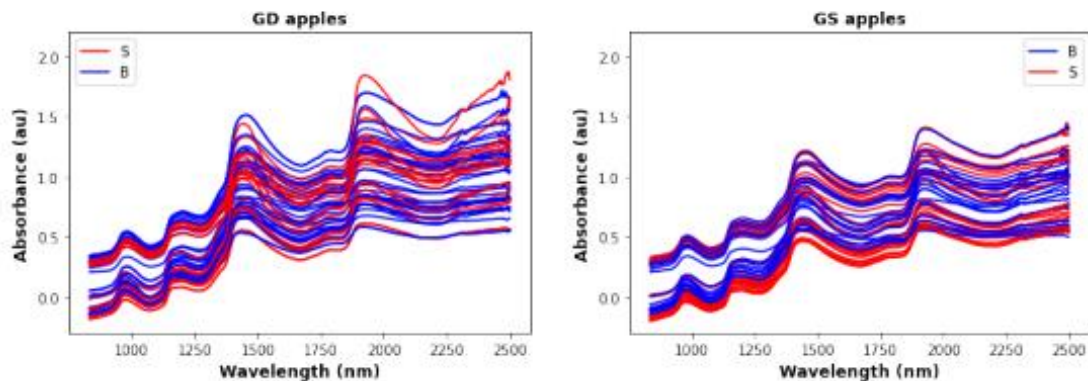


Figure 1. Cont.

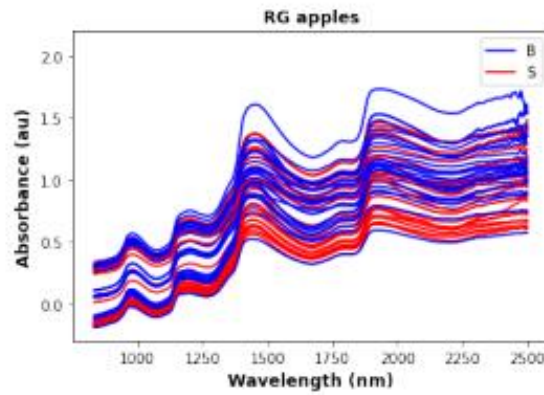


Figure 1. Infrared data for 50 randomly selected GD, GS, and RG, where B are the bad apples, and S are the good apples.

2.4. Data Pre-Processing

Various spectral preprocessing methods including Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV), derivatives, scaling and normalization can enhance the modeling outcome in spectral signals [25] data and were applied to the spectral signals. “Standard scaling” was found to be a most viable option. All three data sets were standardized using the StandardScaler library implemented in scikit-learn [26] and the Python programming language.

Figure 2 shows the standardized wavelength data for the three different apple types. Visual comparison with Figure 1, shows that standardization produces a clearer separation between B and S apples, especially for GS and RG apples.

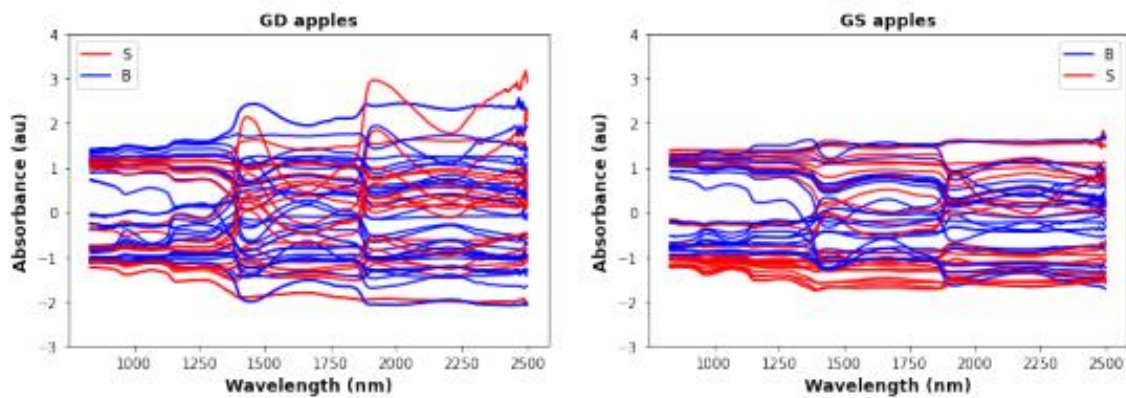


Figure 2. Cont.

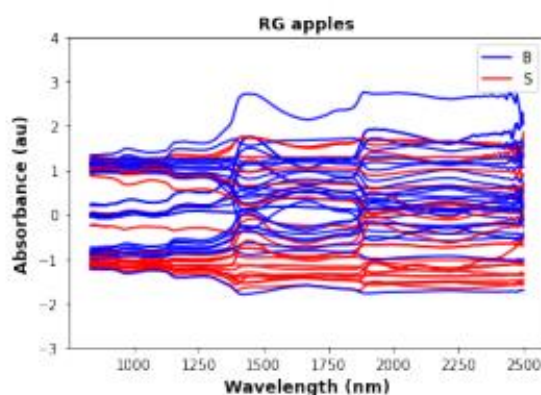


Figure 2. Infrared data after standardization for the same samples as in Figure 1.

2.5. Analytical Workflow

2.5.1. Introduction

Our main goal in this investigation is to reduce the number of features, thus simplifying the classification process. In this section, we describe the workflow used to select optimal features and machine learning methods for bruise classification. The code and the data together with the results, are available on Zenodo [27].

2.5.2. Baseline Method

We also established logistic regression (LR) on all 2000 wavelengths as a baseline predictor against which all other predictors can be compared. As a preliminary test, we measured feature permutation importance on all wavelengths using the `permutation_importance` function from `scikit_learn`. The permutation importance test measures the relative drop in accuracy when each individual feature is shuffled, thus destroying the correlation between that feature and samples [26,28]. This method has the advantage of fast execution (which is necessary since we are examining more than 2000 features). Figure 3 shows the relative importance of the first 500 features. The figure shows that there are clusters of wavelengths with higher feature importance. However, these wavelengths are supplying redundant information, since the intensity values vary only slightly between adjacent wavelengths. Thus we expect that many wavelengths can be eliminated as features without greatly reducing the predictive accuracy.

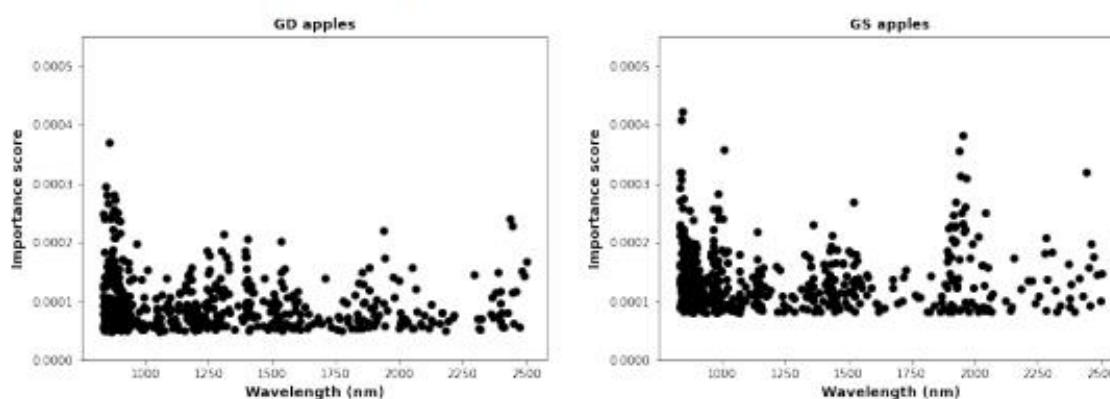


Figure 3. Cont.

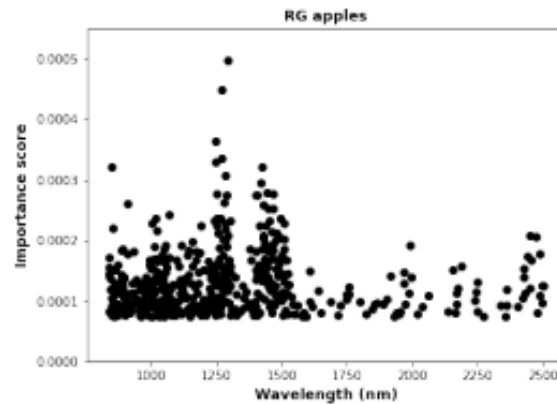


Figure 3. The graph shows the permutation importance for the top 500 features/wavelength.

2.5.3. Feature Selection

The permutation method clearly shows that some wavelengths are more important than others. However, the permutation method itself is unsuitable for feature selection, since it does not take into account the fact that two different relatively important features may be highly correlated and supply similar information.

To deal with selection among correlated features, there are two commonly-used methods, described as follows:

- In recursive feature elimination (RFE), the model is first trained on the entire set of features, and deletes the least important features. This training-deletion procedure is repeated recursively on the remaining features until the desired number of features remain.
- In sequential feature selection (SFS), initially models based on each individual feature are computed, and the feature with the best cross-validation score is selected. Then all models consisting of the selected features and one additional feature are computed, and the best chosen. This process is repeated recursively, adding one feature each time. The default cross validation score was used for this step. There is a backward selection variant of this method that is similar to RFS but we found that execution is too slow.

2.5.4. Machine Learning Classifiers

Mathematical models have been used extensively for chemometric studies of spectral data from non-contact FT-NIR acquired on fruit. Machine learning models are increasingly being used for this purpose [6,29] because of their flexibility and adaptability to a wide variety of applications. A total of six machine learning (ML) tools are used for the binary classification: logistic regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB), *k*-nearest neighbour (Knn), and Artificial neural networks (ANN). These methods are state-of-the-art in the literature for relatively small datasets [30,31] such as those described in Table 1.

Table 1. Repartition of the number of samples per apple cultivar.

Type	Non-Bruised Samples	Bruised Samples	Total
GD	274	273	547
GS	252	251	503
RG	278	284	562

All tools were optimized with 3-fold cross validation to avoid overfitting to make sure that a trained model can generalize on unseen data.

Optimization of ML parameters for all feature sets was performed using random search, as implemented in the Python library scikit-learn [26]. All models except for SVM and the baseline went through 30 randomized searches, whereas SVM went through only 15 because SVM is very computationally intensive.

2.5.5. Comparison of Classifier Performance

In this research, we compared the classification precision of all ML methods for all features, which is computed as

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

where *TP* and *FP* are the numbers of true positive and false positive predictions, respectively. Here a 'positive' result refers to identifying the apple as good: so a *FP* result falsely classifies a bruised apple as good. Precision is used because the most costly error is misidentifying bruised apples.

To make an effective comparison between models, error bars (corresponding to two standard deviations) for the classification precision differences were also calculated. If 0 lay outside the error bars, then we concluded that the difference between the two models was statistically significant. Otherwise, we failed to reject the null hypothesis of no difference between models.

The precision of the different classifiers were evaluated using the testing set. In order to obtain error bars, jackknife with leave-out-one was implemented [32,33]. For the training set error, the left-out instances were 1/3 of the training set, while for the testing set all instances were left out one-by-one to obtain the jackknife estimate of the standard deviation.

3. Results

In our workflow, we used both RFE and SFS methods with LR in order to get the best 50 features, which resulted in 50×2 classifiers, corresponding to taking the best feature, best 2 features, up to the best 50 features. The choice of 50 features was based on preliminary investigations using permutation importance, which showed that no gain in accuracy was achieved from using more than 50 features. Figure 4 shows the precision score for the best 50 feature sets using LR on training data. The figure shows that SFS gives better estimators for all three apples for all features sets. Consequently, for the testing results we only applied SFS and did not consider RFE. Note that the feature selection in Figure 4 employed all 2000+ features, which was a very computationally intensive calculation. Subsequently we found that virtually identical results could be obtained with much less computation time by doing feature selection with only the best 200 features obtained from permutation importance (shown in Figure 3).

Figure 5 shows graphically the wavelengths for the best 10, 20, 30, 40, and 50 features for all the three cultivars: the exact wavelength values are given in Appendix A. The *x* axis on each plot gives the wavelength, while the *y* axis shows the feature ranking in groups of 10. The figure shows that the 10 most important features are somewhat consistent among all three cultivars: wavelengths near 900, 1300, 1500 and 1900 nm appear as top-10 wavelengths for all three cultivars. Nonetheless, there are significant differences between the best features for the three cultivars. For example, only GS has wavelengths above 2400 in the top 10. It should be noted that in an industrial setting cultivars would be sorted separately, hence in practice a cultivar-specific approach to feature selection is desirable. We also note that as feature importance decreases, the features tend to get more clustered. For example, with GD apples the features with importance 41–50 form three clusters, while features with importance 0–30 are more diverse.

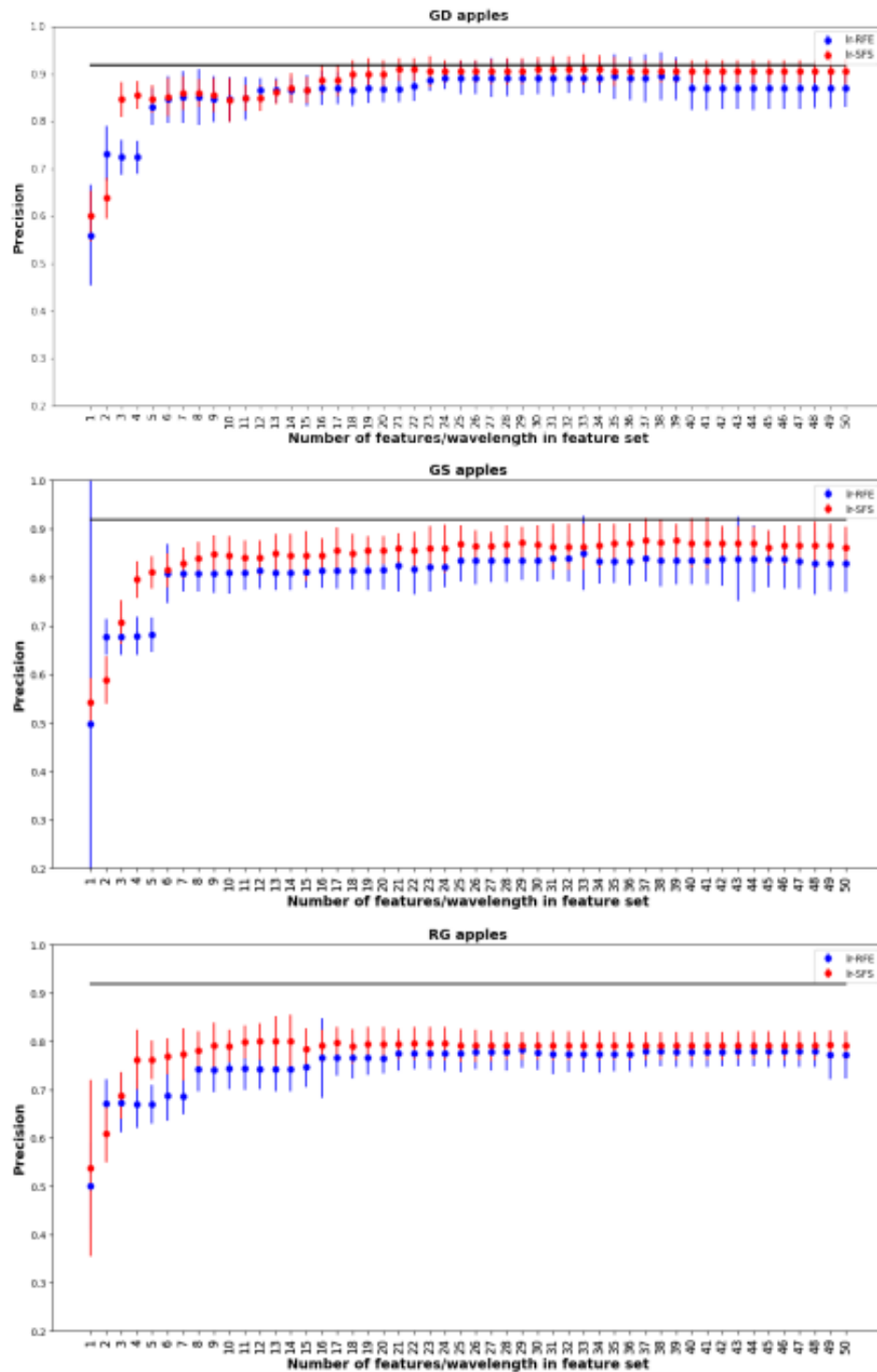


Figure 4. Best 50 feature sets using RFE and SFS on the 3 data sets using LR.

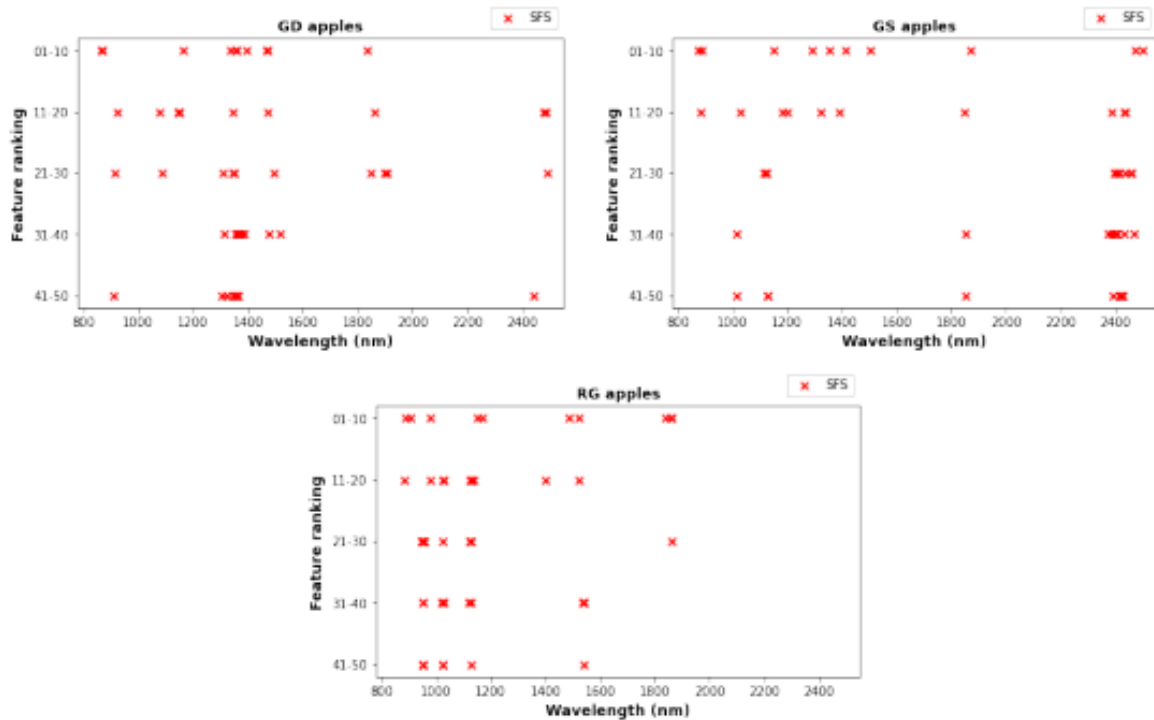


Figure 5. Allocation of the 50 selected features that are most relevant to bruise segregation for the three cultivars. Selected wavelengths are plotted in sets of 10 with the first selected on top. Detailed frequency information is given in Appendix A.

From the 50 different features sets we selected four for testing on the testing set. The four features sets were determined as follows: (a) Features up to and including the first jump in accuracy; (b) Best 10 features; (c) When the model stabilizes; (d) Best 50 features. Figure 6 compares the precision of models based on the four different feature sets. For each feature set, all six ML tools listed in Section 2.5.4 were implemented. In general precisions ranged from 0.7 to 0.9, with GD typically obtaining the best results. For each ML method feature sets (b)–(d) gave nearly the same performance, LR and SVM obtained similar mean precisions, but LR had much smaller error bars. The overall highest mean precisions for the different apple species were 0.86, 0.79, and 0.81 for GD, GS, and RG respectively. All of these best results were obtained with feature set (c).

The horizontal line in Figure 6 shows the precision obtained with the baseline estimator which was LR using all wavelengths. Compared to the baseline, the best-performing estimators were slightly lower mean precisions, but differences were small (between 0.01–0.03) and statistically insignificant. Although the reduced feature sets did not give better precision than the baseline, they require far fewer wavelength measurements and correspondingly are much easier to implement in practice.

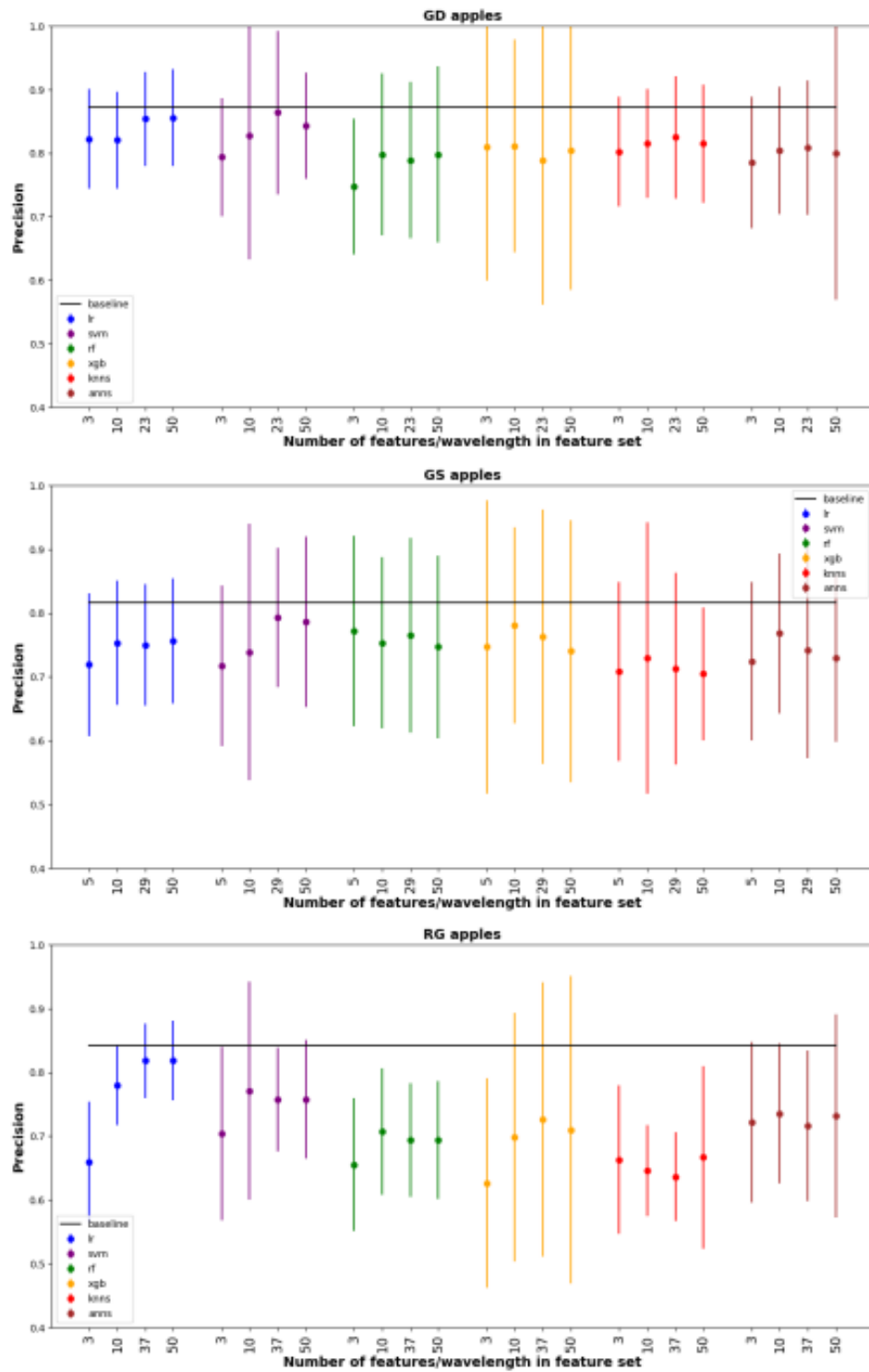


Figure 6. Plot of precision for all models using and all selected feature subsets, on test data with a comparison to the baseline model (horizontal line). The vertical lines represent error bars.

4. Discussion

The precision of classification for all models based on selected features subsets developed in this study did not exceed that of the baseline, which utilized all the wavelengths. A similar relative performance was also reported in a study to classify bruises in four apple cultivars by Luo et al. [16]; authors applied a ROC-AUC based method to select effective wavelengths for spectral imaging and found that reduced models did not exceed the performance expressed by sensitivity, specificity and accuracy of full-spectrum based model. It should be noted that though our binary classification focuses only on the existence or absence of bruises, our sampling has introduced high variability in the sample space by including different levels of bruising, sample origins and sampling times. High variability in the samples can be a challenge for some models to fully capture. However, it is a desirable aspect to build models with high generalisation ability. Furthermore, the reflective mode of spectral acquisition may introduce nonlinear relationships between chemical composition and spectra [34], as opposed to the linear relationships indicated by the Beer-Lambert law.

In all the cultivars, the LR and SVM models based on about 40 wavelengths had the highest average precision values (0.79–0.86). These results are comparable to the common range of classification performance metrics in spectroscopy-based bruise classification. Nonetheless, there has been reports [35] of higher model performance for bruise classification in apples using different absorption wavelength regions. There is some possibility that higher performance may be obtained using deep learning, which has been applied successfully in other agro-product classification scenarios [36]. However, effective training of a deep learning classifier typically requires a much larger dataset than was available for this study [32,37]. Furthermore, many studies that apply machine learning (including deep learning) do not include error bars in their performance evaluations, so there is some question as to whether reported improvements in classification are statistically significant [38].

The generalization ability of machine learning is an important aspect for effective applications. Therefore, all the models were tested on unseen data, previously separated from the original dataset. Nonetheless, given that data were acquired in a controlled manner, it should be noted that factors such as temperature may affect the acquisition of NIR spectra. Also, variability can be introduced by considering fruits from different climatic regions and growing conditions. To improve model generalization ability, these aspects could be considered in future studies.

Previous studies have examined the NIR waveband characteristics of bruises in apples. Geola et al. [39] proposed a procedure for detection of damaged tissue in ‘Golden Delicious’ apples, and a classification function was obtained in the region of 750 to 800 nm. Other apple tissue classification studies have used reflectance spectra, and in most cases the most significant wavelengths were found in the range of 690–850 nm [40–42]. Kleynen et al. [43], working with a VIS-NIR spectrometer developed a method to detect defects on bicolor apple fruit (‘Jonagold’). They found that the most significant wavelengths were in the NIR range (700–920 nm or 14,285–10,869 cm^{-1}). According to Lammertyn and associates, the light penetration depth of NIR radiation in Jonagold apple tissue was the highest in the region 700–900 nm [44]. Hence, Kleyman associated their observations to the latter wavelength region. Although these studies were generally based on short-wave NIR bands, considering the broad spectrum range from 780 nm to 5000 nm can improve detection of bruises with varying depths [10].

The chemistry of apple peel is very complex and has been adapted for its biological function. The peel’s outer surface consists of cuticle waxes composed of fatty alcohol, fatty acids and long chain hydrocarbons [45–47]. Additionally, there are different class of compounds such as flavonoids [48,49], phenolic acids [50,51], triterpene esters [52] and proanthocyanidins [53]. The color chemistry of the peel varies according to the cultivar of the apple under consideration; the major class of compounds contributing to the apple’s colors are chlorophylls (green) [54], carotenoids (yellow) [55] and anthocyanins (Red) [56].

The NIR spectrum has several overtones due to absorption from several different types of chemical bonds including C–H (from methyl, methylene, methoxy, aromatic, and carbonyl associated groups, N–H (from primary and secondary amides, primary, secondary, and tertiary amines, and amine salts) O–H (alcohols and water), S–H, and C=O groups [57].

In our case, it is difficult to localize any specific signals match with any of the functional group; from Figure 1, several broad bands can be identified which are common to all three apple types. The first band ranges from 850–1100 nm and includes two sub-bands: 850–920 nm (C–H methyl/methylene associated with aliphatic and/or aromatic skeletons and 950–1000 nm (OH and NH aliphatic / aromatic groups), A second band ranges from 1150–1300 nm is due to C=O, aromatic C–H and C=C groups, A third band ranges from 1350–1600 nm, and covers OH, free and hydrogen bonded and C–H aliphatics and aromatics, N–H aromatics and aliphatics, C=O aldehydes and ketones, A less prominent band ranges from 1800–1850 nm, and covers OH, CH of carbohydrate polymer. Finally, a band from 1900–2000 covers C=O amides/carboxylic groups, and OH/NH mainly aromatics.

Figure 7 shows how the most important features for bruise identification compare with the NIR spectra for the different apple types. In every case there are top-10 features associated with each band in the spectrum. However, the selected features tend to be located near the steep rising edge of the band, and not in the middle of the band range. The band associated with the most top-10 features is 1350–1600, which as described above includes absorption from several types of chemical bonds. Exact values of the top-10 features/wavelengths for each apple type are given in Appendix A.

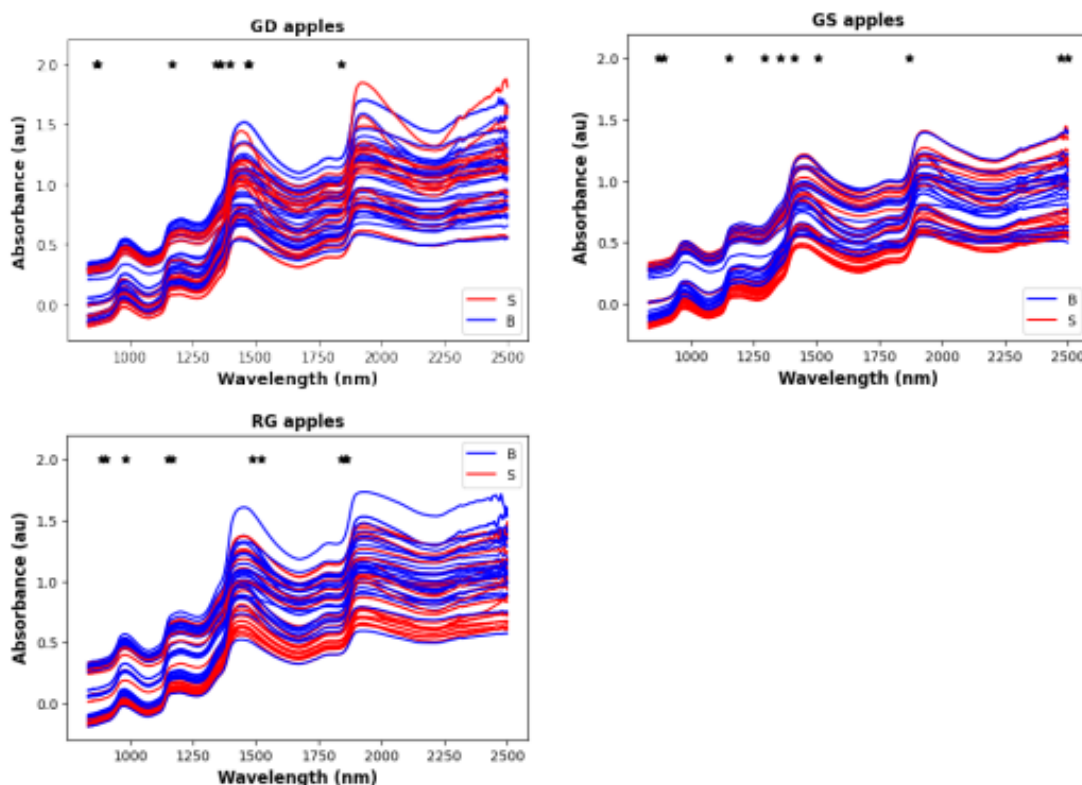


Figure 7. Position of top 10 features/wavelengths (shown as black asterisks) compared to the NIR spectra for the three different apple types.

Applications of spectroscopy to detect bruises seem to have had diminished interest in favor of imaging options. This may be related to the spatial limitations of spectroscopic devices, which offer only a limited spatial representation (up to 22 mm spot diameter) of the sample exposed. This limitation makes it inconvenient for most uses for fast sorting and grading of different types of fresh produce. The latest trends in non-destructive defect detection in agri-food fresh products have been focused on imaging techniques [58], including hyperspectral imaging. However, these imaging techniques rely heavily on image processing for feature extraction which is computationally costly. The emerging use of deep learning for spectroscopy and imaging based evaluation of agro-product quality offers a quick end-to-end modelling process [36], but the acquisition of images is still time consuming and not fit for industrial sorting speeds. Hence, multi-spectral imaging is preferred for such applications, since it is based on a few predetermined wavelengths that are effective for a specific attribute evaluation. This work shows the feasibility, with the flexibility of open-tools software, of wavelength selection on spectral data from fully exposed apples. This kind of exposure simulates sample presentation that would be applicable for inline sorting applications.

5. Conclusions

Building on the possibilities offered by a combination of open-source development tools and a contactless NIR spectrometer with large (100 mm in diameter) sample exposure, a feasibility study of bruise damage classification in apples was conducted, with the aim to determine the informative wavelengths in three apple cultivars. Bruise segregation models were built using six machine learning classification algorithms coupled with both, recursive feature elimination and sequential feature selection methods to determine the most informative wavelengths. A classification precision that matched the full-spectra based models within error bars could be achieved using up to 50 wavelengths selected from 4 main wavebands. The best reduced classification models were based on the LR and SVM machine learning techniques, which gave precision values ranging from 0.7 to 0.9 depending on the cultivar.

Author Contributions: Conceptualization, J.F.I.N., U.L.O. and W.J.P.; methodology, J.F.I.N., E.A.H. and C.T.; software, E.A.H. and M.V.; validation, C.T. and J.F.I.N.; formal analysis, E.A.H. and C.T.; investigation, J.F.I.N. and C.T.; resources, W.J.P., U.L.O. and M.V.; data curation, J.F.I.N.; writing—original draft preparation, J.F.I.N., E.A.H. and C.T.; writing—review and editing, C.T., W.J.P., U.L.O. and M.V.; visualization, E.A.H. and C.T.; project administration, M.V.; funding acquisition, W.J.P., U.L.O. and M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work is based on research supported by the of South African National Research Foundation (NRF Grant Numbers: 64813, 121291) and by the South African Department of Higher Education and Training Historically Disadvantaged Institutions Development Grant (DHET HDI-DG). The opinions, findings and conclusions or recommendations expressed are those of the authors alone, and the funding agencies accept no liability whatsoever in this regard. EAH and MV acknowledge financial support from the Inter-University Institute for Data Intensive Astronomy (IDIA), a partnership of the University of Cape Town, the University of Pretoria, the University of the Western Cape and the South African Radio Astronomy Observatory.

Data Availability Statement: The code, together with the results, is available on Zenodo at <https://zenodo.org/badge/latestdoi/478611734>. We also provided a walk-through educational tutorial which can be found on Zenodo at <https://doi.org/10.5281/zenodo.7018476>.

Acknowledgments: We wish to acknowledge the valuable contribution of Ahmed Hussein, who provided the chemical interpretations of the NIR spectra presented in the Discussion. We also want to thank the reviewers for their helpful comments that enabled us to make significant improvements to the paper. We also acknowledge the use of the ilifu cloud computing facility—(www.ilifu.ac.za), a partnership between the University of Cape Town, the University of the Western Cape, the University of Stellenbosch, Sol Plaatje University, the Cape Peninsula University of Technology and the South African Radio Astronomy Observatory. The ilifu facility is supported by contributions from the Inter-University Institute for Data Intensive Astronomy (IDIA)—a partnership between the University

of Cape Town, the University of Pretoria, the University of the Western Cape and the South African Radio Astronomy Observatory), the Computational Biology division at UCT and the Data Intensive Research Initiative of South Africa (DIRISA).

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

Appendix A

The appendix shows the best 50 features wavelength using the sequential feature selection model.

Table A1. Features/wavelengths for identifying bruised/sound apples in order of priority, for different apple types.

Priority	GD Wavelength (nm)	GS Wavelength (nm)	RG Wavelength (nm)
1	1393.9	2473.9	1840.1
2	864.5	1356.7	1486.6
3	1357.4	872.4	904.6
4	1468.1	1411.3	1168.9
5	1836.1	1292.4	888.5
6	1472.3	1150.2	980.6
7	868.8	1505.6	1152.3
8	1163.1	889.7	1522.4
9	1338.5	1870.6	1857.2
10	1354.6	2500.1	1861.2
11	1144.1	1180.6	976.5
12	1145.2	883.1	882.5
13	1862.5	1201.4	1136.1
14	2488.1	2432.1	1523.3
15	1151.3	2436.7	1399.2
16	1076.2	1391.6	1123.8
17	2478.6	1029.6	1022.3
18	924.0	2387.3	1028.0
19	1346.1	1321.4	1126.7
20	1473.9	1846.6	1127.2
21	1496.9	2455.2	949.0
22	1343.3	2439.0	953.9
23	915.2	2409.5	1862.5
24	2492.9	2396.2	1024.0
25	1309.4	1124.3	1125.3
26	1896.6	2459.8	1122.8
27	1909.2	1125.3	946.6
28	1850.6	2414.0	954.6
29	1351.0	2400.6	951.8

Table A1. Cont.

Priority	GD Wavelength (nm)	GS Wavelength (nm)	RG Wavelength (nm)
30	1086.6	1115.6	1127.7
31	1384.2	2372.0	1125.8
32	1376.9	2389.5	1537.7
33	1312.7	2398.4	1026.8
34	1477.3	1015.5	1541.4
35	1366.7	1851.9	1120.9
36	1518.8	2402.8	1019.5
37	1367.4	2407.3	953.5
38	1361.7	2405.0	952.8
39	1363.1	2434.4	1538.7
40	1356.7	2469.2	1025.6
41	1324.1	1128.2	949.7
42	1344.7	1853.2	952.1
43	1363.8	1126.7	1126.3
44	2441.3	1016.3	1539.6
45	1352.4	2425.3	951.4
46	1320.8	2416.3	950.0
47	1353.2	2391.7	951.1
48	1364.5	2418.5	952.5
49	1306.8	2429.8	1025.2
50	911.3	2423.0	1023.1

References

- Hussein, Z.; Fawole, O.A.; Opara, U.L. Harvest and Postharvest Factors Affecting Bruise Damage of Fresh Fruits. *Hortic. Plant J.* **2020**, *6*, 1–13. [\[CrossRef\]](#)
- Opara, U.L.; Pathare, P.B. Bruise damage measurement and analysis of fresh horticultural produce—A review. *Postharvest Biol. Technol.* **2014**, *91*, 9–24. [\[CrossRef\]](#)
- Al-Dairi, M.; Pathare, P.B.; Al-Yahyai, R.; Opara, U.L. Mechanical damage of fresh produce in postharvest transportation: Current status and future prospects. *Trends Food Sci. Technol.* **2022**, *124*, 195–207. [\[CrossRef\]](#)
- Hussein, Z.; Fawole, O.A.; Opara, U.L. Preharvest factors influencing bruise damage of fresh fruits—A review. *Sci. Hortic.* **2018**, *229*, 45–58. [\[CrossRef\]](#)
- Li, Z.; Thomas, C. Quantitative evaluation of mechanical damage to fresh fruits. *Trends Food Sci. Technol.* **2014**, *35*, 138–150. [\[CrossRef\]](#)
- Nturambirwe, J.F.; Nieuwoudt, H.H.; Perold, W.J.; Opara, U.L. Detecting bruise damage and level of severity in apples using a contactless nir spectrometer. *Appl. Eng. Agric.* **2020**, *36*, 257–270. [\[CrossRef\]](#)
- Nturambirwe, J.F.I.; Perold, W.J.; Opara, U.L. Classification learning of latent bruise damage to apples using shortwave infrared hyperspectral imaging. *Sensors* **2021**, *21*, 4990. [\[CrossRef\]](#)
- Zhou, Y.; Wu, D.; Hui, G.; Mao, J.; Liu, T.; Zhou, W.; Zhao, Y.; Chen, Z.; Chen, F. Loquat Bruise Detection Using Optical Coherence Tomography Based on Microstructural Parameters. *Food Anal. Methods* **2018**, *11*, 2692–2698. [\[CrossRef\]](#)
- Lu, Y.; Lu, R. Non-destructive defect detection of apples by spectroscopic and imaging technologies: A review. *Trans. ASABE* **2017**, *60*, 1765–1790. [\[CrossRef\]](#)
- Baranowski, P.; Mazurek, W.; Wozniak, J.; Majewska, U. Detection of early bruises in apples using hyperspectral data and thermal imaging. *J. Food Eng.* **2012**, *110*, 345–355. [\[CrossRef\]](#)
- Huang, W.; Li, J.; Wang, Q.; Chen, L. Development of a multispectral imaging system for online detection of bruises on apples. *J. Food Eng.* **2015**, *146*, 62–71. [\[CrossRef\]](#)
- Zhang, B.; Gu, B.; Tian, G.; Zhou, J.; Huang, J.; Xiong, Y. Challenges and solutions of optical-based nondestructive quality inspection for robotic fruit and vegetable grading systems: A technical review. *Trends Food Sci. Technol.* **2018**, *81*, 213–231. [\[CrossRef\]](#)

13. Xiaobo, Z.; Jiewen, Z.; Povey, M.J.; Holmes, M.; Hanpin, M. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* **2010**, *667*, 14–32. [CrossRef]
14. Mehmood, T.; Liland, K.H.; Snipen, L.; Sæbo, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69. [CrossRef]
15. Nturambirwe, J.F.I.; Opara, U.L. Machine learning applications to non-destructive defect detection in horticultural products. *Biosyst. Eng.* **2020**, *189*, 60–83. [CrossRef]
16. Luo, X.; Takahashi, T.; Kyo, K.; Zhang, S. Wavelength selection in vis/NIR spectra for detection of bruises on apples by ROC analysis. *J. Food Eng.* **2012**, *109*, 457–466. [CrossRef]
17. Nturambirwe, J.F.I.; Nieuwoudt, H.H.; Opara, U.L.; Perold, W.J. Performance of genetic algorithm in optimization of NIRS PLS models to predict apple fruit quality. *Acta Hort.* **2018**, *1201*, 355–362. [CrossRef]
18. Li, L.; Jang, X.; Li, B.; Liu, Y. Wavelength selection method for near-infrared spectroscopy based on standard-sample calibration transfer of mango and apple. *Comput. Electron. Agric.* **2021**, *190*, 106448. [CrossRef]
19. Abasi, S.; Minaei, S.; Jamshidi, B.; Fathi, D. Dedicated non-destructive devices for food quality measurement: A review. *Trends Food Sci. Technol.* **2018**, *78*, 197–205. [CrossRef]
20. Amuah, C.L.; Teye, E.; Lamptey, F.P.; Nyandey, K.; Opoku-Ansah, J.; Adueming, P.O.W. Feasibility Study of the Use of Handheld NIR Spectrometer for Simultaneous Authentication and Quantification of Quality Parameters in Intact Pineapple Fruits. *J. Spectrosc.* **2019**, *2019*, 5975461. [CrossRef]
21. Srivastava, S.; Shashikant, B.V. Handheld, smartphone based spectrometer for rapid and nondestructive testing of citrus cultivars. *J. Food Meas. Charact.* **2021**, *15*, 892–904. [CrossRef]
22. Opara, L.U.; AlGhafri, A.; Agzoun, H.; Allssai, J.; Aljabri, F. Design and development of a new device for measuring susceptibility to impact damage of fresh produce. *New Zealand J. Crop. Hortic. Sci.* **2007**, *35*, 245–251. [CrossRef]
23. Magwaza, L.S.; Opara, U.L.; Terry, L.A.; Landahl, S.; Cronje, P.J.; Nieuwoudt, H.; Mouazen, A.M.; Saeys, W.; Nicolai, B.M. Prediction of 'Nules Clementine' mandarin susceptibility to rind breakdown disorder using Vis/NIR spectroscopy. *Postharvest Biol. Technol.* **2012**, *74*, 1–10. [CrossRef]
24. Bruker. Matrix-F FT-NIR Spectrometer. Available online: <https://www.bruker.com/en/products-and-solutions/infrared-and-raman/ft-nir-spectrometers/matrix-f-ft-nir-spectrometer.html> (accessed on 15 December 2022).
25. Ning, J.; Ye, H.; Sun, Y.; Zhang, J.; Mei, Z.; Xiong, S.; Zhang, S.; Li, Y.; Hui, G.; Yi, X.; et al. Study on apple damage detecting method based on relaxation single-wavelength laser and convolutional neural network. *J. Food Meas. Charact.* **2022**, *16*, 3321–3330. [CrossRef]
26. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API design for machine learning software: Experiences from the scikit-learn project. *arXiv* **2013**, arXiv:1309.0238.
27. Hussein, E. Apple Classification, first release. *Zenodo v1.0* **2022**. Available online: <https://zenodo.org/record/7007001/export/dcite4#Y7PBHRVBxPY> (accessed on 15 March 2012). [CrossRef]
28. Altmann, A.; Tološi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340–1347. [CrossRef]
29. Nturambirwe, J.F.I.; Nieuwoudt, H.H.; Perold, W.J.; Opara, U.L. Non-destructive measurement of internal quality of apple fruit by a contactless NIR spectrometer with genetic algorithm model optimization. *Sci. Afr.* **2019**, *3*, e00051. [CrossRef]
30. Mendez, K.M.; Reinke, S.N.; Broadhurst, D.I. A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification. *Metabolomics* **2019**, *15*, 1–15. [CrossRef]
31. Aguasca-Colomo, R.; Castellanos-Nieves, D.; Méndez, M. Comparative analysis of rainfall prediction models using machine learning in islands with complex orography: Tenerife Island. *Appl. Sci.* **2019**, *9*, 4931. [CrossRef]
32. Hussein, E.A.; Thron, C.; Ghaziasgar, M.; Vaccari, M.; Marnewick, J.L.; Hussein, A.A. Comparison of Phenolic Content and Antioxidant Activity for Fermented and Unfermented Rooibos Samples Extracted with Water and Methanol. *Plants* **2022**, *11*, 16. [CrossRef]
33. Hussein, E.A.; Ghaziasgar, M.; Thron, C.; Vaccari, M.; Bagula, A. Basic Statistical Estimation Outperforms Machine Learning in Monthly Prediction of Seasonal Climatic Parameters. *Atmosphere* **2021**, *12*, 539. [CrossRef]
34. Nadler, B.; Coifman, R.R. Partial least squares, Beer's law and the net analyte signal: Statistical modeling and analysis. *J. Chemom.* **2005**, *19*, 45–54. [CrossRef]
35. Siedliska, A.; Baranowski, P.; Mazurek, W. Classification models of bruise and cultivar detection on the basis of hyperspectral imaging data. *Comput. Electron. Agric.* **2014**, *106*, 66–74. [CrossRef]
36. Zhang, X.; Yang, J.; Lin, T.; Ying, Y. Food and agro-product quality evaluation based on spectroscopy and deep learning: A review. *Trends Food Sci. Technol.* **2021**, *112*, 431–441. [CrossRef]
37. Purushotham, S.; Meng, C.; Che, Z.; Liu, Y. Benchmarking deep learning models on large healthcare datasets. *J. Biomed. Inform.* **2018**, *83*, 112–134. [CrossRef]
38. Hussein, E.A.; Ghaziasgar, M.; Thron, C.; Vaccari, M.; Jafta, Y. Rainfall Prediction Using Machine Learning Models: Literature Survey. In *Artificial Intelligence for Data Science in Theory and Practice*; Alloghani, M., Thron, C., Subair, S., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 75–108. [CrossRef]
39. Geolla, F.; Geolla, F.; Peiper, U.M. A spectrophotometric method for detecting surface bruises on "Golden Delicious" apples. *J. Agric. Eng. Res.* **1993**, *58*, 47–51. [CrossRef]

40. Upchurch, B.L.; Affeldt, H.A.; Hruschka, W.R.; Norris, K.H.; Throop, J.A. Spectrophotometric study of bruises on whole, 'Red Delicious' apples. *Trans. Am. Soc. Agric. Eng.* **1990**, *33*, 585–589. [[CrossRef](#)]
41. Miller, W.M.; Throop, J.A.; Upchurch, B.L. Pattern recognition models for spectral reflectance evaluation of apple blemishes. *Postharvest Biol. Technol.* **1998**, *14*, 11–20. [[CrossRef](#)]
42. Wen, Z.; Tao, Y. Building a rule-based machine-vision system for defect inspection on apple sorting and packing lines. *Expert Syst. Appl.* **1999**, *16*, 307–313. [[CrossRef](#)]
43. Kleynen, O. Selection of the most efficient wavelength bands for 'Jonagold' apple sorting. *Postharvest Biol. Technol.* **2003**, *30*, 221–232. [[CrossRef](#)]
44. Lammertyn, J.; Peirs, A.; De Baerdemaeker, J.; Nicolai, B. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biol. Technol.* **2000**, *18*, 121–132. [[CrossRef](#)]
45. Lu, P.L.; Lin, C.H. Physiology of fruit cracking in wax apple (*Syzygium samarangense*). *Bot. Orient. J. Plant Sci.* **2011**, *8*, 70–76. [[CrossRef](#)]
46. Wang, B.H.; Cao, J.J.; Zhang, B.; Chen, H.Q. Structural characterization, physicochemical properties and α -glucosidase inhibitory activity of polysaccharide from the fruits of wax apple. *Carbohydr. Polym.* **2019**, *211*, 227–236. [[CrossRef](#)] [[PubMed](#)]
47. Liu, Y.P.; Xie, Z.; Guan, R.Q.; Du, M.R.; Qiao, Z.H.; Suo, X.Y.; Liu, Z.Y.; Bian, Y.; Qiang, L.; Fu, Y.H. Syzysmalactone, an Unusual 11-Carbon δ -Lactone Derivative from the Fresh Ripe Fruits of *Syzygium samarangense* (Wax Apple). *J. Nat. Prod.* **2022**, *85*, 2100–2103. [[CrossRef](#)] [[PubMed](#)]
48. Van der Sluis, A.; Dekker, M.; Jongen, W. Flavonoids as bioactive components in apple products. *Cancer Lett.* **1997**, *114*, 107–108. [[CrossRef](#)]
49. Hagen, S.F.; Solhaug, K.A.; Bengtsson, G.B.; Borge, G.I.A.; Bilger, W. Chlorophyll fluorescence as a tool for non-destructive estimation of anthocyanins and total flavonoids in apples. *Postharvest Biol. Technol.* **2006**, *41*, 156–163. [[CrossRef](#)]
50. Ceymann, M.; Arrigoni, E.; Schärer, H.; Nising, A.B.; Hurrell, R.F. Identification of apples rich in health-promoting flavan-3-ols and phenolic acids by measuring the polyphenol profile. *J. Food Compos. Anal.* **2012**, *26*, 128–135. [[CrossRef](#)]
51. Lancaster, J.E.; Reay, P.F.; Norris, J.; Butler, R.C. Induction of flavonoids and phenolic acids in apple by UV-B and temperature. *J. Hort. Sci. Biotechnol.* **2000**, *75*, 142–148. [[CrossRef](#)]
52. Poirier, B.C.; Buchanan, D.A.; Rudell, D.R.; Mattheis, J.P. Differential partitioning of triterpenes and triterpene esters in apple peel. *J. Agric. Food Chem.* **2018**, *66*, 1800–1806. [[CrossRef](#)]
53. Lancaster, J.; Dougall, D.K. Regulation of skin color in apples. *Crit. Rev. Plant Sci.* **1992**, *10*, 487–502. [[CrossRef](#)]
54. Schwartz, S.J.; Lorenzo, T.V. Chlorophylls in foods. *Crit. Rev. Food Sci. Nutr.* **1990**, *29*, 1–17. [[CrossRef](#)] [[PubMed](#)]
55. Schweiggert, R.M.; Vargas, E.; Conrad, J.; Hempel, J.; Gras, C.C.; Ziegler, J.U.; Mayer, A.; Jiménez, V.; Esquivel, P.; Carle, R. Carotenoids, carotenoid esters, and anthocyanins of yellow-, orange-, and red-peeled cashew apples (*Anacardium occidentale* L.). *Food Chem.* **2016**, *200*, 274–282. [[CrossRef](#)] [[PubMed](#)]
56. Marais, E.; Jacobs, G.; Holcroft, D.M. Postharvest irradiation enhances anthocyanin synthesis in apples but not in pears. *HortScience* **2001**, *36*, 738–740. [[CrossRef](#)]
57. Jamrógiewicz, M. Application of the near-infrared spectroscopy in the pharmaceutical technology. *J. Pharm. Biomed. Anal.* **2012**, *66*, 1–10. [[CrossRef](#)]
58. Mahanti, N.K.; Pandiselvam, R.; Kothakota, A.; Ishwarya S., P.; Chakraborty, S.K.; Kumar, M.; Cozzolino, D. Emerging non-destructive imaging techniques for fruit damage detection: Image processing and analysis. *Trends Food Sci. Technol.* **2022**, *120*, 418–438. [[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.

Az almatermésű gyümölcsök sérüléseinek osztályozására szolgáló tulajdonságok csökkentése az érintésmentes FT-NIR spektroszkópia és a gépi tanulás segítségével

Jean Frederic Ingingizwe Nturambirwe ^{1,*}, Eslam A. Hussein ², Mattia Vaccari ^{1,2,3}, Christopher Thron ⁴, Willem Jacobus Perold ⁵ and Umezuruike Linus Opara ⁶,

Absztrakt: A spektroszkópiai adatok alkalmasak biológiai rendszerek modellezésére, például a kertészeti termékek minőségi paramétereinek előrejelzésére. Azonban a széles hullámhosszspektrum használata nem praktikus a termelésben. Az ilyen adatok magas dimenziószámúak, és általában összetett, nehezen érthető modelleket eredményeznek. Továbbá a különböző hullámhosszok közötti kollinearitás azt eredményezi, hogy egyes adatváltozók redundánsak, és akár a zajhoz is hozzájárulhatnak. A változóselektációs módszerek alkalmazása az optimális modell elérésének egyik hatékony módja, és ez a célja ennek a munkának is. Egy érintésmentes spektrométer előnyeit kihasználva a 800-2500 nm közötti tartományban a közeli infravörös spektrális adatokat három almafajta, nevezetesen a "Golden Delicious", a "Granny Smith" és a "Royal Gala" sérüléseinek osztályozásához használták. Hat kiemelkedő gépi tanulási osztályozó algoritmust alkalmaztak, és két különböző kiválasztási módszert használtak a sérült és nem sérült gyümölcsök megkülönböztetésének szempontjából legrelevánsabb hullámhosszok meghatározására. A kiválasztott hullámhosszok a 900 nm, az 1300 nm, az 1500 nm és az 1900 nm tartományokban csoportosultak.

Kulcsszavak: változó szelekció; modelloptimalizálás; hibaosztályozás; gépi tanulás; alapvonal; mérési bizonytalanság számszerűsítése; tulajdonságok csökkentése; minőségellenőrzés; sérülések; alma

1. Bevezetés

Az almatermésűek nagyon érzékenyek a betakarítás alatti és utáni kezelési gyakorlatokból eredő mechanikai sérülésekre. Az ilyen sérülések a szövetek és sejtek károsodásával járnak, és megkönnyítik a mikroorganizmusok általi fertőzést és a betegségek terjedését, ami a gyümölcs romlásához és így a betakarítás utáni veszteségekhez vezet. A sérülések megelőzésére irányuló, a feldolgozásra alkalmazható módszerek segíthetnek a zúzódások előfordulásának csökkentésében [1-3]. Az ilyen jellegű intézkedéseket azonban korlátozza, hogy a kezeléshez szaktudással rendelkező munkaerőre van szükség, ami nem mindig megvalósítható, különösen a fejlődő országokban [4]. A termékek osztályozása és válogatásuk a sérülések jelenléte és mértéke alapján segíthet a megfelelő felhasználási célokra, például állati takarmányozásra vagy

feldolgozásra való újra felhasználásban, ha a termék fogyasztói elfogadhatósága nem ideális a piaci forgalomba hozatalhoz, és így csökkenthető a gyümölcshéj sérülése esetén a betegség tovább terjedésének valószínűsége [5,6]. Ez alternatív és/vagy kiegészítő megoldás lehet a károk további csökkentésére, valamint a friss termékek minőségének és biztonságának biztosítására.

A gyümölcsök sérülésének értékelésére szolgáló roncsolásmentes technikák ("Non-destructive techniques", NDT) az évek során sokat fejlődtek [7], és többek között az optikai koherencia tomográfia [8], a multispektrális képalkotás [9] és a hőképalkotás [10] ígéretesnek bizonyulnak a hatékony szétválogatás és osztályozás szempontjából. A multispektrális képalkotás néhány, rögzített hullámsávra támaszkodik, amelyek a legjobban jellemzik a célzott hiányosságokat vagy minőségi paramétereket, és előnye, hogy lehetővé teszi a gyors észlelést ipari szortírozási sebesség mellett [11]. Minden egyes felhasználási területre vonatkozóan azonban meg kell határozni azokat a konkrét hullámsávokat, amelyek a legjobban meghatározzák az adott felhasználás szempontjából releváns tulajdonságokat [12]. A rezgésspektroszkópia szintén kiemelkedő lehetőség a hibák felderítésére, de nagy dimenziójú spektrális adatokat eredményez, ahol sok olyan változót tartalmazhat információt, amely az adott probléma szempontjából irreleváns. A teljes spektrumok használata olyan modelleket eredményez, amelyek összetettek, és amelyek teljesítménye romolhat a kevésbé informatív változók bevonásával. A modellek egyszerűség és teljesítmény szempontjából történő optimalizálásának hatékony módja az, hogy csak az alábbiakat válasszuk ki és vegyük figyelembe a modell szempontjából leginkább információt adó változókat [13].

Többféle változó szelekciós módszer létezik, azonban egyik sem alkalmazható minden célra [14]. A különböző hibák kimutatásával foglalkozó kutatások azonban sikeres modellfejlesztésekről számoltak be a változószelekció alkalmazásával [15,16]. Huang és munkatársai (2015) a legjobb számítógépes képek súlyozási együtthatóit felhasználva a 325-1100 nm-es tartományon belül a 780, 850 és 960 nm-t javasolták eredményes hullámhosszként az almán lévő horzsolások felismerésére, egy online használatra szánt multispektrális képalkotó ("multispectral imaging", MSI) rendszer kifejlesztésére tett kísérlet keretében. Nturambirwe és munkatársai (2018) azt tapasztalták, hogy a GA-PLS következetesen javította a teljes spektrumalapú horzsolásosztályozó modelleket, 10% és 30% közötti különbséggel az osztályozási pontosság vonatkozásában, amikor kontakt üzemmódban és 22 mm-es pontszerű mintamérettel végzett szkennelést alkalmaztak [17].

Az iparban történő alkalmazást támogató szempontok közé tartoznak az ipari rendszerekhez illeszkedő műszerkonstrukciók, mint például a nagy mintaexponícióra való képesség és a gyors letapogatás az adatgyűjtéshez, a robusztus detektálási modellek és a kalibrációs transzferek egyszerűsége [18], valamint a kollaboratív fejlesztésekhez szükséges nyílt platformok. Az élelmiszer-minőség és az anyagok azonosítására szolgáló NDT (roncsolásmentes vizsgálat) más felhasználási módjai a különböző felhasználók, például a fogyasztók és a kutatók részéről szintén növekvő érdeklődést mutatnak [19], és a felhasználóbarát kézi és mobil spektrométerek kialakítását részesítik előnyben. Az ilyen alkalmazások hatékonysága az alkalmazásspecifikus és releváns hullámhosszok megfelelő azonosításán múlik [20,21].

A kutatás célja az volt, hogy három almafajta esetében meghatározzuk a sérülések megkülönböztetése szempontjából jelentős hullámhosszokat egy Fourier-transzformációs közeli infravörös (FT-NIR) spektrométer segítségével, amely online mintaszolgáltatást modellez (nagy, akár 100 mm átmérőjű minták érintkezés nélküli exponícióját). A célkitűzések a következők voltak: először egy nyílt forráskódú szoftveralapú gépi tanulási csővezeték kifejlesztése az FT-NIR spektrális adatok modellezésére, másodsorban a hullámhosszok fontosságának megállapítása az alma zúzódásos sérüléseivel kapcsolatban, végül pedig az alkalmazás lehetőségeinek áttekintése.

2. Anyagok és módszerek

2.1. Gyümölcsfajták

Három almafajtát, azaz a "Golden Delicious" (GD) (sárgászöld), a "Granny Smith" (GS) (zöld) és a "Royal Gala" (RG) (túlnyomórészt piros) almát 2019-ben két különböző helyi kiskereskedelmi üzletből (két egymást követő hónapban) szereztük be két különböző helyi kiskereskedelmi üzletből, Stellenbosch városából (Western Cape, Dél-Afrika). Először egy 100 almából álló tételt szereztünk be (S1 beszerzési forrás), másodjára pedig 114 almát biztosítottunk (S2 beszerzési forrás), melyek közel azonos arányban tartalmazták a három almafajtát. Olyan gyümölcsöket választottunk ki és használtunk fel a sérülési kísérletben, amelyek látható hibáktól mentesek voltak.

2.2. A kísérlet

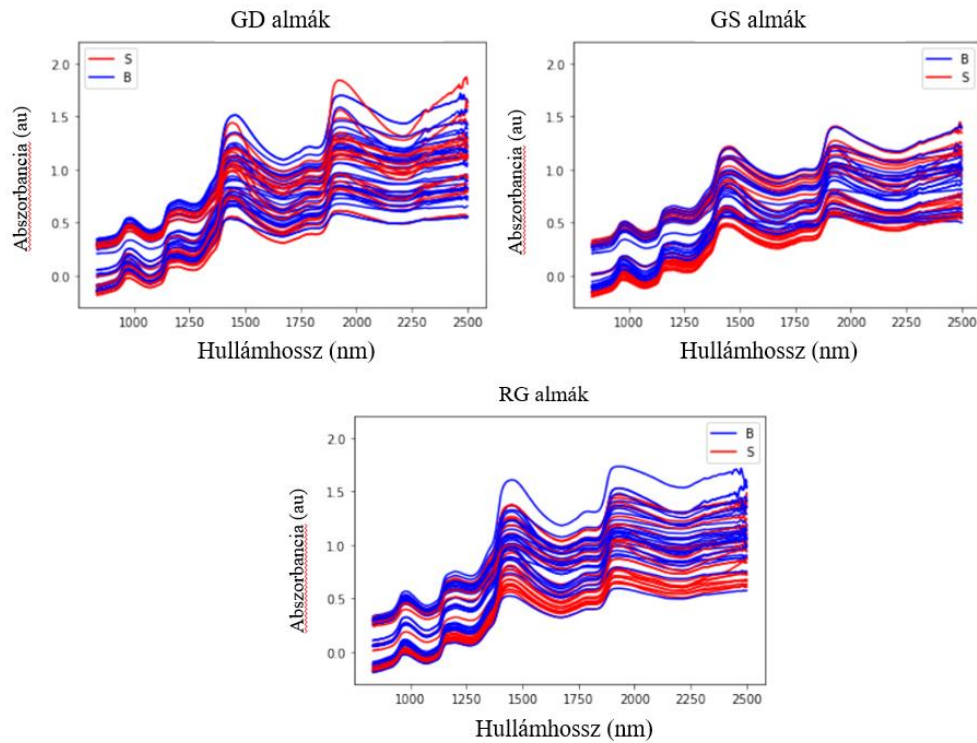
Az almákat hűtött helyen (5°C, 85%-os relatív páratartalom) tároltuk a sérülési kísérletig és az azt követő Fourier-transzformációs (FT)-NIR spektrális mérések elvégzéséig. Az egyes

zúzódási kísérleteket megelőzően három órán át szobahőmérsékleten tartottuk őket, hogy a méréseket környezeti, laboratóriumi körülmények között (25°C, 65% relatív páratartalom) végezhessük el. A sérüléseket úgy idéztük elő, hogy egy rozsdamentes acélgolyót - különböző magasságból (20, 35 és 65 cm) - az egyes almák két ellentétes oldalára ejtettük, így három különböző súlyosságú sérülést okoztunk. A kísérleti elrendezés a [6,22] alapján történt. Minden alma ekvatoriális síkja körül két ellentétes oldali (sérült és nem sérült) területet vizsgáltunk a Matrix-F spektrométer (Matrix-F duplex, Bruker Optics, Ettlingen, Németország) érintésmentes emissziós feje (EH) alatt. Minden egyes mérésnél a spektrumot 64 szkennelésen keresztül átlagoltuk. A NIR pásztázási tartomány 12 500-4000 cm^{-1} között volt, 4 cm^{-1} intervallumokban [23]. A MATRIX-F FT-NIR spektrométer száloptikás NIR megvilágító és detektáló fejjel volt felszerelve (185 mm magasságú és 230 mm átmérőjű, legfeljebb 100 mm átmérőjű mintákhoz), egyetlen expozícióval lehetővé téve a teljes almafelület feléig kiterjedő mérést. A száloptikai megvilágítófej 4 léghűtésű volfrám NIR fényforrást tartalmaz (volfrám halogén, 12 V, 20 W). A mintáról szórtan visszaverődő fényt összegyűjtöttük és egy száloptikai kábelon keresztül a spektrométer detektorához (egy nagy érzékenységű, termoelektromos hűtésű és hőmérséklet-szabályozott InGaAs diódadetektor) vezettük [24].

2.3. Mintavétel

A három almafajtából két fő mintakategóriát hoztunk létre, nevezetesen a sérült ("B" - bruised) és a nem sérült ("S" – non-bruised) gyümölcsöket. A B kategóriából három alkategóriát hoztunk létre a zúzódások különböző súlyossági szintjeinek (L1, L2 és L3 - levels of bruise severity) megfelelően, így hozzájárulva az adathalmaz nagyobb variabilitásához.

A két csoport közötti arány azt jelzi, hogy az adatok kiegyensúlyozottak. Az 1. ábra a 3 alma adatsorát mutatja 50 véletlenszerűen kiválasztott mintából.

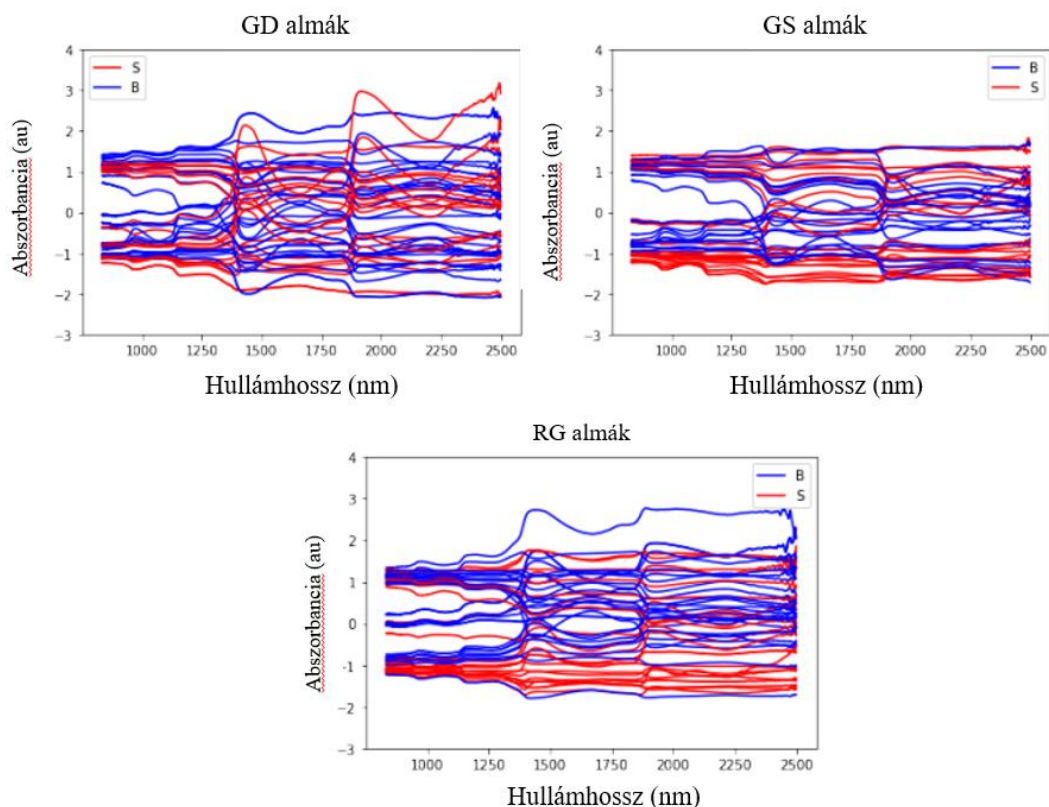


1. ábra: Infravörös adatok 50 véletlenszerűen kiválasztott GD, GS és RG esetében, ahol B a sérült alma, S pedig a nem sérült alma

2.4. Adatok előzetes feldolgozása

Különböző spektrális előfeldolgozási módszerek - beleértve a multiplikatív szórás-korrekciót (MSC), a standard normálváltozót (SNV), a deriválásokat, a skálázást és a normalizálást, javíthatják a modellezés eredményét a spektrális jelek [25] adatainál, és ezeket a spektrális jelekre már korábban alkalmazták. A "standard skálázás" a leginkább megvalósítható lehetőségnek bizonyult. Mindhárom adatkészletet a scikit-learn (nyílt forráskódú gépi tanulási projekt) [26] és a Python programozási nyelvben implementált StandardScaler könyvtár segítségével standardizáltuk.

A 2. ábra a három különböző almatípusra vonatkozó standardizált hullámhosszadatokat mutatja. Az 1. ábrával való vizuális összehasonlítás azt szemlélteti, hogy a standardizálás egyértelműbb különbséget tesz a „B” és „S” almák között, különösen a GS és RG almák vonatkozásában.



2. ábra: Infravörös adatok standardizálás után az 1. ábrán látható mintákkal összehasonlítva

2.5. Analitikai munkafolyamat

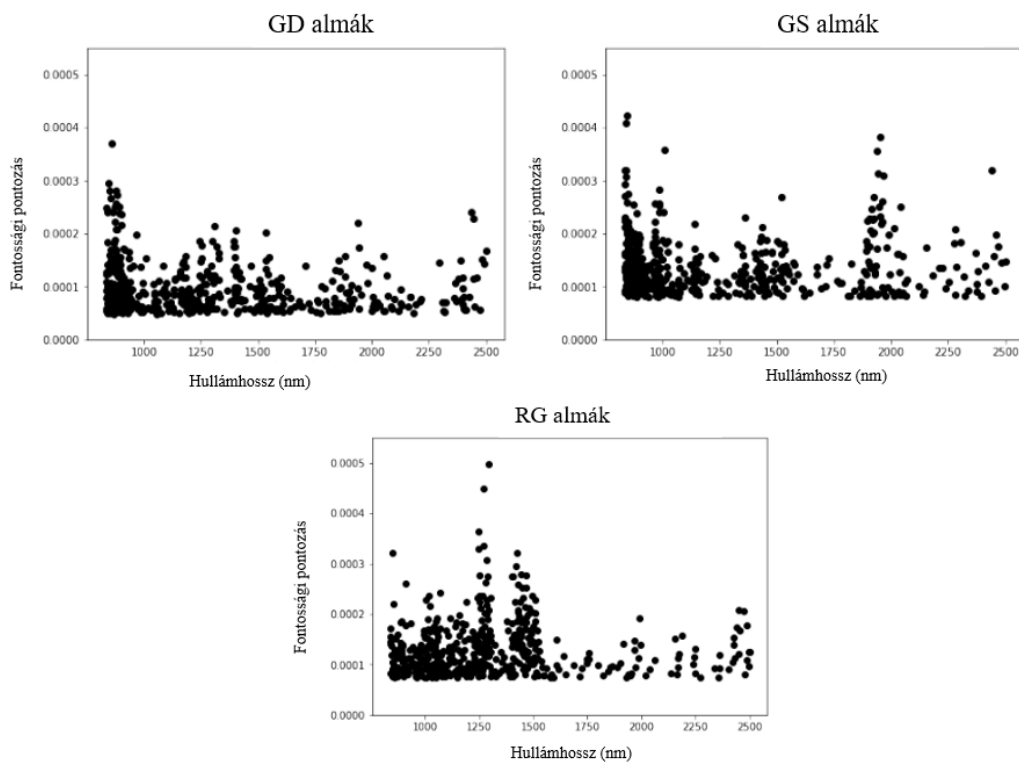
2.5.1. Bevezetés

A kutatás során a legfőbb célunk a tulajdonságok számának redukálása, ezáltal egyszerűsítve az osztályozási műveletet. E szakasz során ismertetjük az optimális tulajdonságok kiválasztásához használt munkafolyamatokat és a sérülések osztályozásához használt gépi tanulási módszereket. A kód és az adatok az eredményekkel együtt elérhetőek a Zenodo [27] weboldalon.

2.5.2. A kiindulási módszer

A logisztikus regressziót (LR) mind a 2000 hullámhosszra létrehoztuk, mint alap előrejelző tényezőt, amellyel az összes többi előrejelző tényezőt össze lehet hasonlítani. Előzetes tesztként a `scikit_learn` `permutation_importance` függvényének segítségével megmértük a permutációs jellemző fontosságát az összes hullámhosszon. A permutációs jellemző fontossági teszt a pontosság relatív csökkenését méri, amikor minden egyes jellemzőt összekeverünk, és így megbontjuk az adott jellemző és a minták közötti korrelációt [26,28]. Ennek a módszernek az

az előnye, hogy gyorsan végrehajtható ("ami nélkülözhetetlen, hiszen több mint 2000 jellemzőt vizsgálunk"). A 3. ábra az első 500 tulajdonság relatív fontosságát mutatja. Az ábrán látható, hogy fontosabb hullámhosszú klaszterek állnak rendelkezésre a jellemzők vonatkozásában. Ezek a hullámhosszok azonban redundáns információt nyújtanak, mivel az intenzitásértékek csak kis mértékben változnak a két egymás melletti hullámhossz között. Így arra számítunk, hogy számos hullámhosszt ki lehet zárni, mint jellemzőt az előrejelzési pontosság jelentős csökkentése nélkül.



3. ábra: A grafikon az 500 legfontosabb jellemző/hullámhossz permutációs fontosságát mutatja

2.5.3. Tulajdonságok kiválasztása

A permutációs módszer egyértelműen rámutat arra, hogy egyes hullámhosszok jelentősebbek a többihez képest. Ugyanakkor a permutációs módszer önmagában nem alkalmas a tulajdonságok kiválasztására, mivel nem veszi figyelembe, hogy két különböző, viszonylag fontos tulajdonság nagymértékben korrelálhat egymással, és hasonló információhoz vezethetnek.

A korrelált jellemzők közötti kiválasztás kezelésére két széles körben használt módszer létezik, amelyeket a következőekben ismertetünk:

- A rekurzív jellemző eltávolítás (RFE) során a modellt először a teljes jellemzőhalmazon betanítjuk, majd töröljük a legkevésbé fontos jellemzőket. Ezt a betanítási-törlési eljárást rekurzív módon megismételjük a fennmaradó jellemzőkkel, amíg a kívánt számú jellemző meg nem marad.
- A szekvenciális jellemzőválasztás (SFS) során eleinte minden egyes jellemzőn alapuló modelleket számolunk ki, és kiválasztjuk a legmegfelelőbb keresztellenőrzési pontszámmal rendelkező jellemzőt. Ezután a kiválasztott jellemzőkből és egy további jellemzőből álló összes modellt kiszámítjuk, majd kiválasztjuk a legjobbat. Ezt a folyamatot rekurzívan meg kell ismételni, minden alkalommal egy jellemzőt hozzáadva. Ehhez a lépéshez az alapértelmezett keresztellenőrzési pontszámot használjuk. A módszernek létezik egy visszafelé történő kiválasztási módszere, amely hasonló az RFS-hez, de úgy találtuk, hogy a kivitelezés túl hosszú ideig tartana.

2.5.4. Gépi tanulási osztályozók

Matematikai modelleket széles körben használtak már a gyümölcsökön végzett, érintésmentes FT-NIR spektrális adatok kemometriai vizsgálatára. A gépi tanulási modelleket egyre gyakrabban használják erre a célra [6,29], mivel rugalmasak és sokféle felhasználási területre adaptálhatóak. A bináris osztályozáshoz összesen hat gépi tanulási (ML) eszközt használunk: logisztikus regresszió (LR), szupport vektor gép (SVM), véletlen erdők (RF), extreme gradient boosting (XGB), K-nearest Neighbour (Knn) és mesterséges neurális hálózat (ANN). Ezek a módszerek a szakirodalomban viszonylag kis adathalmazokra a legkorszerűbbek [30,31], mint ahogy ezt mutatja az 1. táblázat is.

1. táblázat: A minták számának almafajtánkénti megoszlása

Típus	Nem sérült minta	Sérült minta	Összes
GD	274	273	547
GS	252	251	503
RG	278	284	562

Minden módszert 3-szoros keresztellenőrzéssel optimalizáltunk, hogy elkerüljük a túlzott illeszkedést, és megbizonyosodjunk arról, hogy a képzett modell képes általánosítani a nem ismert adatokra.

Az ML paraméterek optimalizálása minden tulajdonsághalmaz esetében a Python könyvtárban implementált scikit-learn [26] véletlenszerű keresésével valósult meg. Az SVM és az alapmodell kivételével minden modell 30 véletlenszerű lekérdezésen ment keresztül, az SVM viszont csak 15-ön, mivel az SVM rendkívül számításigényes.

2.5.5. Az osztályozó teljesítmény összehasonlítása

A kutatásban az összes ML-módszer osztályozási pontosságát hasonlítottuk össze az összes jellemzőre vonatkozóan, amely a következőképpen került kiszámításra

$$\text{Precizitás} = \frac{TP}{TP+FP}$$

ahol a TP a valódi pozitív (TP), az FP a tévesen pozitív (FP) előrejelzések száma. Itt a "pozitív" eredmény az alma jó minőségére utal: tehát az FP eredmény tévesen jónak minősíti a sérült almát. A precizitást azért használjuk, mert a legnagyobb költséggel járó hiba a sérült almák téves azonosítása.

A modellek közötti hatékony összehasonlíthatóság érdekében az osztályozás pontossági különbségeire vonatkozó hibasávokat (amelyek két standard eltérésnek felelnek meg) is kiszámítottuk. Ha a 0 a hibasávokon kívül esett, akkor arra a megállapításra jutottunk, hogy a két modell közötti eltérés statisztikailag szignifikáns. Ellenkező esetben nem tudtuk volna elvetni azt a nullhipotézist, hogy nincs különbség a modellek között.

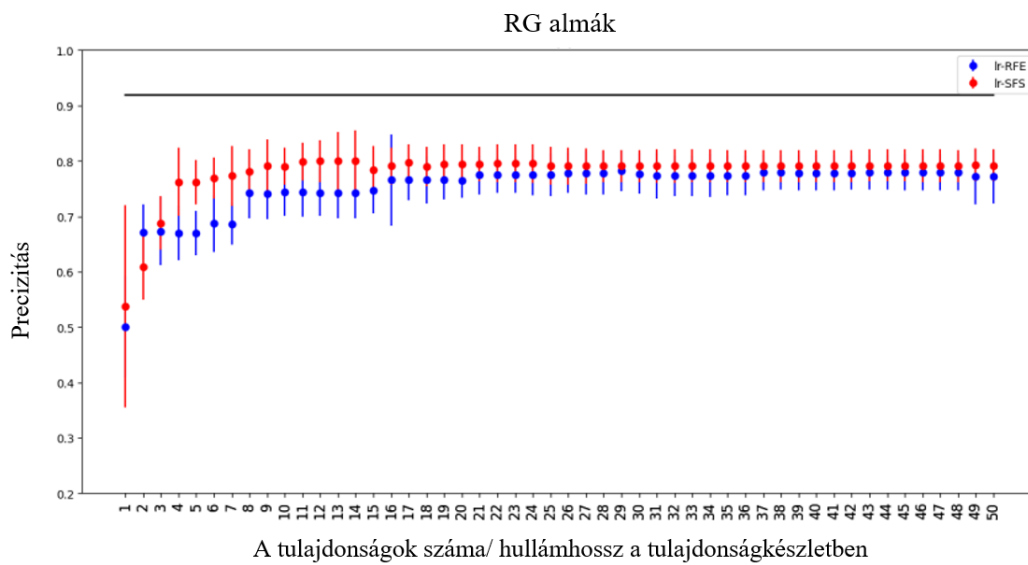
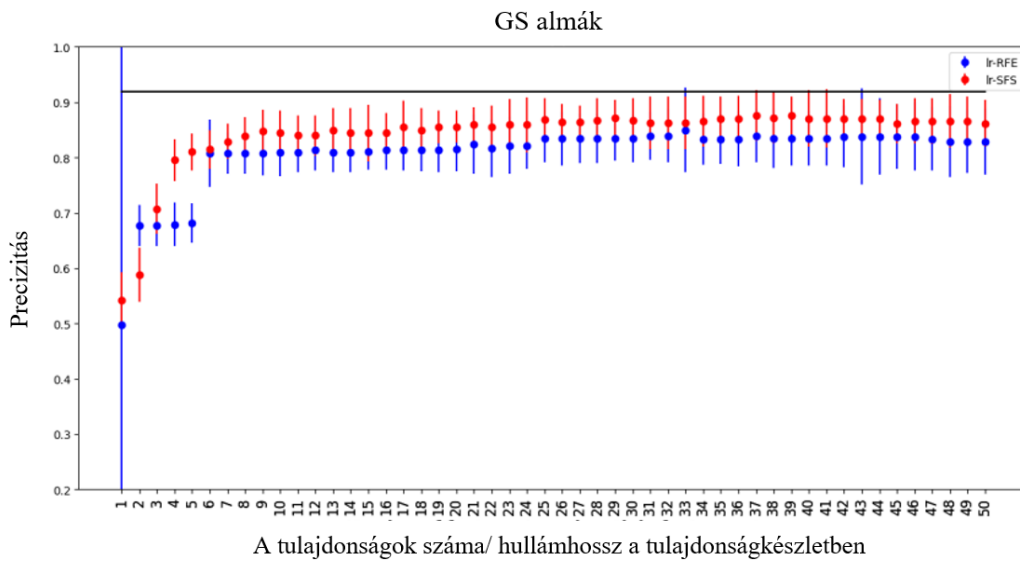
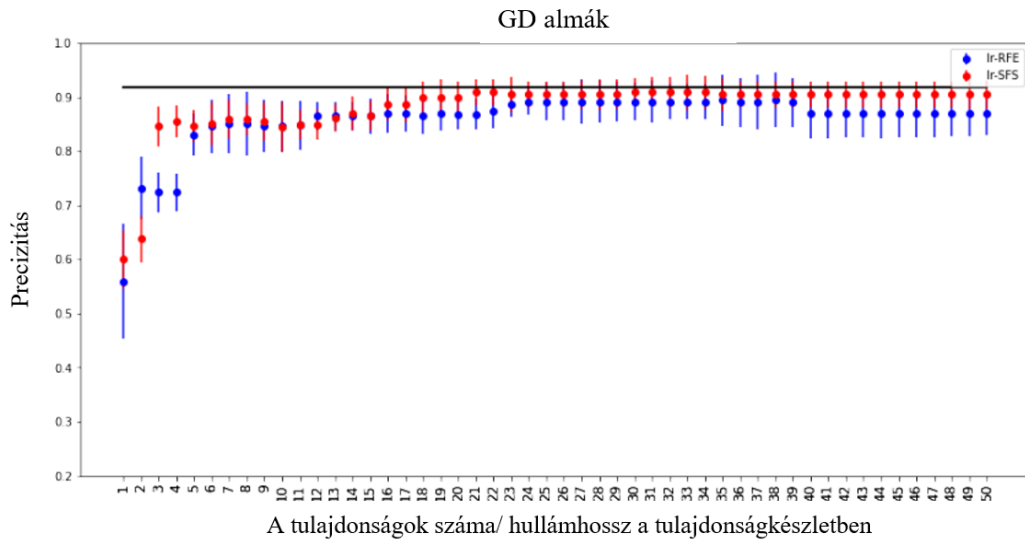
A különböző besorolások pontosságát a tesztelési halmaz segítségével értékeltük. A hibasávok kiszámításához „jackknife” módszert alkalmaztunk „leave-out-one” keresztvalidációval [32,33]. A tanulóhalmaz hibája esetében a kihagyott esetek a tanulóhalmaz egyharmadát tették ki, míg a tesztkészlet esetében az összes esetet egyenként hagytuk ki, hogy megkapjuk a standard eltérés „jackknife” becslését.

3. Eredmények

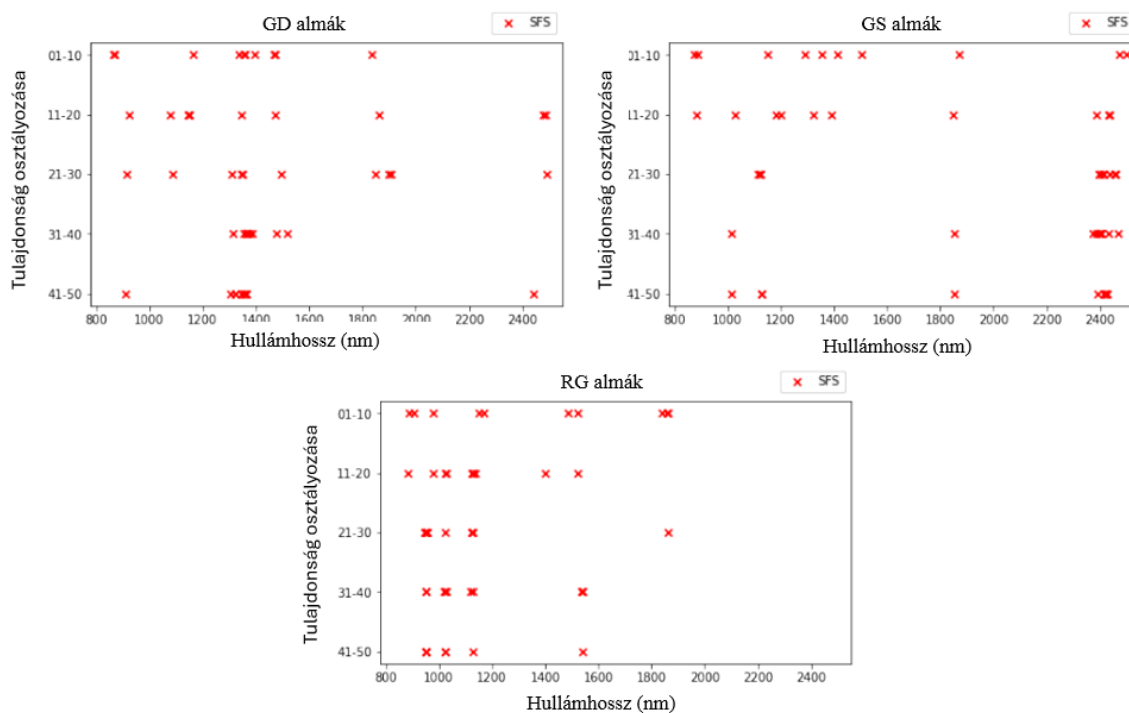
A kutatómunka során mind az RFE, mind az SFS módszereket LR-rel együtt használtuk, hogy a legjobb 50 jellemzőt megkapjuk, ami 50×2 osztályozási paramétert eredményezett, ami megfelel a legjobb jellemző, a legjobb 2 jellemző és a legjobb 50 jellemző figyelembevételének. Az 50 tulajdonság kiválasztása a permutációs jellemző fontosságával végzett előzetes vizsgálatokon alapult, amelyek azt mutatták, hogy az 50-nél több jellemző használata nem jár a pontosság javulásával. A 4. ábra a legjobb 50 tulajdonságkészlet pontosságát mutatja az LR használatával, a tanulási adatok alapján. Az ábrán jól látható, hogy az SFS mindhárom alma

esetében jobb becsléseket ad az összes tulajdonságkészlet esetében. Az eredmények tesztelésénél ezért csak az SFS-t alkalmaztuk, és nem vettük figyelembe az RFE-t. Meg kell jegyezni, hogy a 4. ábrán látható tulajdonságválasztás az összes, több mint 2000 tulajdonságot felhasználta, ami rendkívül számításigényes művelet volt. Ezt követően megállapítottuk, hogy gyakorlatilag megegyező eredményeket lehet elérni sokkal rövidebb idő alatt, ha csak a permutációs jellemző fontosságából kapott legjobb 200 jellemzővel végezzük el a tulajdonságválasztást (lásd a 3. ábrát).

Az 5. ábra grafikusán mutatja a legjobb 10, 20, 30, 40 és 50 tulajdonság hullámhosszát mindhárom fajta esetében: a pontos hullámhosszértékek az "A" függelékben szerepelnek. Az egyes ábrákon az x tengely a hullámhosszt, az y tengely pedig a tulajdonságok 10-es csoportokban való rangsorolását ábrázolja. Az ábrán az látható, hogy a 10 legfontosabb tulajdonság valamelyest konzisztens mindhárom fajta esetében: a 900, 1300, 1500 és 1900 nm közeli hullámhosszok között szerepelnek. Mindemellett a három fajta legfontosabb tulajdonságai között lényeges különbségek mutatkoznak. Például csak a GS fajtánál vannak 2400 feletti hullámhosszok a 10 legjobb érték között. Figyelembe kell venni, hogy ipari körülmények között a fajtákat külön válogatnák, ezért a gyakorlatban a tulajdonságok kiválasztásának fajtaspecifikus megközelítése a célravezető. Megfigyelhető továbbá, hogy a tulajdonságok fontosságának csökkenésével a jellemzők egyre inkább csoportosulnak. A GD almák esetében például a 41-50 fontosságú jellemzők három klasztert alkotnak, míg a 0-30 fontosságú jellemzők sokkal szerte ágazóbbak.



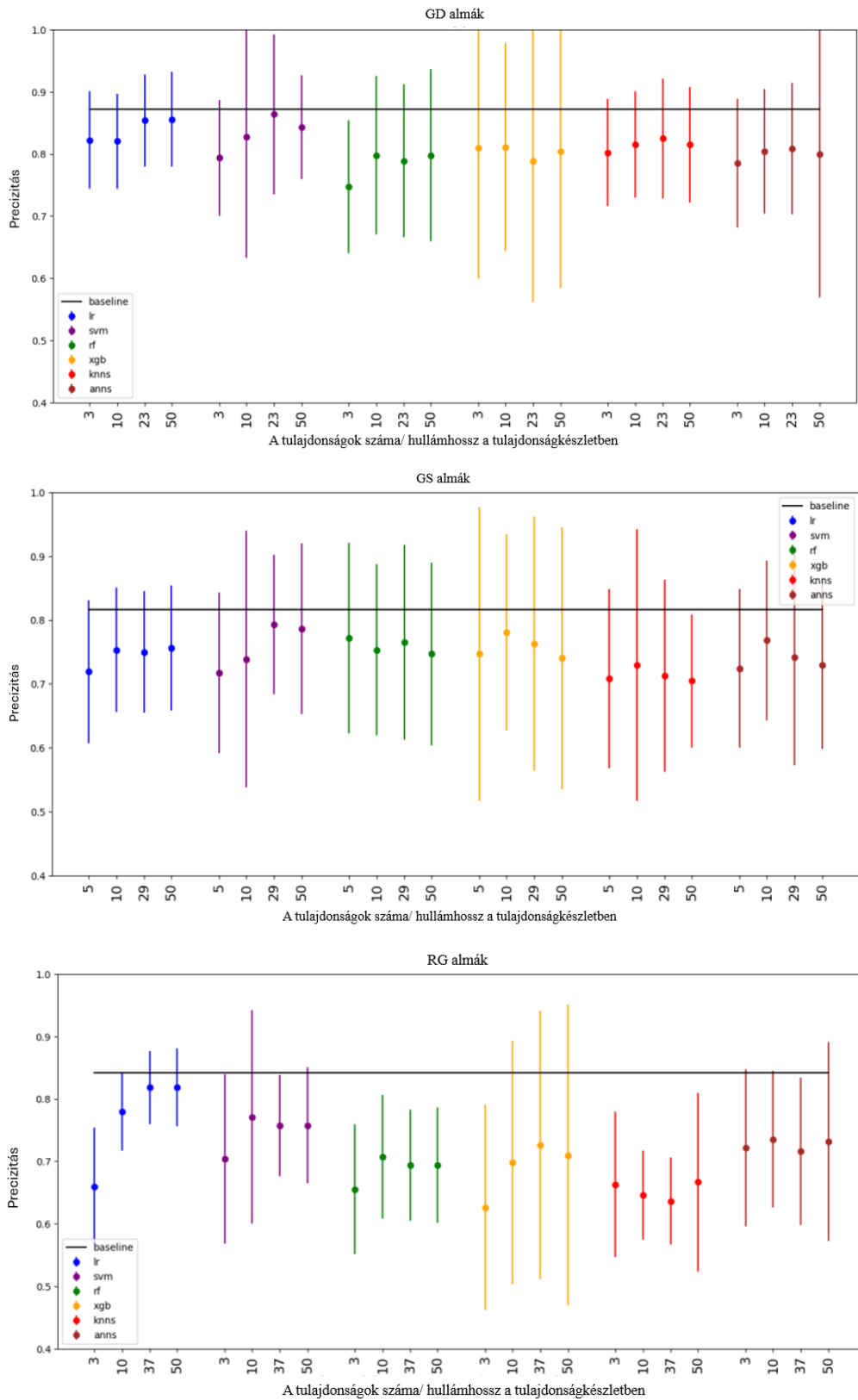
4. ábra: A legjobb 50 tulajdonságkészlet RFE és SFS használatával a 3 adathalmazon LR használatával



5. ábra: A három fajta esetében a sérülések szegregációja szempontjából legfontosabb 50 kiválasztott tulajdonság eloszlása. A kiválasztott hullámhosszok 10-es csoportokban vannak ábrázolva, az elsőként kiválasztott van legfelül. A részletes gyakorisági információk az A. függelékben találhatóak

Az 50 különböző tulajdonságkészletből négyet választottunk ki a teszteléshez a vizsgálati halmazban. A négy tulajdonságkészletet a következőképpen határoztuk meg: (a) A tulajdonságok az első pontossági ugrásig bezárólag; (b) A legjobb 10 tulajdonság; (c) A modell stabilizálódásakor; (d) A legjobb 50 tulajdonság. A 6. ábra a négy különböző jellemzőkészleten alapuló modellek pontosságát hasonlítja össze. Minden egyes tulajdonságkészlethez a 2.5.4. szakaszban felsorolt összes ML-eszközt alkalmaztuk. Összességében a precizitás 0,7 és 0,9 között mozgott, a legjobb eredményt jellemzően a GD adta. A b)-d) pontokban szereplő tulajdonságkészletek minden ML-módszer esetében közel azonos eredményt hoztak, az LR és az SVM esetében hasonló átlagos precizitást eredményeztek, viszont az LR-nek jelentősen kisebbek voltak a hibásávjai. A legmagasabb átlagos precizitások a különböző almafajtákra átlagosan 0,86, (GD); 0,79 (GS) és 0,81 (RG) voltak. Összességében a legjobb eredményeket a (c) tulajdonságkészlet esetében kaptuk. A vízszintes vonal a 6. ábrán az alapvonalai becsléssel kapott precizitást mutatja, amelyet az LR minden hullámhosszt használva ért el. Az alapvonalhoz képest a legjobban teljesítő becslések átlagpontossága kissé alulmaradt, azonban a különbségek kismértékűek (0,01-0,03 közöttiek) és statisztikailag elhanyagolhatóak. A

csökkentett tulajdonságkészletek ugyan nem adtak jobb precizitást, mint az alapvonal, azonban sokkal kevesebb hullámhosszmérést követelnek meg, így a gyakorlatban lényegesen könnyebben megvalósíthatóak.



6. ábra: A kiválasztott jellemzők összes részhalmazát használó összes modell precizitásának ábrázolása a tesztadatokra, összehasonlítva az alapmodellel (vízszintes vonal). A hibásávokat a függőleges vonalak jelölik

4. Vita

A tanulmányban kidolgozott, kiválasztott tulajdonságok részalmazain alapuló modellek osztályozási precizitása nem haladta meg az alapmodellét, amely az összes hullámhosszt alkalmazta. Hasonló viszonylagos eredményről számoltak be Luo és munkatársai [16] egy négy almafajta sérüléseinek osztályozására vonatkozó kutatásban is [16]; a szerzők egy „ROC-AUC” alapú módszert alkalmaztak a spektrális képalkotáshoz szükséges effektív hullámhosszok kiválasztására, és megállapították, hogy a szűkített modellek nem haladták meg a teljes spektrumon alapuló modell érzékenységi, specificitási és pontossági értékeit. Fontos megemlíteni, hogy bár a bináris osztályozásunk csak a sérülések meglétére vagy hiányára összpontosít, a mintavételi eljárásunk a sérülések különböző szintjeinek, a minta eredetének és a mintavételezési időpontoknak a figyelembevételével magas variabilitást hozott a mintatartományba. Az egyes modellek számára kihívást jelent a minták nagyfokú változékonyságának teljes mértékű figyelembevétele. Ugyanakkor jó általánosítási képességgel rendelkező modellek készítése mindenképpen szükséges. Emellett a spektrális felvétel reflektív módszere nemlineáris kapcsolatokat eredményezhet a kémiai összetétel és a spektrumok között [34], ellentétben a Beer-Lambert-törvény által leírt lineáris kapcsolatokkal.

Az összes fajta esetében a körülbelül 40 hullámhosszon alapuló LR és SVM modellek mutatták a legmagasabb átlagos precizitási értékeket (0,79; 0,86). Ezek az eredmények összehasonlíthatóak a spektroszkópia-alapú sérülésosztályozásban használt osztályozási teljesítménymérők általános tartományával. Mindazonáltal léteznek olyan tanulmányok [35], amelyek szerint az alma sérüléseinek osztályozására szolgáló modellek jobb eredményt értek el a különböző abszorpciós hullámhosszú tartományok alkalmazásával. Elképzelhető, hogy nagyobb hatékonyság érhető el a mélytanulás alkalmazásával, amelyet más agrártermékek osztályozási esetekben már sikeresen alkalmaztak [36]. A mélytanulós osztályozók hatékony képzéséhez azonban jellemzően lényegesen szélesebb adatállományra van szükség, mint ami ehhez a tanulmányhoz rendelkezésre állt [32,37]. Ráadásul számos, gépi tanulást (beleértve a mély tanulást is) alkalmazó kutatás nem tartalmaz hibásavokat a teljesítményértékelésekben, így kétséges, hogy a leírt osztályozási fejlesztések statisztikailag szignifikánsak-e [38].

A hatékony alkalmazásokat illetően fontos szempont a gépi tanulás általánosítási képessége. Éppen ezért az összes modellt az eredeti adathalmazból korábban elkülönített, nem ismert adatokkal teszteltük. Mindazonáltal, mivel az adatgyűjtés ellenőrzött módon történt, meg kell jegyezni, hogy olyan tényezők, mint például a hőmérséklet, befolyásolhatják a NIR-spektrumok

felvételét. Ezenkívül a különböző éghajlati régiókból és termesztési körülményekből származó gyümölcsök figyelembevételével variabilitás is bevezethető. A modell általánosítási képességének javítása céljából ezeket a szempontokat a jövőbeli vizsgálatokban figyelembe lehetne venni.

Korábban már vizsgálták az alma sérüléseinek a NIR-hullámsávok jellemzőit. Geola és munkatársai [39] javaslatot tettek a "Golden Delicious" fajtájú almák sérült szövetének kimutatására, és 750-800 nm-es tartományban osztályozó függvényt határoztak meg. Más almaszövet-osztályozási vizsgálatok reflexiós spektrumokat alkalmaztak, és az esetek többségében a legmeghatározóbb hullámhosszakat a 690-850 nm-es tartományban találták [40-42]. Kleynen és munkatársai [43] VIS-NIR spektrométerrel dolgozva módszert dolgoztak ki a bicolor alma ("Jonagold") terméshibáinak kimutatására. A legfontosabb hullámhosszakat a NIR-tartományban találták (700-920 nm vagy $14\ 285-10\ 869\ \text{cm}^{-1}$). Lammertyn és munkatársai szerint a Jonagold almaszövetben a NIR-sugárzás fénybehatolási mélysége a 700-900 nm-es tartományban a legnagyobb [44]. Kleyman ezért megfigyeléseiket az utóbbi hullámhosszú tartományhoz társította. Annak ellenére, hogy ezek a vizsgálatok rendszerint a rövidhullámú NIR-sávokra épültek, a 780 nm-től 5000 nm-ig terjedő széles spektrumtartomány figyelembevétele javíthatja a különböző mélységű zúzódások felismerését [10].

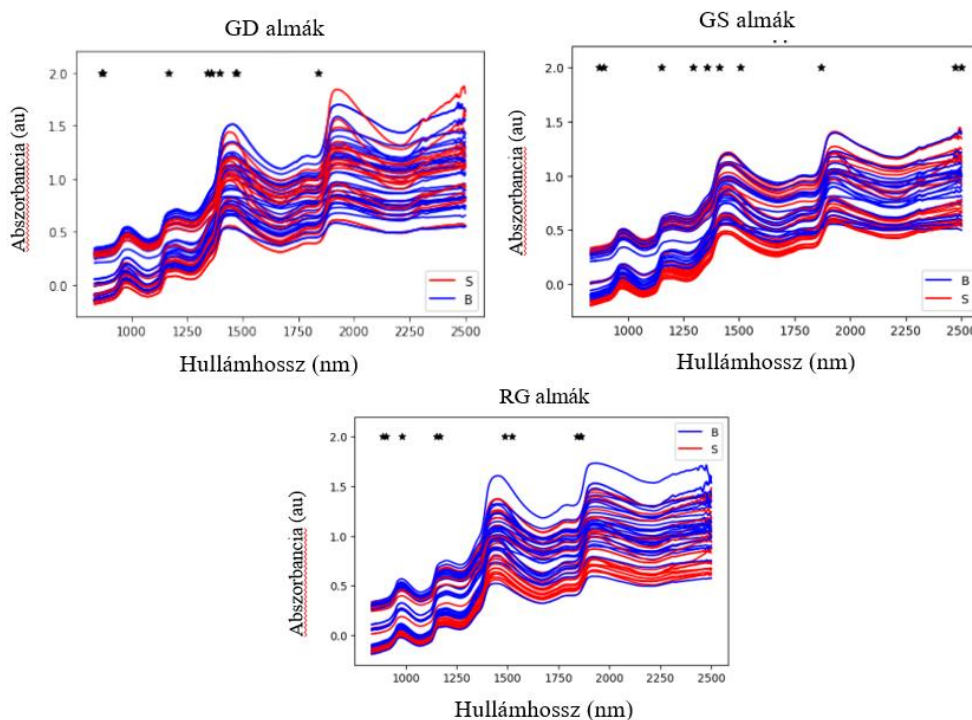
Az almahéjának kémiai összetétele nagyon összetett, és a biológiai funkciójának megfelelően alkalmazkodott. A héj külső felületét zsíralkoholokból, zsírsavakból és hosszú szénhidrogénekből álló kutikula viaszok alkotják [45-47]. Továbbá különböző vegyületek különböző csoportjai, például flavonoidok [48,49], fenolsavak [50,51], triterpén-észterek [52] és proantocianidinek [53] találhatóak benne. A héj színkémiája a vizsgált almafajtától függően változik; az alma színét adó vegyületek fő csoportjába a klorofillok (zöld) [54], a karotinoidok (sárga) [55] és az antocianinok (piros) [56] tartoznak.

A NIR spektrum számos felhanggal rendelkezik a különböző típusú kémiai kötések elnyelése miatt, beleértve a C-H (metil-, metilén-, metoxi-, aromás- és karbonilcsoportok), a N-H (primer és szekunder amidok, primer, szekunder és terciar aminok és aminsók), a O-H (alkoholok és víz), a S-H és a C=O csoportokat [57].

Esetünkben nehéz meghatározni bármelyik funkciós csoporthoz tartozó egyedi jeleket; az 1. ábrán több széles sáv azonosítható, amelyek mindhárom almafajtára jellemzőek. Az első tartomány 850-1100 nm között mozog, és két alsávot foglal magába: 850-920 nm (C-H metil/metilén, amely alifás és/vagy aromás vázakhöz kapcsolódik, és 950-1000 nm (OH és NH

alifás/aromás csoportok). A második, 1150-1300 nm közötti sáv a C=O, az aromás C-H és a C=C csoportoknak köszönhető. A harmadik tartomány 1350-1600 nm között mozog, és OH, szabad és hidrogénkötésű és C-H alifás és aromás, N-H aromás és alifás, C=O aldehideket és ketonokat takar. Egy kevésbé feltűnő tartomány 1800-1850 nm között mozog, és a szénhidrátpolimer OH-ját, CH-ját fedi le. Végül egy 1900-2000 nm közötti tartományban található sáv a C=O amidokat/karboxilcsoportokat és a többnyire aromás OH/NH vegyületeket fedi le.

A 7. ábra azt mutatja, hogy a sérülések azonosítása szempontjából legfontosabb jellemzők hogyan viszonyulnak a különböző almatípusok NIR-spektrumához. A spektrum minden egyes sávjához minden esetben a 10 legjobb jellemző tartozik. A kiválasztott jellemzők azonban inkább a sáv meredeken emelkedő széléhez közel helyezkednek el, és nem a sáv tartományának közepén. A 10 legfontosabb jellemzővel társított sáv az 1350-1600-as sáv, amely a fentiek szerint többféle kémiai kötésből származó abszorpciót tartalmaz. A top 10 jellemző/hullámhossz pontos értékeit az egyes almatípusok esetében az „A” függelék tartalmazza.



7. ábra: A 10 legfontosabb jellemző/hullámhosszúság (fekete csillagokkal jelölve) helyzete a három különböző almatípus NIR-spektrumával összehasonlítva

Úgy tűnik, hogy a spektroszkópia alkalmazása egyre kisebb érdeklődést mutat a sérülések kimutatására a képalkotó lehetőségek javára. Ez összefügghet a spektroszkópai eszközök

térbeli korlátaival, hiszen a spektroszkópai eszközök csak korlátozott térbeli reprezentációt (legfeljebb 22 mm-es foltátmérő) biztosítanak az érintett mintáról. Ez a korlát a különböző típusú friss termények gyors válogatására és osztályozására nem elegendő. Az agrár-élelmiszeripari friss termékek roncsolásmentes hibazonosításának legújabb trendjei a képalkotó technikákra [58], köztük a hiperspektrális képalkotásra összpontosítanak. Azonban ezek a képalkotó technikák nagyban támaszkodnak a képfeldolgozásra a jellemzők kinyeréséhez, ami nagy számítási ráfordítást igényel. A mélytanulás egyre elterjedtebb alkalmazása az agrártermékek minőségének spektroszkópai és képalkotáson alapuló értékelésére gyors végpontok közötti modellezési folyamatot kínál [36], azonban a képek rögzítése még mindig időigényes és nem alkalmas az ipari osztályozási sebesség elérésére. Ennélfogva az ilyen alkalmazásokhoz a többspektrális képalkotást részesítik előnyben, mivel az néhány előre meghatározott hullámhosszon alapul, amelyek hatékonyak egy adott tulajdonság értékeléséhez. Ebben a munkában bemutatásra kerül a hullámhossz kiválasztásának megvalósíthatósága a teljes mértékben exponált almák spektrális adatai alapján, a rugalmas, nyílt eszközökkel működő szoftverek segítségével. Ez a fajta exponálás olyan mintakínálatot modellez, amely az inline válogatási alkalmazásokhoz alkalmazható lenne.

5. Következtetések

A nyílt forráskódú fejlesztő eszközök és egy érintésmentes, nagy (100 mm átmérőjű) mintát exponáló NIR-spektrométer kombinációja által kínált lehetőségekre építve megvalósíthatósági tanulmányt végeztünk az alma sérüléseinek osztályozására, azzal a céllal, hogy három almafajtában megállapítsuk az információval szolgáló hullámhosszokat. A sérülések elkülönítésére hat gépi tanulási osztályozó algoritmus segítségével modelleket hoztunk létre, amelyekhez rekurzív jellemzőeliminációs és szekvenciális jellemzőválasztási módszereket is alkalmaztunk a leginkább informatív hullámhosszok meghatározása érdekében. A teljes spektrumon alapuló modellekkel hibahatáron belül megegyező osztályozási pontosságot lehetett elérni a 4 fő hullámsávból kiválasztott legfeljebb 50 hullámhossz használatával. A legjobban lecsökkentett osztályozási modellek az LR és SVM gépi tanulási technikákon alapultak, amelyek a fajtától függően 0,7 és 0,9 közötti pontossági értékeket eredményeztek.

„A” függelék

A melléklet a legjobb 50 tulajdonságot tartalmazza a szekvenciális tulajdonságválasztási modell segítségével.

A1. táblázat: A sérült/zúzott alma azonosítására alkalmas tulajdonságok/hullámhosszok prioritási sorrendben, különböző almatípusok esetében.

Prioritás	GD hullámhossz (nm)	GS hullámhossz (nm)	RG hullámhossz (nm)
1	1393,9	2473,9	1840,1
2	864,5	1356,7	1486,6
3	1357,4	872,4	904,6
4	1468,1	1411,3	1168,9
5	1836,1	1292,4	888,5
6	1472,3	1150,2	980,6
7	868,8	1505,6	1152,3
8	1163,1	889,7	1522,4
9	1338,5	1870,6	1857,2
10	1354,6	2500,1	1861,2
11	1144,1	1180,6	976,5
12	1145,2	883,1	882,5
13	1862,5	1201,4	1136,1
14	2488,1	2432,1	1523,3
15	1151,3	2436,7	1399,2
16	1076,2	1391,6	1123,8
17	2478,6	1029,6	1022,3
18	924,0	2387,3	1028,0
19	1346,1	1321,4	1126,7
20	1473,9	1846,6	1127,2
21	1496,9	2455,2	949,0
22	1343,3	2439,0	953,9
2	915,2	2409,5	1862,5
24	2492,9	2396,2	1024,0
25	1309,4	1124,3	1125,3
30	1086,6	1115,6	1127,7
31	1384,2	2372,0	1125,8
32	1376,9	2389,5	1537,7
33	1312,7	2398,4	1026,8
34	1477,3	1015,5	1541,4
35	1366,7	1851,9	1120,9
36	1518,8	2402,8	1019,5
37	1367,4	2407,3	953,5
38	1361,7	2405,0	952,8
39	1363,1	2434,4	1538,7
40	1356,7	2469,2	1025,6
41	1324,1	1128,2	949,7
42	1344,7	1853,2	952,1

43	1363,8	1126,7	1126,3
44	2441,3	1016,3	1539,6
45	1352,4	2425,3	951,4
46	1320,8	2416,3	950,0
47	1353,2	2391,7	951,1
48	1364,5	2418,5	952,5
49	1306,8	2429,8	1025,2
50	911,3	2423,0	1023,1

Hivatkozások

1. Hussein, Z.; Fawole, O.A.; Opara, U.L. Harvest and Postharvest Factors Affecting Bruise Damage of Fresh Fruits. *Hortic. Plant J.* 2020, 6, 1–13.
2. Opara, U.L.; Pathare, P.B. Bruise damage measurement and analysis of fresh horticultural produce—A review. *Postharvest Biol. Technol.* 2014, 91, 9–24.
3. Al-Dairi, M.; Pathare, P.B.; Al-Yahyai, R.; Opara, U.L. Mechanical damage of fresh produce in postharvest transportation: Current status and future prospects. *Trends Food Sci. Technol.* 2022, 124, 195–207.
4. Hussein, Z.; Fawole, O.A.; Opara, U.L. Preharvest factors influencing bruise damage of fresh fruits—A review. *Sci. Hortic.* 2018, 229, 45–58.
5. Li, Z.; Thomas, C. Quantitative evaluation of mechanical damage to fresh fruits. *Trends Food Sci. Technol.* 2014, 35, 138–150.
6. Nturambirwe, J.F.; Nieuwoudt, H.H.; Perold, W.J.; Opara, U.L. Detecting bruise damage and level of severity in apples using a contactless nir spectrometer. *Appl. Eng. Agric.* 2020, 36, 257–270.
7. Nturambirwe, J.F.I.; Perold, W.J.; Opara, U.L. Classification learning of latent bruise damage to apples using shortwave infrared hyperspectral imaging. *Sensors* 2021, 21, 4990.
8. Zhou, Y.; Wu, D.; Hui, G.; Mao, J.; Liu, T.; Zhou, W.; Zhao, Y.; Chen, Z.; Chen, F. Loquat Bruise Detection Using Optical Coherence Tomography Based on Microstructural Parameters. *Food Anal. Methods* 2018, 11, 2692–2698.
9. Lu, Y.; Lu, R. Non-destructive defect detection of apples by spectroscopic and imaging technologies: A review. *Trans. ASABE* 2017, 60, 1765–1790.
10. Baranowski, P.; Mazurek, W.; Wozniak, J.; Majewska, U. Detection of early bruises in apples using hyperspectral data and thermal imaging. *J. Food Eng.* 2012, 110, 345–355.
11. Huang, W.; Li, J.; Wang, Q.; Chen, L. Development of a multispectral imaging system for online detection of bruises on apples. *J. Food Eng.* 2015, 146, 62–71.

12. Zhang, B.; Gu, B.; Tian, G.; Zhou, J.; Huang, J.; Xiong, Y. Challenges and solutions of optical-based nondestructive quality inspection for robotic fruit and vegetable grading systems: A technical review. *Trends Food Sci. Technol.* 2018, 81, 213–231.
13. Xiaobo, Z.; Jiewen, Z.; Povey, M.J.; Holmes, M.; Hanpin, M. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* 2010, 667, 14–32.
14. Mehmood, T.; Liland, K.H.; Snipen, L.; Sæbø, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* 2012, 118, 62–69.
15. Nturambirwe, J.F.I.; Opara, U.L. Machine learning applications to non-destructive defect detection in horticultural products. *Biosyst. Eng.* 2020, 189, 60–83.
16. Luo, X.; Takahashi, T.; Kyo, K.; Zhang, S. Wavelength selection in vis/NIR spectra for detection of bruises on apples by ROC analysis. *J. Food Eng.* 2012, 109, 457–466.
17. Nturambirwe, J.F.I.; Nieuwoudt, H.H.; Opara, U.L.; Perold, W.J. Performance of genetic algorithm in optimization of NIRS PLS models to predict apple fruit quality. *Acta Hort.* 2018, 1201, 355–362.
18. Li, L.; Jang, X.; Li, B.; Liu, Y. Wavelength selection method for near-infrared spectroscopy based on standard-sample calibration transfer of mango and apple. *Comput. Electron. Agric.* 2021, 190, 106448.
19. Abasi, S.; Minaei, S.; Jamshidi, B.; Fathi, D. Dedicated non-destructive devices for food quality measurement: A review. *Trends Food Sci. Technol.* 2018, 78, 197–205.
20. Amuah, C.L.; Teye, E.; Lamptey, F.P.; Nyandey, K.; Opoku-Ansah, J.; Adueming, P.O.W. Feasibility Study of the Use of Handheld NIR Spectrometer for Simultaneous Authentication and Quantification of Quality Parameters in Intact Pineapple Fruits. *J. Spectrosc.* 2019, 2019, 5975461.
21. Srivastava, S.; Shashikant, B.V. Handheld, smartphone based spectrometer for rapid and nondestructive testing of citrus cultivars. *J. Food Meas. Charact.* 2021, 15, 892–904.
22. Opara, L.U.; AlGhafri, A.; Agzoun, H.; AlIssai, J.; AlJabri, F. Design and development of a new device for measuring susceptibility to impact damage of fresh produce. *New Zealand J. Crop. Hortic. Sci.* 2007, 35, 245–251.

23. Magwaza, L.S.; Opara, U.L.; Terry, L.A.; Landahl, S.; Cronje, P.J.; Nieuwoudt, H.; Mouazen, A.M.; Saeys, W.; Nicolai, B.M. Prediction of 'Nules Clementine' mandarin susceptibility to rind breakdown disorder using Vis/NIR spectroscopy. *Postharvest Biol. Technol.* 2012, 74, 1–10.
24. Bruker. Matrix-F FT-NIR Spectrometer. Available online: <https://www.bruker.com/en/products-and-solutions/infrared-andraman/ft-nir-spectrometers/matrix-f-ft-nir-spectrometer.html> (accessed on 15 December 2022).
25. Ning, J.; Ye, H.; Sun, Y.; Zhang, J.; Mei, Z.; Xiong, S.; Zhang, S.; Li, Y.; Hui, G.; Yi, X.; et al. Study on apple damage detecting method based on relaxation single-wavelength laser and convolutional neural network. *J. Food Meas. Charact.* 2022, 16, 3321–3330.
26. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API design for machine learning software: Experiences from the scikit-learn project. *arXiv* 2013, arXiv:1309.0238.
27. Hussein, E. Apple Classification, first release. Zenodo v1.0 2022. Available online: <https://zenodo.org/record/7007001/export/dcite4#.Y7PBHRVBxPY> (accessed on 15 March 2012).
28. Altmann, A.; Tolo,si, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* 2010, 26, 1340–1347.
29. Nturambirwe, J.F.I.; Nieuwoudt, H.H.; Perold, W.J.; Opara, U.L. Non-destructive measurement of internal quality of apple fruit by a contactless NIR spectrometer with genetic algorithm model optimization. *Sci. Afr.* 2019, 3, e00051.
30. Mendez, K.M.; Reinke, S.N.; Broadhurst, D.I. A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification. *Metabolomics* 2019, 15, 1–15.
31. Aguasca-Colomo, R.; Castellanos-Nieves, D.; Méndez, M. Comparative analysis of rainfall prediction models using machine learning in islands with complex orography: Tenerife Island. *Appl. Sci.* 2019, 9, 4931.
32. Hussein, E.A.; Thron, C.; Ghaziasgar, M.; Vaccari, M.; Marnewick, J.L.; Hussein, A.A. Comparison of Phenolic Content and Antioxidant Activity for Fermented and Unfermented Rooibos Samples Extracted with Water and Methanol. *Plants* 2022, 11, 16.

33. Hussein, E.A.; Ghaziasgar, M.; Thron, C.; Vaccari, M.; Bagula, A. Basic Statistical Estimation Outperforms Machine Learning in Monthly Prediction of Seasonal Climatic Parameters. *Atmosphere* 2021, 12, 539.
34. Nadler, B.; Coifman, R.R. Partial least squares, Beer's law and the net analyte signal: Statistical modeling and analysis. *J. Chemom.* 2005, 19, 45–54.
35. Siedliska, A.; Baranowski, P.; Mazurek, W. Classification models of bruise and cultivar detection on the basis of hyperspectral imaging data. *Comput. Electron. Agric.* 2014, 106, 66–74.
36. Zhang, X.; Yang, J.; Lin, T.; Ying, Y. Food and agro-product quality evaluation based on spectroscopy and deep learning: A review. *Trends Food Sci. Technol.* 2021, 112, 431–441.
37. Purushotham, S.; Meng, C.; Che, Z.; Liu, Y. Benchmarking deep learning models on large healthcare datasets. *J. Biomed. Inform.* 2018, 83, 112–134.
38. Hussein, E.A.; Ghaziasgar, M.; Thron, C.; Vaccari, M.; Jafta, Y. Rainfall Prediction Using Machine Learning Models: Literature Survey. In *Artificial Intelligence for Data Science in Theory and Practice*; Alloghani, M., Thron, C., Subair, S., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 75–108.
39. Geolla, F.; Geolla, F.; Peiper, U.M. A spectrophotometric method for detecting surface bruises on “Golden Delicious” apples. *J. Agric. Eng. Res.* 1993, 58, 47–51.
40. Upchurch, B.L.; Affeldt, H.A.; Hruschka, W.R.; Norris, K.H.; Throop, J.A. Spectrophotometric study of bruises on whole, ‘Red Delicious’ apples. *Trans. Am. Soc. Agric. Eng.* 1990, 33, 585–589.
41. Miller, W.M.; Throop, J.A.; Upchurch, B.L. Pattern recognition models for spectral reflectance evaluation of apple blemishes. *Postharvest Biol. Technol.* 1998, 14, 11–20.
42. Wen, Z.; Tao, Y. Building a rule-based machine-vision system for defect inspection on apple sorting and packing lines. *Expert Syst. Appl.* 1999, 16, 307–313.
43. Kleynen, O. Selection of the most efficient wavelength bands for ‘Jonagold’ apple sorting. *Postharvest Biol. Technol.* 2003, 30, 221–232.

44. Lammertyn, J.; Peirs, A.; De Baerdemaeker, J.; Nicolai, B. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biol. Technol.* 2000, 18, 121–132.
45. Lu, P.L.; Lin, C.H. Physiology of fruit cracking in wax apple (*Syzygium samarangense*). *Bot. Orient. J. Plant Sci.* 2011, 8, 70–76.
46. Wang, B.H.; Cao, J.J.; Zhang, B.; Chen, H.Q. Structural characterization, physicochemical properties and α -glucosidase inhibitory activity of polysaccharide from the fruits of wax apple. *Carbohydr. Polym.* 2019, 211, 227–236.
47. Liu, Y.P.; Xie, Z.; Guan, R.Q.; Du, M.R.; Qiao, Z.H.; Suo, X.Y.; Liu, Z.Y.; Bian, Y.; Qiang, L.; Fu, Y.H. Syzygiamalactone, an Unusual 11-Carbon δ -Lactone Derivative from the Fresh Ripe Fruits of *Syzygium samarangense* (Wax Apple). *J. Nat. Prod.* 2022, 85, 2100–2103.
48. Van der Sluis, A.; Dekker, M.; Jongen, W. Flavonoids as bioactive components in apple products. *Cancer Lett.* 1997, 114, 107–108.
49. Hagen, S.F.; Solhaug, K.A.; Bengtsson, G.B.; Borge, G.I.A.; Bilger, W. Chlorophyll fluorescence as a tool for non-destructive estimation of anthocyanins and total flavonoids in apples. *Postharvest Biol. Technol.* 2006, 41, 156–163.
50. Ceymann, M.; Arrigoni, E.; Schärer, H.; Nising, A.B.; Hurrell, R.F. Identification of apples rich in health-promoting flavan-3-ols and phenolic acids by measuring the polyphenol profile. *J. Food Compos. Anal.* 2012, 26, 128–135.
51. Lancaster, J.E.; Reay, P.F.; Norris, J.; Butler, R.C. Induction of flavonoids and phenolic acids in apple by UV-B and temperature. *J. Hort. Sci. Biotechnol.* 2000, 75, 142–148.
52. Poirier, B.C.; Buchanan, D.A.; Rudell, D.R.; Mattheis, J.P. Differential partitioning of triterpenes and triterpene esters in apple peel. *J. Agric. Food Chem.* 2018, 66, 1800–1806.
53. Lancaster, J.; Dougall, D.K. Regulation of skin color in apples. *Crit. Rev. Plant Sci.* 1992, 10, 487–502.
54. Schwartz, S.J.; Lorenzo, T.V. Chlorophylls in foods. *Crit. Rev. Food Sci. Nutr.* 1990, 29, 1–17.
55. Schweiggert, R.M.; Vargas, E.; Conrad, J.; Hempel, J.; Gras, C.C.; Ziegler, J.U.; Mayer, A.; Jiménez, V.; Esquivel, P.; Carle, R. Carotenoids, carotenoid esters, and anthocyanins of yellow-

, orange-, and red-peeled cashew apples (*Anacardium occidentale* L.). *Food Chem.* 2016, 200, 274–282.

56. Marais, E.; Jacobs, G.; Holcroft, D.M. Postharvest irradiation enhances anthocyanin synthesis in apples but not in pears. *HortScience* 2001, 36, 738–740.

57. Jamrógiewicz, M. Application of the near-infrared spectroscopy in the pharmaceutical technology. *J. Pharm. Biomed. Anal.* 2012, 66, 1–10.

58. Mahanti, N.K.; Pandiselvam, R.; Kothakota, A.; Ishwarya S., P.; Chakraborty, S.K.; Kumar, M.; Cozzolino, D. Emerging nondestructive imaging techniques for fruit damage detection: Image processing and analysis. *Trends Food Sci. Technol.* 2022, 120, 418–438.

Forrásnyelvi szöveg 2



OPEN ACCESS

EDITED BY
Miguel Angel Prieto Lage,
University of Vigo, Spain

REVIEWED BY
Marta Barral Martínez,
University of Vigo, Spain
Sepidar Seyyedimansour,
Islamic Azad University of Tabriz, Iran

*CORRESPONDENCE
R. Pandiselvam
anbupandi1989@yahoo.co.in
M. R. Manikantan
manicpcr@gmail.com
Monica Trif
monica_trif@hotmail.com

SPECIALTY SECTION
This article was submitted to
Food Chemistry,
a section of the journal
Frontiers in Nutrition

RECEIVED 20 June 2022
ACCEPTED 02 August 2022
PUBLISHED 12 October 2022

CITATION
Pandiselvam R, Prithviraj V,
Manikantan MR, Kothakota A, Rusu AV,
Trif M and Mousavi Khaneghah A
(2022) Recent advancements in NIR
spectroscopy for assessing the quality
and safety of horticultural products: A
comprehensive review.
Front. Nutr. 9:973457.
doi: 10.3389/fnut.2022.973457

COPYRIGHT
© 2022 Pandiselvam, Prithviraj,
Manikantan, Kothakota, Rusu, Trif and
Mousavi Khaneghah. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Recent advancements in NIR spectroscopy for assessing the quality and safety of horticultural products: A comprehensive review

R. Pandiselvam^{1*}, V. Prithviraj², M. R. Manikantan^{1*},
Anjineyulu Kothakota³, Alexandru Vasile Rusu^{4,5}, Monica Trif^{6*}
and Amin Mousavi Khaneghah⁷

¹Physiology, Biochemistry and Post-Harvest Technology Division, ICAR – Central Plantation Crops Research Institute, Kasaragod, Kerala, India, ²Department of Food Engineering, National Institute of Food Technology Entrepreneurship and Management, Sonapat, Haryana, India, ³Agro-Processing and Technology Division, CSIR-National Institute for Interdisciplinary Science and Technology (NIIST), Trivandrum, Kerala, India, ⁴Life Science Institute, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Cluj-Napoca, Romania, ⁵Animal Science and Biotechnology Faculty, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Cluj-Napoca, Romania, ⁶Food Research Department, Centre for Innovative Process Engineering (CENTIV) GmbH, Stuhr, Germany, ⁷Department of Fruit and Vegetable Product Technology, Prof. Wacław Dąbrowski Institute of Agriculture and Food Biotechnology-State Research Institute, Warsaw, Poland

The qualitative and quantitative evaluation of agricultural products has often been carried out using traditional, i.e., destructive, techniques. Due to their inherent disadvantages, non-destructive methods that use near-infrared spectroscopy (NIRS) coupled with chemometrics could be useful for evaluating various agricultural products. Advancements in computational power, machine learning, regression models, artificial neural networks (ANN), and other predictive tools have made their way into NIRS, improving its potential to be a feasible alternative to destructive measurements. Moreover, the incorporation of suitable preprocessing techniques and wavelength selection methods has arguably proven its practical feasibility. This review focuses on the various computation methods used for processing the spectral data collected and discusses the potential applications of NIRS for evaluating the quality and safety of agricultural products. The challenges associated with this technology are also discussed, as well as potential future perspectives. We conclude that NIRS is a potentially useful tool for the rapid assessment of the quality and safety of agricultural products.

KEYWORDS

cultivar authentications, damage detection, maturity, rapid method, spectroscopy, texture

Introduction

Agro-based processing industries are trending toward the production of fresh and minimally processed commodities. In this context, new processes and products are now available on the market, driven by consumer interest. Agricultural products are the main raw material for many food industries (1, 2). They are highly perishable in ambient conditions and thus have a limited shelf life, which can be extended under refrigerated storage conditions. Because fruits and vegetables cannot be stored for a long time due to their susceptibility to deterioration, particularly microbial and chemical spoilage, the various processes in the post-harvest supply chain, such as grading and sorting, should be completed immediately after harvesting.

The use of proper analytical methods to ensure the quality and safety of end products is necessary before processing and throughout the post-harvest supply chain (3). However, conventional methods for detecting defects in fruits and vegetables have several disadvantages: The evaluation of raw fruit quality, cultivar authenticity, and damage to the products pose challenges to The evaluation of the quality of the raw fruit and the authenticity of the cultivar, as well as the assessment of any damage to the product, pose challenges to collecting good quality inputs from farmers or wholesale dealers. Many kinds of damage cannot be detected with a visual inspection alone, thus rendering the process of selecting high-quality products insecure. Conventional testing methods, such as sampling from big lots, are time-consuming and expensive (4–7). Moreover, a significant amount of product is often destroyed during conventional testing and sampling.

Specific techniques are needed to, for example, detect sun scaling in apples and weed out the defectives, as such injuries do not have any chemical treatments (8). Apples can also have internal browning issues that are not visible on the surface (9), thus again ruling out the possibility of visual detection. Similarly, fruits are often kept in refrigerated conditions to extend shelf life, but this poses a risk of causing chilling injury, which cannot be detected by peel color or any other aspect of external appearance (10). Therefore, the standard practice of local markets is to determine the price of fruits and vegetables based on their physical attributes, estimated quantitatively and qualitatively (11).

Non-destructive analytical methods thus play a vital role in overcoming the challenges of conventional laboratory methods. In this aspect, near infrared spectroscopy (NIRS), combined with predictive algorithms, is best suited to assessing product quality and damage detection and for identifying cultivars (12).

A NIR spectroscopic system consists mainly of an interconnected light source, spectrometer, and computer, as shown in Figure 1. (This diagram shows the simplest representation possible, and does not account for a more practical NIRS system). The light source emits photons in the

infrared region, which come into contact with the sample, and their further interactions create vibrations or stretching (13) in the molecules in the interior of the sample. These vibrations create a spectrum dependent on the properties of the molecule and their corresponding chemical bonds. These spectral profiles are dominant in specific parts of the spectrum. For instance, molecules like chlorophyll vibrate in the 500–750 nm region (8, 9) and O-H bonds in water in the 970–1,150 nm region (9). The spectra have to be analyzed with specific preprocessing techniques to eliminate noise and unwanted and redundant information. Later on, preprocessing techniques is implemented to find the effective wavelength, which spans certain portions of the spectrum in a way that can precisely classify the product, as shown in Figure 2. A properly built model can ably classify the product with little or no mistakes. Care should be taken when designing a NIR spectroscopic system to conduct a perfect analysis, using standard methods, with minimum time delay and high accuracy. The standard error in a laboratory (SEL) may lead to a standard error in prediction (SEP), which should be minimized during quantitative prediction with proper methods.

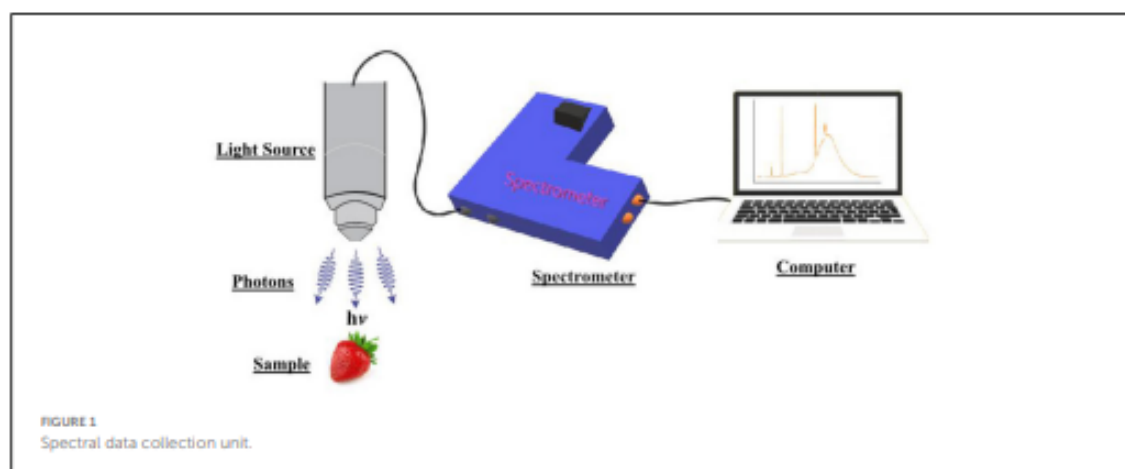
NIR spectroscopy applications have been reviewed for the processing of cereals processing (14) and seafood (15), for various applications with dairy products (16), and for quality analysis of meat and spices (4, 17). Nevertheless, it is paramount to understand the various applications of NIRS for agricultural products, as these applications could pave the way to preparing high-quality end products from these raw materials. In this context, the present review discusses various NIR spectroscopy measurement techniques and optimization strategies used in agricultural products.

Applications of spectroscopic techniques

Quality parameters

Food products, and their reliability, depend on the quality of the products from which they are derived. NIR spectroscopy is one of the best methods for predicting the primary characteristics of these products, *viz.* total soluble solids (TSS), soluble solids content (SSC), titratable acidity (TA), and pH, using spectral signatures and algorithms. The various applications of NIRS for determining the quality of agricultural products are shown in Table 1.

A study conducted by Lan et al. on apple quality evaluated the characteristics of the puree produced from the same. The apples' spectral details were used to predict viscosity, cell wall content, dry matter, SSC, and puree product TA. The parameters mentioned above showed R^2 values greater than 0.8, indicating the method's accuracy (18). The study's spectral measurements were carried out in a range of 800–2,500 nm using an automatic sampling wheel with 18 different positions. However, methods



using six optical fibers (9) or manual positioning are more realistic and accurate. This may be due to the consistency of the sampling wheel setup. The SSC content determination resulted in an R^2 value of 0.92 due to the homogeneity of the puree product (18).

The algorithms perform best when the product is homogeneous rather than inconsistent. Similar investigations were conducted on calçot onions and on apple purees, and also obtained accurate predictions for soluble solids, glucose, malic acid, and dry matter (33, 34). These studies indicate that the quality evaluation of purees can be successfully performed using spectral analysis.

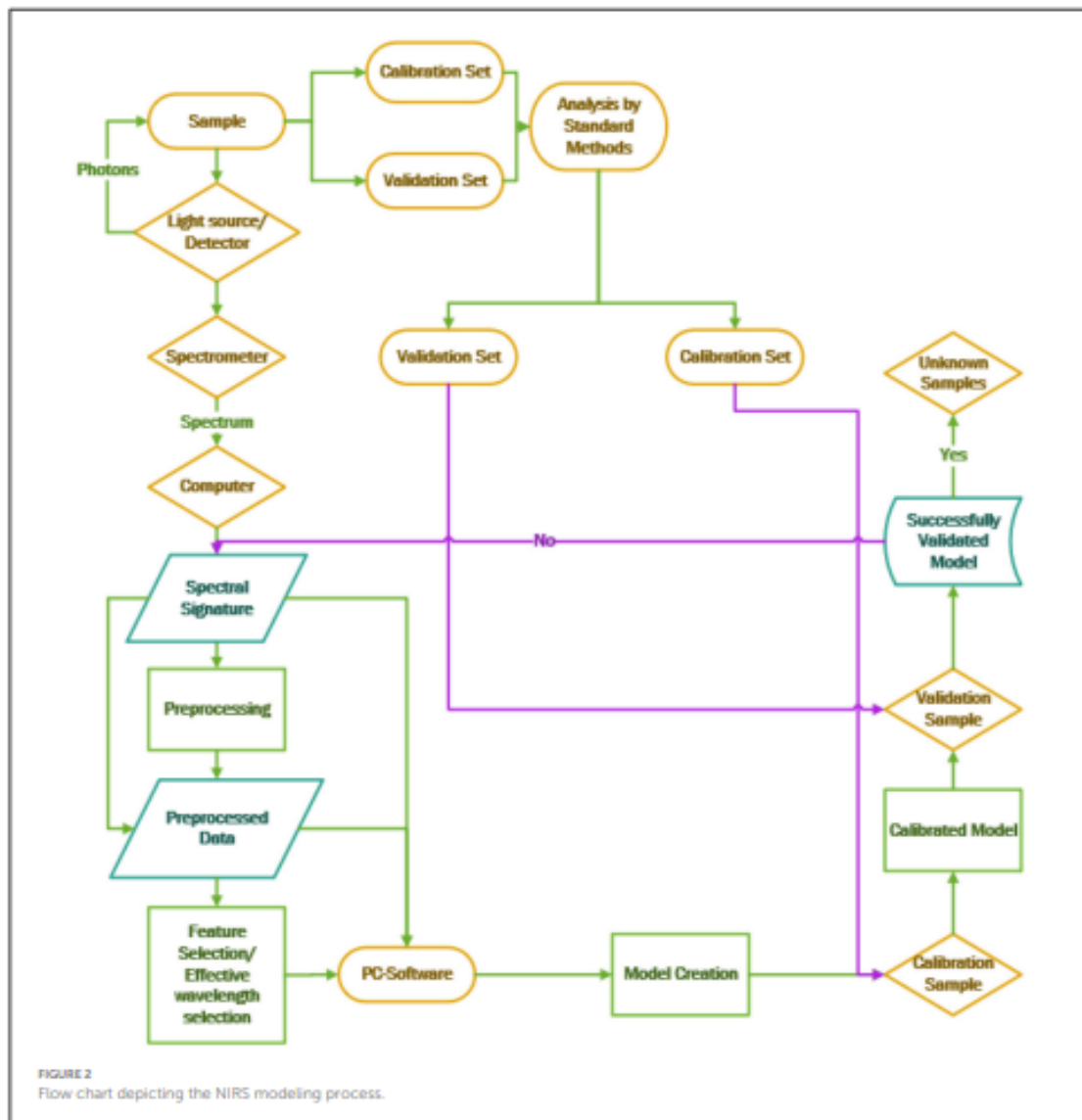
In these scenarios, the use of partial least squares (PLS) models with TSS prediction resulted in $R^2 = 0.95$, better than any other parameters (27). However, rheological constants and color values were found to underperform. This may be due to the non-linear nature of rheological variations or the overlapping of spectral bands (33). Prediction of puree characteristics by evaluating the quality of intact apples was challenging and semi-quantitative, suitable for the industrial process (33). NIRS works based on the chemical compositions and photon response; thus, processing commodities and converting raw produce to value-added end product will drastically alter the product's chemical nature. This may be why more latent variables and lower accuracy are obtained in predicting the quality characteristics of processed products.

Khodabakhshian et al. conducted internal quality studies on pomegranates using both transmittance and reflectance modes in the 400–1,100 nm range. Their study evaluated the model performance by using the predicted residual error sum of squares (PRESS) method as a cross-validation technique in the regression analysis. The usual pretreatments, particularly standard normal variate (SNV) and multiplicative scatter correction (MSC), were carried out, thus accounting for the morphological variation in the pomegranate varieties (27).

Practical case scenarios always need to account for morphological variation. Therefore, while SNV and MSC preprocessing can be used, it may not be the same for all situations. The overall characteristics of both the reflectance and transmittance spectra were similar and formed peaks in the 750 and 970 nm ranges, though more noise was observed in the transmittance spectra (27). The spectra obtained with wavelength depicted irregular spikes all over the data, indicating the noise problems. Distinctive peaks for spectral signature are the preferred method for NIRS analysis. Interestingly, preprocessing techniques remained similar for most of the quality traits irrespective of the crop, demonstrating the potential for efficiency at the point of analysis. Both transmittance and reflectance methods performed similarly, though the latter had a slight upper hand due to its higher penetrating power (21, 35, 36) and more robust system. Thin-peeled fruits such as pomegranate obtained improved results in reflectance mode, suggesting that this method can be used for thick-walled fruits such as coconut, areca nut, and cocoa.

An extensive study on SSC determination conducted by Hu et al. studied various measuring configurations, variable selection algorithms, and classification models with Hami melons. In this context, the calyx model performed the best, which could be due to the higher SSC content in the calyx region (23). This result can be further interpolated to obtain the best prediction results after all preprocessing. The measurement should be taken at the most prominent part of the attribute.

A similar study performed on apples incorporated a greater number of orientations, which resulted in the selection of stem-calyx vertical, with stem upward, as the optimal direction (29). This suggests that whether one orientation performs better than others is due to the prominence of the factor in the oriented region. Various combinations of variable selections and prediction algorithms suggest that the Monte Carlo-uninformative variable elimination-successive



projections algorithm (MC-UVE-SPA) attained the best results with all prediction algorithms for SSC determination ($R_p > 0.8$) (23). Xia et al. found that CARS-SPA-PLS performed best for SSC prediction in apples, obtaining a low root mean square error of prediction (RMSEP) of $< 0.573^\circ$ brix. They also found that the determination of effective wavelength from a global model could help reduce the effect of orientation to a certain limit (29).

Selecting effective wavelengths that correspond well to the regions that classify samples can reduce equipment expenses, as demonstrated by a study on a banana quality evaluation

where the lowest possible wavelength window was suggested (37). The effective wavelength of each agricultural product varies and therefore needs to be standardized during spectral analysis. In bananas and apples, this variation of change in the effective wavelength may be due to differences in color and overall composition. We can thus deduce that orientation, effective wavelength, products used, and equipment design should be considered while designing an online measuring system for optimum performance. Various studies have been able to predict SSC content with $R_p^- > 0.9$, obtaining better

TABLE 1 Uses of NIRS to determine the quality aspects of various agricultural products.

Agricultural product	Spectral range	Software package	Number of samples	Accuracy	Findings	References
Apple and apple purees	800–2,500 nm	OPUS v. 5.0, XLSTAT, MATLAB v. 7.5, R v. 3.5.2	240	Classification accuracy of apples = 82% and purees = 88%	Viscosity, cell wall content, dry matter, SSC, and titratable acidity showed best prediction ($R^2 > 0.8$). Final apple puree conditions were predicted from those of apples.	(18)
Apple	350–2,500 nm	MATLAB 2014a, Unscrambler v. 10.5x	120	SSC: ($R^2 = 0.87$, RMSEP = 0.55) pH: ($R^2 = 0.72$, RMSEP = 0.009)	Denosing with wavelet transform before pretreatment was found effective.	(19)
Tangerine	700–1,100 nm	MATLAB v. 7.0	275	94.0% using SSOM	SSOM is a non-linear classifier successfully used to detect MC, SSC, TA, and granulation.	(20)
Elderberry	800–2,500 nm	OPUS v. 7.2, Unscrambler 10.4, Statistica v. 8.0	Fruits from 11 orchards, blended	97.06%	TSS and SSC correlated well with spectral data, and pattern recognition was possible.	(21)
Apple	800–2,500 nm	OPUS v. 7.2	214	TA: ($R^2 = 68.17\%$, RMSEP = 0.12) TSS/TA: ($R^2 = 82.62\%$, RMSEP = 0.43) TSS: ($R^2 = 90.93\%$, RMSEP = 0.61)	Spectral intensities varied according to the spectroscope used. GA significantly improved the prediction quality.	(22)
Hami melon	550–950 nm	Unscrambler v. 9.7	120	RMSEP = 0.95–0.99	SSC determination obtained the best result for the MC-UV-SPA-MLR combined preprocessing and non-linear prediction algorithm.	(23)
Grapes	800–1,100 nm	Unscrambler v. 10.5	120	TA: ($R^2 = 0.716$, RMSEP = 0.103) pH: ($R^2 = 0.547$, RMSEP = 0.395) SSC: ($R^2 = 0.971$, RMSEP = 0.522)	Considerable variation was observed in connection with various pretreatments. Improvement in calibration results does not always improve validation results.	(24)
Pineapple	740–1,070 nm	MATLAB v. 9.5.0	90	85%	Good prediction accuracy for TSS with RMSEC = 0.95 and RMSEP = 0.84.	(25)
Olives	2,307–2,348 nm	UCal	100 (12 cultivars)	$R^2 = 0.964$ and 959	Oil content was predicted prior to oil extraction.	(26)
Pomegranate	400–1,100 nm	ParLeS v. 3.1, AvaSoft	100	TSS: ($R = 0.95$, RMSEC = 0.522)	Reflectance mode was better than transmission mode for determining TSS and pH.	(27)

(Continued)

TABLE 1 (Continued)

Agricultural product	Spectral range	Software package	Number of samples	Accuracy	Findings	References
Persimmon	1,000–2,500 nm	NIRware v. 1.2	147	$R > 0.75$, RPD < 1.5	SSC predictions obtained the best result with MSC pretreatment.	(28)
Apple	550–950 nm	Unscrambler v. 10.1, MATLAB 2016a, SpectraSuite, Visual Studio 2010	180	$r_p = 0.842$, RMSEP = 0.453	Keeping the stem–calyx axis vertical, with stem upward, was found to be the best orientation. PLS and LS-SVM were used to create compensation models, CARS and SPA to select effective wavelengths.	(29)
Grape	450–2,500 nm	Vision software, GenStat	120	TSS: ($R^2 = 0.896$, RMSEP = 0.308) TA: ($R^2 = 0.835$, RMSEP = 0.066) TSS/TA: ($R^2 = 0.812$, RMSEP = 0.451)	The balance between sugar and acid was taken as a quantitative parameter, which was correlated with the perception of taste.	(11)
Tomato	400–1,100 nm 900–1,700 nm	MATLAB R2016	600	pH: ($r_p = 0.819$) SSC: ($r_p = 0.800$)	Spatially resolved spectroscopy was found effective for agricultural crops with heterogeneous structures and chemical compositions.	(30)
Tomato	930–1,650 nm	SpectraWiz, ParLe5 v. 3.1	120	Lycopene: ($r_{cv} = 0.840$, RMSECV = 2.256) Vitamin C: ($r_{cv} = 0.818$, RMSECV = 1.087)	MSC combined with first derivative was able to predict lycopene and vitamin C content through the PLS model. The NIR process used here does not afford high accuracy.	(31)
Oranges and grapes	450–2,500 nm	Vision TM v. 3.5.0.0, GenStat	120 Grapefruit 120 Oranges	TSS: ($R^2 = 0.927$, RMSEP = 0.283) TA: ($R^2 = 0.929$, RMSEP = 0.017) TSS/TA: ($R^2 = 0.958$, RMSEP = 0.2) TSS: ($R^2 = 0.896$, RMSEP = 0.308) TA: ($R^2 = 0.835$, RMSEP = 0.066) TSS/TA: ($R^2 = 0.812$, RMSEP = 0.451)	The sweetness and flavor attributes of oranges and grapes were studied. The organoleptic parameter BrixA (Brix minus acids) was evaluated.	(32)

SSC, soluble solid content; RMSEP, root mean square error of prediction; SSOM, supervised self-organizing map; MC, moisture content; TA, titratable acidity; TSS, total soluble solids; GA, genetic algorithm; RPD, residual predictive deviation; MSV, multiplicative scatter correction; MC-UVI, Monte Carlo–uninformative variable elimination; SPA-MLR, successive projection algorithm–multiple linear regression; CARS, competitive adaptive reweighted sampling; RMSECV, root-mean-square error of cross-validation.

results than with other factors, such as TA and pH (24). Similarly, the TSS content of pineapple was compared by combining various algorithms (25). This latter study, which used a handheld spectrometer, obtained an accuracy rate

of 85% with an RMSEC of 0.95 and an RMSEP of 0.84 (25). This technology offers the best practical results in terms of feasibility and future perspectives, even with using handheld equipment.

Kanchanomai et al. (24) investigated potential rapid evaluation techniques by determining the quality of grapes using SSC, pH, and TA. The SSC prediction obtained an R_p value of 0.97, whereas TA ($R_p = 0.71$) and pH ($R_p = 0.54$) were comparatively less accurate. Firmness and seedlessness were also studied, but these factors also had low prediction accuracy. The researchers concluded that a NIR range of 800–1,100 nm could be suitable for predicting internal quality in grapes. Similarly, table grapes were analyzed using various preprocessing techniques and a NIR range of 400–1,000 nm (27, 38). The proper preprocessing technique improved the R_{Pred} from a range of 0.6 to 0.8052 and above when Savitzky-Golay's second derivative (SG2) was applied. This suggests that various factors such as pH, TSS, and firmness can be predicted by applying SG2 preprocessing in table grapes. The better results obtained can be justified using the proper wavelength range and preprocessing technique used on table grapes.

A combination of back propagation neural network (BPNN), generalized regression neural network (GRNN), and particle swarm optimization (PSO) was used in order to determine SSC and total acidic content (TAC) in apples (39). This study focused on developing a hybrid artificial neural network (ANN) model, as mentioned with earlier techniques; this was necessary to overcome the inherent limitations within ANN. The results showed that, during SSC prediction, the hybrid model and BPNN model had nearly the same RMSEC and RMSEP values (<0.6 and <0.7, respectively), whereas the GRNN did not perform well (values of >0.6 and >0.9, respectively). However, during TAC determination, the GRNN and hybrid models had low RMSEC and RMSEP values (<0.1), whereas BPNN did not perform well (>0.2). This confirms that an adaptation of hybrid models indeed increased the versatility of the algorithms to determine various quality attributes rather than having them to be changed with the parameter.

In terms of online determination methods, various systems have been explored for use with apples, mangoes, and bananas. Proper orientation of the product, as indicated in the previous finding, proved to be one of the key challenges facing the online system design. In an effort to address this challenge, the use of the CARS-SPA-PLS model after SGS was studied and was found to be effective in apples (29), with an $r_p > 0.8$. A similar system of MCARS-SPA-PLS was suggested for online apple prediction with higher accuracy. These algorithms can handle biological variability and tackle orientation issues during prediction. Online systems were found highly suitable for squash (40) and juice products, irrespective of the orientation challenges. This is likely due to the inconsistency in the fruits and homogeneity in the value-added products.

Highly perishable produce such as banana needs rapid quality analysis measures. An online conveyor system for the same was designed that took TSS, pH, dry matter (DM), and acid-brix ratio (ABR) into account (37). During the validation phase, corresponding R^2 values of 0.81, 0.78, 0.78, and 0.87

were found, indicating highly accurate prediction results. The study also selectively carried out preprocessing for both pH and ABR, which improved the prediction results. The accuracy of online detection was higher for apples, as the combination of algorithms 'successfully managed the difficulties of product orientation. Therefore, similar combination algorithms need to be adapted for online systems, given the challenges of performing these measurements with precise orientation.

Textural properties

Texture is one of the key factors influencing the quality, feel, and appearance of agricultural products. The most common parameters used to measure the textural quality of fresh produce are firmness and penetrometer readings (41, 42). Some studies have analyzed a wider variety of parameters, viz. fracture force, hardness, apparent modulus of elasticity, compressive energy (43), initial firmness, average firmness, rupture force of peel, rupture distance, penetration energy, and penetration force (42). These parameters can distinguish varieties (41), as different species and biological conditions have unique ranges. It is also worth noting that some studies have opted to use a universal testing machine (UTM) rather than a textural analyzer for evaluating the firmness and penetration depth of more oversized objects.

The summary of the various applications of NIRS for measuring the textural qualities of agricultural products is depicted in Table 2. Sánchez et al. (44) analyzed spinach texture using a handheld NIR device connected to a micro-electrical mechanical system (MEMS). The researchers used a simple handheld device for NIR measurement, thus demonstrating the utility of compact, practical technologies over more complex, high-cost instruments. The response of the NIR spectroscopy to CH, OH, and NH bonds and their absorbance patterns creates crests and troughs (41) in the spectral signature, which is the base of analysis. It was observed that the use of the handheld spectrometer resulted in an R^2 value nearly equal to 1. The commodities used in the experiment were taken from different farms in Spain. Practices like this (i.e., that aggregate a greater amount of varieties and biological variations) allow for better calibration and data processing, which in turn contributes to the building of more robust prediction systems and, ultimately, yields better results.

This assertion is supported by findings from other studies, an investigation of methods for determining tomato texture in which tomatoes in similar maturity groups gave distinct clusters in a scatter plot (41). Distinct clusters will not account for the various maturity stages and varieties present in the sample; therefore, the sample size should be increased to create a better performing model. The textural tests were performed using a universal testing machine rather than a textural analyzer (44, 47). The food matrix tested for puncture force, toughness,

TABLE 2 Uses of NIRS to determine the textural properties of agricultural products.

Agricultural product	Spectral range	Software package	Number of samples	Accuracy	Findings	References
Mango	800–2,500 nm	OPUS v. 7.0.129	85	$R = 0.7–0.75$	Penetration parameters, firmness, and rupture force were predicted.	(42)
Spinach	1,600–2,400 nm	WinISI II v. 1.50	140	$R^2_c = 0.74$	Consumer acceptance parameters, dry matter, and textural properties were identified.	(44)
Tomato	800–2,500 nm	OPUS	30	$R^2 = 0.7–0.97$ RMSECV = 0.4% Brix	SSC content was predicted using PLS analysis. The textural prediction was accurate.	(41)
Tomato	950–1,650 nm	SPSS 13.0, PermutMatrix v. 1.9.3, Unscrambler v. 9.7, MATLAB 2017	90	$R^2_p = 0.85–0.966$ RMSEP < 0.5	Raw spectra outperformed preprocessed ones. ELM obtained the best prediction over PLS and SVM.	(43)
Kiwi	800–900 nm	ImageJ, BGS-AvaCam v. 3.7.0, MATLAB R2016b, Unscrambler v. 9.7	116	$R^2 = 0.68–0.77$	Firmness was influenced by similar gray pixels. PLS give slightly better results than ANN.	(46)
Olive	1,100–2,300 nm	SNAP 2.03, Unscrambler v. 9.7, SigmaPlot v. 10.0	100	$R^2 = 0.8–0.99$	A large data set can account for several factors and thus achieve accurate prediction. A rapid and inexpensive technique was used.	(47)
Pear	380–1,030 nm	Spectral Image System, Unscrambler v. 9.7, ENVI v. 5.0, MATLAB v. 8.1	135	$R^2_p = 0.89–0.92$	Scanning speed > 1.5 m/s is needed for real-time application. Deep learning using SAE-FNN was performed.	(48)
Pistachio	200–1,100 nm	AvaSoft7, ParLe5 v. 3.1	81	$R^2 = 0.754–0.91$ RMSEP = 0.253–26.049	The SNV model performed better than MSC. Among textural parameters, fracture force was predicted the most accurately.	(41)

SSC, soluble solid content; RMSEP, root mean square error of prediction; RMSECV, root-mean-square error of cross-validation; PLS, partial least square; ELM, extreme learning machine; SVM, support vector machine; ANN, artificial neural network; SAE, stacked auto encoder; FNN, fully connected neural network; MSC, multiplicative scatter correction.

stiffness, fracture point (44), and firmness (47). Maximum force to puncture the leaf ($r^2_{cv} = 0.67$; RPD $_{cv} = 1.72$), toughness ($r^2_{cv} = 0.62$; RPD $_{cv} = 1.62$), stiffness ($r^2_{cv} = 0.69$, RPD $_{cv} =$

1.79) and the displacement of the probe necessary to fracture each leaf ($r^2_{cv} = 0.62$, RPD $_{cv} = 1.61$) (44). Similarly, a firmness measurement performed on olives recorded a high R^2 value of

0.997, indicating excellent-quality predictions of this parameter (47). There is a significant improvement in R^2 value with suitable preprocessing, as illustrated by a study on pistachio kernels comparing raw data and various preprocessed data (43).

The selection of the preprocessing method to be used with textural analysis depends on the parameter under study and the instruments used in the measurement. A study conducted by Mohammadi-Moghaddam et al. (43) was able to model better-performing PLS models for textural analysis after various pretreatments. It is worth noting that pretreatment is not always necessary, and that better results were obtained in determining the firmness of pears (via PLS regression) without preprocessing (48). Direct comparison between different pretreatments to determine the best choice is not possible, as it depends on several factors.

In olives, firmness is understood to correlate with maturity and oil content (47) due to the relationship between oil and dry matter. As these factors are mechanically and structurally connected, defining the mouth feel via NIRS can offer a non-destructive method of measurement. Although firmness prediction is a bit complicated, one study on olives obtained readings of 1,248, 1,449, 1,758, 1,917, 1,990, and 2,238 nm at peak regions using PLS regression, thus delivering high correlation and low error rates (47). A similar study using cherry tomato resulted in a firmness prediction of $R^2 = 0.966$ (45). The algorithm for the extreme learning machine used in this study works by optimizing the hidden layers in the cross-validation step, thus solving practical problems (45).

Compression and penetrometer tests were performed together on tomatoes by Camps and Gilli (41), with the compression test demonstrating better prediction performance ($R^2 = 0.85$ – 0.97). This observation is further supported by a study conducted on juicy stone fruits that found that the compression test was significant as the fruit matured (49). A textural analysis on tomatoes revealed that there are about 7–10 latent variables for compression parameters and anywhere between 3 and 8 for penetrometer parameters (41), and that data clustering may be the reason behind better R^2 results. However, although clustered data can improve R^2 values, the RMSE values may not differ much.

Studies on the correlation between texture and quality in mango suggest that, based on the R^2 value, the parameter should either be used for rough screening or for thorough evaluation. On this basis, one study found that the parameters of pulp penetration force, peel rupture force, and penetration energy in the pulp can be used for rough screening (42). The authors also discussed the relationship between textural parameters and other factors such as maturity, storage time, and processing stage. Processing operations such as roasting can also cause a decrease in moisture content, resulting in more reflectance (43); similarly, a difference in amplitude of peak has been observed in pistachio kernels after processing. Camps and Gilli (41) studied three tomato varieties with 90 samples outperformed this study

on the same parameters, although the sample size was 80 on a single variety. These results confirm both the highly variable characteristics of tomatoes and the research gaps regarding these characteristics. Yu et al. (48) proposed a deep learning model with a stacked auto-encoder (SAE) and fully connected neural network (FNN), where SAE was used to extract the features which were given as FNN input. An investigation conducted on pear fruit compared the PLS regression with a deep learning algorithm, and significantly better results were obtained for SAE-FNN (48). Deep learning predicted the R^2 value in the range of 0.9, whereas the PLS regression could only reach values near 0.8 (48). Experiments on firmness prediction in kiwi fruit found that, on average, PLS ($R^2 = 0.77$) performed better than ANN ($R^2 = 0.72$) (46).

Variety/cultivar identification and authentication

Authentication and traceability throughout the supply chain are essential, especially where technologies such as block-chain methods are implemented. Today's consumers are focused on authenticity; they demand traceability and strict quality controls in production. In recent years, NIRS, along with classification algorithms, has been integrated into supply and production chains to meet these demands.

The development of models that can discriminate between different genotypes of apples (50–53), bell peppers, (54), mangoes (55), pears (56), potatoes (57), and mulberries (7) using the NIR spectrum is summarized in Table 3. In practical terms, the sorting of these commodities should be done on a fast-moving conveyor system, but designing a spectrum-collecting system for this environment is troublesome. Because even agricultural products of the same variety can have different surface properties, obtaining multiple spectral samples, as large as 250 readings (7, 55, 59), is necessary. The sample size and the number of measurements needed are determined by the heterogeneity and complexity of the sample, such as whether the product is sliced (59). The heterogeneous nature of the sample is accounted for by techniques such as taking three or four different types of measurements; by the use of 120 and 90-degree rotations (54, 58); and by the use of fabricated fruit holders (59, 61) or by using the arrangement in the spectrometer. Standard NIR models require whole fruit cover scanning using arrangements of diode-array instruments (44), the integration of the sphere around them to recover information that is otherwise lost (52), and methods to keep the distance from fruit to measurement probe constant, irrespective of the size (50).

Most researchers are using the existing spectrometer system directly, and therefore more emphasis has been given to the classification methods. A broader range of photometers and a larger sample size can significantly improve the model (54).

TABLE 3 Summary of variety and cultivar identification for various agricultural products using NIRS.

Agricultural product	Spectral range	Software package	Number of samples	Accuracy	Findings	References
Bell pepper	1,600–2,400 nm	Win ISI II v. 1.5, MATLAB 2015a	394	88.28–91.37%	Preliminary screening using SSC and dry matter was a success. The importance of SEP and SEL was discussed.	(54)
Mulberry	909–1,649 nm	PLS Toolbox v. 6.21, MATLAB R2009	468	84.1%	<i>Dendrobium officinale</i> Kimura et Miyo (DOK) was distinguished from <i>Dendrobium devonianum</i> Paxt (DDP).	(7)
Apple	1,000–2,500 nm	MATLAB 7.11, Antaris II System	180	74.44%	Among PCA, PCA+LDA, SDA, and DPLS, SDA was found to have better performance for feature extraction.	(58)
Apple	1,000–2,500 nm	Fiber Optic Solids cell, NIRWare Unscrambler,	410	77.9%	Classification of apples according to various terrain types.	(51)
Tangerine, red cabbage, cornichons, kale and applesauce	1,100 nm and 2,100 nm	PAS LABS v. 1.2, SIMCA v. 14.1	15	99%	NIRS prediction was possible for commodities kept inside glass. OPLS-DA outperformed PCA and PLS-DA.	(59)
Potato	964.13–1645.01 nm and 2502.50–16666.67 nm	SpectralCube, OPUS v. 7.2, PLS-toolbox v. 8.6, Unscrambler v. 10.1, MATLAB R2017b	240	RP = 0.954 RMSEP = 0.421	A PLSR model was used to find the degree of doneness and predict the variety.	(57)
Apple	300–1,100 nm	ModelBuilder, R Statistical software	640	R ² values were 0.90 and 0.92 and RMSE were 0.67%	Individual models for cultivars performed better than the combined model.	(33)
Mango	1,200–2,200 nm	Unscrambler	1,310	Alphonso and Banganapalli (99.07%, 99.58%), Dusheri and Malda (98.37%, 94%)	A distinct score plot allowed for more accurate classification.	(55)
Apple	400–1,021 nm	Ocean View, MATLAB R2014b	300	SPA-SVM 85.83% SPA-ELM 95%	Among BPNN, SVM and ELM models, ELM performed better. Feature selection with SPA combined with ELM produced better results than PCA.	(60)

(Continued)

TABLE 3 (Continued)

Agricultural product	Spectral range	Software package	Number of samples	Accuracy	Findings	References
Pears	350–1,800 nm 350–1,000 nm 1,000–1,800 nm	Unscrambler v. 9.7	110	R^2 0.90–0.92 RMSEP 0.23–0.30	Feature selection was obtained better with CARS than with MC-UVE and SPA. CARS-MLR and CARS-PLS accurately determined SSC.	(56)
Apple	1,000–2,500 nm	MATLAB R2014a	208	98.1%	Geographical region had a significant effect on SSC. CARS feature selection and PLS-DA had good prediction accuracy.	

SSC, soluble solid content; SEP, standard error in prediction; SEL, standard error in laboratory; PCA, principle component analysis; LDA, linear discriminant analysis; SDA, stacked denoising autoencoder; DPLS, dynamic partial least squares; OPLS, orthogonal partial least-squares; PLS-DA, partial least-squares discriminant analysis; PLSR, partial least squares regression; BPNN, back propagation neural network; SVM, support vector machines; ELM, extreme learning machines; SPA, successive projection algorithm; CARS, competitive adaptive reweighted sampling; MC-UVE, Monte Carlo–uninformative variable elimination; SDA-MLR, successive projection algorithm–multiple linear regression.

Broader models are useful for obtaining unique peaks in the spectrum that can discriminate between varieties. Because fluctuations in temperature and light play a vital role in creating the NIR model, care should be taken to keep the surrounding environment the same for all measurements (53).

For cultivar prediction, the input data are preprocessed using methods such as multiplicative scatter correction (MSC) (51, 55, 60), EMSC (50, 62), standard normal variate (SNV) (7, 51, 56), detrend (51, 55), normalization (58, 59), Savitzky-Golay (7, 50, 51, 61), and Norris gap (55). These are independent reference techniques for eliminating the unwanted effects of irrelevant information in the spectra (52). Without these techniques, the noise would be learned along with the true calibration data, causing over-fitting (53). The application of the various treatments is based on the corrections required by the scenario, and it is up to the researcher to decide which corrections are suitable. Preprocessing is a necessary step in variety detection because the produce being measured is in raw form, and the noise level can therefore be high. However, it should be kept in mind that the use of smoothing techniques, such as those listed above, to process raw data can result in the loss of valuable information.

The investigation of fruit and vegetable spectra using produce housed in glass containers during preprocessing did not yield better results than raw data, at least in a scenario where a reference-dependent orthogonal partial least squares (OPLS) model contained an integrated OSC filter (59). The filter's effectiveness was evaluated by a response permutation set and performance matrices, viz. sensitivity, specificity, efficiency, false negative, false positive, and true negative (59).

A direct spectral comparison may not always be possible due to crossovers and overlapping. Li et al. (60) measured 300 samples of apple and observed spectra between 400 and

1,021 nm, and all of which were largely similar to each other. The observed spectra consisted of various attributes, so dimension reduction techniques such as PCA and the score plot were later used for classification. Dimension reduction was necessary to eliminate irrelevant and redundant spectral variables (56). In the context of cultivar identification, PCA was able to retain 98% of the data for apple cultivation (51), which makes it an ideal feature extraction method.

Jha et al. (55) investigated the classification of apple varieties and obtained 99.07% and 99.58% accuracy for Alphonso and Banganapalli mangoes, respectively. Score plot variance typically decreases from PC1, PC2, and PC3; thus, most researchers use the first two (51) or the first three (50). The variety inherent within the identification process needs a score plot where similar results are grouped or clustered in one area. However, choosing PC2 and PC3 regardless of PC1 can result in better models with a better grouping (54). Variable selection should be made using the genetic algorithm (GA), the successive projection algorithm (SPA), Monte Carlo–uninformative variable elimination (MC-UVE), and competitive adaptive reweighted sampling (CARS), and these selected variables should be combined with PLS or multiple linear regression (MLR) in a later stage for predicting (7, 54, 61). Comparative evaluations of these techniques have suggested that CARS has the best performance, resulting in R_p values near 0.9 and RMSEP values near 0.5 (7, 52, 56). Pattern recognition can also be achieved using machine learning methods such as backpropagation neural networks (BPNN), support vector machines (SVM), and extreme learning machines (ELM). These methods are suitable for identifying the variety and geographical origin of fruits based on the spectrum. The number of factors and groups plays a substantial role in the performance of the model. Reducing the number of possible predictions increases the model's accuracy. In one study of

apples from different orchard levels, for example, the models performed well only for classification between valley and mountain—i.e., a 2-variable model (51).

Individual models are more accurate than multivariate models (53), but multivariate models need to be developed in practical scenarios. In cases where individual models are deliberately selected for cultivar predictions, data overfitting and insufficient external validations are possible pitfalls (53). Model development requires a suitable training and testing set, and it is ideal to have the same number of training and testing sets for each of the varieties that need discrimination. If the number of elements is not equal between the training and testing sets, there can be a class imbalance, which can lead to biases within the model. Imbalances are usually dealt with by using under- and oversampling techniques, though other methods are possible; Zimmer and Schneider (59), for example, evaluated a well-handled model without treating class imbalance. Models are trained with data from destructive methods and laboratory evaluation. Therefore, the model's standard error in prediction (SEP) depends highly on the standard error in the laboratory (SEL). Sánchez et al. (54) suggested that SEP values greater than or equal to 5 times the SEL value indicate a model of unacceptable quality. For this reason, in a quantitative prediction, it is better to estimate the repeatability and reproducibility of measurements for better understanding. Most of the research comes with the significant caveat that external evaluation is lacking (53), but evaluation of this kind is necessary for a model used to predict cultivars and varieties from various places.

In the case of variety and cultivar identification, the models are used to classify qualitative data, and this automatically enables a quantitative classification (54), i.e., different cultivation locations confer a set of internal qualities that are unique to each variety (60). Eisenstecken et al. (51) and Li et al. (61) conducted experiments to determine varieties grown at different elevations, where quantitative factors (viz., carbohydrate, fructose, sucrose, glucose, sorbitol, and citric acid) displayed a predictable variation according to the elevation at which the fruit was cultivated. This knowledge, combined with a multi-origin regression model, can predict quality parameters such as pH, firmness, acidity, and moisture (52). Nevertheless, the models created during one season are not suitable for the next season due to natural variation in the environment, which in turn creates variation in crop attributes (53).

Maturity detection

Maturity detection is one of the main uses of NIR in the context of agricultural products. Due to variations in biological factors, many crops do not have standard maturity indices (63). While there is no way to directly predict the maturity of the fruit, factors such as color, firmness, TSS, TA, and dry matter are often

correlated with maturity. Among these parameters, color and the pigment component responsible for color are frequently used for maturity detection. This may be because the maturity of fruit was traditionally determined by analyzing the color with the naked eye. One NIR study of lipophilic antioxidants and firmness in tomatoes found a particular trend in the spectral lines as they approached maturity (64). Although this study is not directly related to maturity prediction, the results suggest the potential of NIR for maturity detection.

Fuzzy logic algorithms are among the most recent technologies introduced for maturity detection. Chen et al. (65) evaluated pomelo maturity using a least square support vector regression (LSSVR) model and PCA algorithm with "fuzzy-optimized" NIR data. Compared to ordinary PCA, the fuzzy transform PCA inhibits original data noise and emphasizes PC (65). The LSSVR model is represented by Equation 1.

$$Y = \omega \cdot \varphi(x) + b \quad (1)$$

where ω = matrix of regression weight; $\varphi(x)$ = kernel function, which is taken as radial basis function (RBF); b = threshold.

The study divided the samples into three categories, viz. calibration, validation, and test. The test samples are not related to calibration or validation; therefore, they can be used to check whether the model is representative (65). The maturity parameter, which reflects color (L, a, b) values, is thus obtained using the model.

The various applications of NIRS for maturity detection are presented in Table 4. A study of maturity detection in grapes using PLS modeling included variable factors such as cultivar, location, ripeness level, and season (66). Compared with a study that considered only varieties collected from a single vendor for classification, validation, and testing, the former approach is more practical and industry-relevant. Creating the model on the 2016 variety and validating it with the 2017 variety (66) proved the model's long-term robustness. Obtaining the best possible prediction results for each of the parameters under study required the use of different preprocessing techniques; some examples are MSW+MSC for TSS and SNV for TA (66).

The maturity of kiwi fruit and mango was predicted by measuring dry matter (DM) content and solid soluble content (SSC) (63, 73). Although SSC can be used for maturity detection by measuring the starch in the mesocarp converted to sugar, this method is not recommended in mangoes (63). An investigation of avocado fruits predicted their maturity by estimating moisture content (MC), but this method did not fit well due to the poor spectra obtained (67). The experiment underperformed at first due to poor peel penetration, but improved once the peel was removed. This suggests that NIR cannot be used for all purposes. Evaluations carried out with and without flesh had the same number of latent variables, but the former's R^2 value was 0.6, which later improved to 0.8 (67). In light of these issues, the corresponding challenges in fruits with similar properties to

TABLE 4 Uses of NIRS for the maturity detection of various agricultural products.

Agriculture produces	Spectral range	Software package	Number of samples	Accuracy	Findings	References
Grape	800–2,500 nm	OPUS v. 7.2	267 + 71	$R^2_y = 0.28\text{--}0.77$, RMSEP = 0.14–7.83	TSS and TA parameters can discriminate between grape varieties. Consumer needs and sensory parameters are correlated.	(66)
Avocado	940–1,798 nm	Latentix v. 2.12	> 10,000	$R^2 = 0.732$, RMSEP = 1.83	Evaluated the performance of a portable NIR device. Conducted external validation with different seasons.	(67)
Mango	306–1,140 nm	Unscrambler v. 10.3	149	$R^2_{cv} = 0.84\text{--}0.87$, RMSE _{cv} = 139%	PLS prediction worked well for SSC and DM. Firmness had poor calibration.	(63)
Watermelon	802–805 nm	OOIBase32, Unscrambler v. 9.7, MATLAB v. 7.10 R2010b	200	76.7–85.1%	RPP and NDIP techniques were used. Over-maturity was identified.	(68)
Tomato	285–1,200 nm	Model Builder v. 1.1.0.105, SAS v. 9.2		$R^2 = 0.67\text{--}0.86$, RMSCV = 0.64–1.082	Readings of lipophilic antioxidants were only partially correct, an destructive measurement techniques were used. Online measurement is possible with suitable improvements.	(64)
Banana	400–1,000 nm		45	$R^2 = 0.89$, RMSE = 0.000598	Chlorophyll content was predicted. Skin color was related to maturity.	(69)
Pomelo	400–2,500 nm		168	$R^2 = 0.893\text{--}0.912$, RMSE _T = 0.87–7.28	Fuzzy logic application in LSSVR with RBF kernel. Online quality determination with Vis-NIR is possible.	(65)
Pomelo	380–2,520 nm	QualitySpec Pro		$R_p = 0.913\text{--}0.997$, RMSE _p = 0.59–5.82	MOPLEC preprocessing and MWPLS were combined to form a robust model. Color parameters L*, a*, and b* were close to LabMeas values.	(70)
Blueberry	400–2,500 nm	Sieva 3.0, IBM SPSS Statistics 19	300	$R = 0.714\text{--}0.970$	L _{AD} variations and anthocyanin levels were related.	(71)

(Continued)

TABLE 4 (Continued)

Agriculture produces	Spectral range	Software package	Number of samples	Accuracy	Findings	References
Apple	200–1,100 nm	Spectrawiz	170	93.27–99.62%	The color and maturity of apples were correlated. Prediction was carried out using ANN-SA and was successful.	(72)
Kiwi	729–975 nm	SAS v. 8.2, F-750 Model Builder	100	$R^2 = 0.73$, $R = 0.48-0.74$	Dry matter and SSC can be estimated from NIRS. The edible quality of baby kiwi fruit was predicted from the unripe stage.	(73)
Mango	700–990 nm	Spectrasuite, XLSTAT 2014.1, SAS v. 9.4	1,200	70–72%	Maturity can be determined using NIR measurements of DM and TSS.	(74)
Pineapple	740–1,070 nm	MATLAB v. 9.5.0	90	100%	Organic and conventional cultivation was perfectly determined.	(25)

TSS, Total soluble solids; TA, Titratable acidity; RMSEP, Root mean square error of prediction; RMSECV, Root mean square error of cross validation; PLS, Partial least squares; SSC, Soluble solids content; DM, Dry matter; RPP, Ratio of intensity between peak1 and peak2; NDDI, Normalized difference intensity of peak; LSSVR, Least square support vector regression; RBF, Radial basis function; MOPLEC, Modified optical path length estimation and correction; MWPLS, Moving window partial least squares; ANN-SA, Artificial neural network simulated annealing algorithm.

avocado need to be addressed, and proper methods need to be formulated.

Temperature has a direct impact on spectral data due to the changing behavior of chemical compounds. Therefore, temperature should be stabilized before measurement, as has been done with kiwi fruit spectra (73). Kim et al. (73) also evaluated the use of Ca-Chitosan coating on kiwi fruit and its effect on maturity. The application of NIRS was able to increase SSC content due to a decrease in the ripening rate (73). The model was able to predict maturity, with R^2 values of 0.73 (73), an acceptable result. Fascinatingly, these results show that applying coatings can change the spectral behavior due to the interaction between photons and the new chemical compounds. However, due to the use of algorithms and processing techniques, all coated fruits with the same compounds will show spectral differences only in other variables, such as SSC and DM, given that a uniform application is performed (Figure 3).

Investigations of banana maturity by Saputro et al. (69) correlate chlorophyll content with maturity determination. A combination of PCA, PCR, and SVM achieved an RMSE value of 0.000598%, implying a good performance. The chlorophyll bands are situated in the 680–700 nm range, which was estimated with the help of NIRS (69). In mangoes, SSC

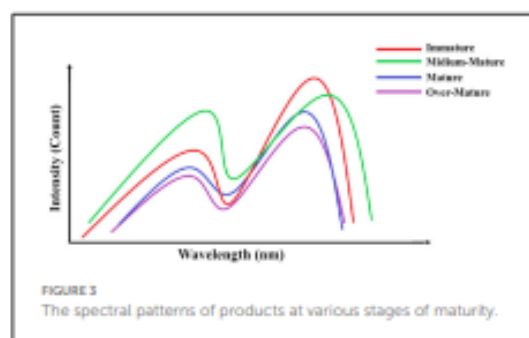


FIGURE 3
The spectral patterns of products at various stages of maturity.

determination was in the range of 699–999 nm and DM at 699–981 nm, which resulted in $R^2 = 0.874-0.87$ in cross-validation (63).

Ribera-Fonseca et al. (71) evaluated the potential of a non-invasive tool to predict the fruit harvest date using maturity parameters. The device measures an index of absorbance difference (IAD), which has a high correlation with fruit quality parameters (71). This approach could not make use of any conventional algorithms; therefore, it relies on IAD values only. Maturation studies of mangoes by Polinar et al. (74) created a

classification model that predicts mature and immature fruit. It was based on days after flower induction (DAFI) by setting the threshold as floaters <25%. Prediction of the ripe stage using these techniques was able to detect artificially ripened fruit with a bland taste (74). These innovations can be used to create auto sorting, identify higher-quality fruits, and prevent fraud in markets.

Watermelon maturity prediction was carried out using the ratio of peak (RP) method. The RP is used when two peaks have opposite intensity variation as the fruit matures. In watermelon, after calculating the ratio of the intensity of peak 1 and peak 2, the normalized difference intensity of peak (NDIP) was used to determine the maturity stage of the fruit (68). The SPA successive projection algorithm (SPA) was used for feature extraction, and least square support vector machines (LS-SVM) was used for classification purposes (68). The plot of the ratio of intensity between peak 1 and peak 2 (RPP) vs. maturity stages was able to distribute the fruit according to different maturity stages (68). Thus, the RPP alone distinguished various maturity stages, without performing PCA or LDA to plot with the highest variations. The correct classification rate (CCR) was highest for corrected RPP values (88.1%) with NDIP (85.1%) and lowest for LS-SVM (76.7%) (68). A high level of CCR is desirable, but at the same time, error parameters also need to be evaluated, which was not mentioned in this study.

Many studies rely on SSC and DM models that use PLS to predict maturity. Taking the spectral analysis to the next level requires the use of different applications, such as those that are used in robotics or that can process video. One study of Fuji apples applied NIR spectral analysis to video frames to detect maturation (72). Rather than the use of quantitative data and correlation, expert panelists were chosen to determine maturity. The processing computer used in the study had "entry-level" specifications by 2020 standards, thus demonstrating the affordability of the analysis, in contrast to "high-end" specifications. The classification model used was ANN combined with simulated annealing (SA). The SA sends out different vectors to ANN and finds the most suitable ANN structure (72). The method performed well, with accuracy rates above 90%, using the wavelength regions of 535–560 nm, 835–855 nm, or 950–975 nm (72).

Chen et al. (70) determined Shatian pomelo maturity by correlating color and maturity levels. Rather than conventional preprocessing methods, the multiplicative effect correction (MOPLEC), which corrects the change in optical path length, was used (70). This technique performed better in comparison to Savitzky-Golay smoothing and increased the signal-to-noise ratio (70). The preprocessing was followed by a moving window partial least square (MWPLS), which found a high signal wavelength combination for PLS modeling (70). MWPLS varies the location and number of adjacent wavelengths to find the optimum combination for PLS modeling (70). The study compared the raw data, SG-processed, and MOPLEC-processed

data and found that R_v was highest for MOPLEC-processed data, showing on-par or better values than SG-processed data (70). The new techniques were able to perform better overall in predicting the maturity of pomelo from color. Therefore, these techniques need to be explored for other fruits to determine whether color prediction can be improved for them as well.

Damage detection

Fresh and minimally processed food commodities are in high demand these days, which raises the stakes on delivering commodities with minimal damage. Well sorted and less-damaged products increase consumer satisfaction, too. There is, however, a great deal of variation among horticultural products: Highly perishable commodities such as tomatoes can show the effects of bruises or damage within 48 h (75), whereas apples can take more than 50 days (8) to show any signs of damage on the outer surface. Harvested commodities should be promptly evaluated for damage to receive the commodity and then convert to any value-added products. Due to the short evaluation period, many food researchers and industries use non-destructive techniques such as NIR spectroscopy that can detect damage without any effects on the products (Table 5). Defects such as internal bruising in blueberries can be detected as soon as 30 min after mechanical impact, with an R^2 of 0.7 (76). Because we can estimate the spoilage period and extent of damage, characterizing spectral signatures so as to detect damage or abnormalities could improve crop utilization (75).

Damage predictions are conducted by spectral examination of the outer skin, known as the epidermis. Because the epidermis is the "interface" of the fruit structure, any unfavorable conditions can cause small changes (75) that may not be apparent in the 500–900 nm range (i.e., visible to the naked eye) (79). These changes can, however, be detected using NIR spectroscopy. However, some changes, such as oxidative browning, can be detected by the human eye due to the significant time lag involved (76).

In terms of collecting input for spectral processing, various configurations have been studied, viz. taking measurements in 90-degree rotations (10) or 120-degree rotations (13), contacting the epidermis (8), using 360-degree measurements (9), and taking multiple measurements at different points and averaging the values (75). In a comparison of apples held at different orientations for measurement, orienting apples along the equatorial region showed the highest potential for defect detection (9). This may be because this orientation exposes a greater amount of curved surface area, which allows for more sensitive measurements. However, the equatorial alignment yielded the worst results in detecting chilling injury in kiwifruit; in this context, results were better with alignment at the stylar end, likely because chilling symptoms begin at the end of the fruit (10). It was observed during a hyperspectral analysis of

TABLE 5 Summary of NIRS applications for damage detection in agricultural products.

Agricultural product	Spectral range	Software package	Number of samples	Accuracy	Findings	References
Peach	400–2,500 nm 325–1,100 nm 930–2,548 nm	ENVI v. 4.6, Istra Optics	200	Bruise detection = 96.5, Sound sample detection = 97.5	SW-NIR was found more suitable than LW-NIR. Improved watershed segment algorithm used.	(61)
Blueberry	950–1,650 nm	LabVIEW, MATLAB 2016b, ENVI v. 4.7	320	70.8%–100%	Prediction accuracy increased with time. Calyx side bruise is difficult to detect.	(76)
Apple	400–1,100 nm	MATLAB, Neural Network Toolbox v. 4	550	Correlation of prediction = 0.87, SEP = 5.8N	Back-propagation neural network, combined with input ratios of scattering profile, predicted fruit firmness and SSC.	(77)
Olive	1,100–2,300 nm	R v. 3.1.3 SNAP! V. 2.04	744	96%	QDA was found better than LDA.	(78)
Apple	550–1,650 nm	MATLAB R2017a, PLS Toolbox v. 8.6	430	92.9%	The influence of different positioning on spectra acquisition was studied. The equatorial orientation was found best.	(9)
Kivi	400–1,100 nm	MATLAB R2018a, PLS toolbox v. 8.7	129		Water-soaked tissues have lower potential for damage detection after the SNV process.	(10)
Coconut	900–2,500 nm	OPUS v. 6.5, Unscrambler v. 9.8	202	94.03%	The NIR model performed better than the acoustic method for detecting cracks in coconut shells.	(13)
Apples	350–1,100 nm	Unscrambler v. 10.5.1, MATLAB R2017b, PLS Toolbox	393	$R^2 = 0.59$	Sun scaling causes changes in chlorophyll content.	(8)
Cherry	350–2,500 nm	RS3, MATLAB 2011a, ENVI v. 4.6, ViewSpec Pro v. 6.2.0, Origin6 SR0, Unscrambler X10.1	300	93.3%	Color, firmness, and SSC were consistent with the Vis-NIR reflectance. LS-SVM combined with SPA detected bruises within the sample.	(79)
Peach	300–1,150 nm	OMNIC v. 8.2, R Studio, SPSS v. 22.0, OriginPro 2017	840	$R_p = 0.71$ – 0.92 , RMSEP = 0.17–20.34	Good correlation between physiological indicators and absorption spectra. GA-PLS performed the best.	(80)
Tomato	5,555.56– 11,111.11 nm	MATLAB 2016		83–97%	PCA-LDA identified <i>G.</i> <i>canadum</i> infection.	(73)

SW-NIR, Shortwave near infrared; LW-NIR, Longwave near infrared; SSC, Soluble solid content; QDA, Quadratic discriminant analysis; LDA, Linear discriminant analysis; SNV, Standard normal variate; LS-SVM, Least square support vector machine; GA-PLS, Genetic algorithm partial least square; PCA-LDA, Principal component analysis and linear discriminant analysis.

blueberries that the calyx end was misclassified as bruised; these data needed to be excluded to reduce the false positive rate and reduce the RMSEP of 0.1 to 0.13 (76). Therefore, the design of NIR systems to measure fruit characteristics should take into account both the defect to be detected and the algorithm used for analysis.

A 6-fiber individual measurement system performed better than the conventional single-point measurement system (9), as accuracy increased from 0.8% to 1.7% when average spectra were used. There is no single method best suited for measurement, because all measurements are post-processed using different methods. Re-orienting can take up a large amount of time needed for developing a commercial online system, which is not suitable. During the training, care should be taken to avoid any outliers, which can affect the classification accuracy by causing overfitting problems. Outliers, in the context of damage detection, refers to commodities that are severely damaged. These cause the PCA and LDA algorithms to underperform. An analysis of sunscald apples followed this procedure to avoid the highly damaged ones, thus increasing the classification accuracy (8). In practical terms, this is possible because severe defects are visible to the naked eye.

An experiment conducted to detect cracked coconut shells found that a PCA plot itself can be used to find the outliers in cases where it is impossible to observe the shell inside the husk (13). The PCA analysis will yield more relevant results once it is run only with effective wavelengths and with outliers eliminated. A PCA analysis of bruising on peaches, with effective wavelengths of 781, 816, 840, 945, and 1,000 nm, yielded a PC4 with high-contrast bruised area (60). When selecting the optimal wavelength, the successive projection algorithm (SPA) follows an iterative process, adding one more wavelength with each step, which reduces the complexity of dealing with the whole spectrum (79). During a bruise analysis of cherry, SPA with 3 variables had an accuracy rate of 96.6%, while the full spectra with 2 variables had a 96% accuracy rate with less computational load (79). A lower computational load, i.e., a shorter analysis time, is essential for the design of an inline system. During the bruising analysis, it was found that even slight damage to the control samples can cause a partiality in the classification model (76). One more caveat is the need to use an independent set of crops from a different region or season. The influences of spatial and temporal variability are still not accounted for by the usual classification algorithms, due to high biological variability. As seen in the detection of sunscald in apples from different years, the RMSEP for prediction was three times that of the calibration (8). The variability caused by biological variation, temperature fluctuation, and measurement positions can be accounted for, to an extent, by using classification models such as least square support vector machines (LS-SVM) (76).

The classification of defect detection is usually done subjectively, i.e., not quantitatively, meaning that human error can occur (10). When the spectra are used for chemometric or

qualitative measurement, they should be immediately assessed after the spectral measurement to reduce errors caused by the increase of time (80).

The detection regions are explored by comparing the sound and damaged fruits using their spectral reflectance curves. These depict the reflectance vs. wavelength plot for both samples. Due to the changes caused by damage, there will be significant variation in some parts of the plot between sound and damaged produce, and these differing bands are used for classification. As far as the classification is considered, sound fruits are more homogeneous, and therefore easier to classify, than damaged ones (61). A consistent variation in the spectral sign is desirable for accurate separation (52). This approach also reduces redundant information and the time needed for processing. For SW-NIR and visible NIR, wavebands are selected instead of wave points, to reduce the complexity (80).

In an analysis of bruising on peaches, it was found that SW-NIR (7,801–1,000 nm) was more efficient than LW-NIR (1,000–2,500 nm) due to better contrast difference in the bruised area (60). During the analysis of HIS, wave points and regions of interest (ROI) are selected for analysis (61, 76), and they also provide the spatial distribution of the bruises. Upon correlating the significant effect of the epidermis and spectral curve on research, we observe that pigments such as carotenoids and chlorophyll bands provide promising regions of variation, viz. at around 670 nm for the chlorophyll band (8, 10, 52), 550 nm for the anthocyanin band (79), and 740 nm for the carotenoid band (8). Defects in fruits with lower chlorophyll content may not show significant chlorophyll band variation, as is the case with internal browning in apples (60). Evaluating specific chemical bonds and determining the wave points is significantly easier than characterizing TSS or SSC, each of which contains several compounds (80) and is thus tedious to identify on a spectral reflectance plot. The wavelength range of SSC for cherries was 900–2,500 nm, a large waveband that needs further analysis (79).

One study of bruising in blueberries followed the analysis of the cut piece's pixels to evaluate the degree of bruising (76), which was more accurate than a subjective human evaluation. Supplementary techniques such as inverse adding doubling (IAD) can also be conducted on the optical properties of water-soaked tissues (10). The water absorption peaks were found to be at the 970 nm and 1,200 nm regions. A bruise capable of causing cell membrane rupture will undoubtedly differ significantly in this region (76).

The traditional methods followed the PCA and PLS methods, which project all variables to reduce dimensions and increase separability, respectively. During PCA analysis, the first few PC are selected for classification due to their feature-rich plots. Nevertheless, an analysis of bruising on peaches found that PC4 had the best bruise distinction, with retention of effective wavelengths (61). Therefore, while performing a PCA, the resultant plot should be observed to check whether

it fits the purpose. To further enhance the efficiency, non-linear optimizations should be carried out, such as the genetic algorithm (GA), where the feature selection follows the natural selection procedure (80). One study of an online system suggested that the band ratio method can be used as a real-time and cost-effective method, with accuracy rates above 90% (76).

Detection of microbial/fungus contamination

Horticultural crops are prone to microbial contamination. Here, we explore early detection of microbial and fungal contamination using NIRS. Early non-destructive detection within 24 h, in the case of *Botrytis cinerea* and *Colletotrichum acutatum*, can stop cross-contamination and economic loss (81). Studies conducted by Matulaprungsan et al. (82) on cabbage contamination by *E. coli* and *S. typhimurium* created artificial conditions and inoculated samples of shredded cabbage for various time periods to explore the potential of NIR analysis (Table 6). During spectral analysis, this method showed a spectral shift due to the leaking from cabbage cells and the growth of bacteria (82). This shifting behavior was treated using the SG-derivative method (82), followed by tomato pathogen analysis (83). Spectral analysis revealed that the shredded samples had better separation between inoculation times than the ground samples (82), which makes the former the method of choice for analysis. Shredded cabbage, therefore, resulted in a higher *t*-value ($r = 0.91\text{--}0.95$), indicating reliable prediction (82). *E. coli* detection in lettuce also resulted in an accuracy rate of 100% in the validation set using NIR spectroscopy and partial least-squares discriminant analysis (PLS-DA) (36). Various concentrations of *E. coli* were able to be classified using five different techniques, viz. PLS-DA, SVM, PCA, hierarchical cluster analysis (HCA), and soft independent modeling by class analogy (SIMCA). PCA and HCA trended toward a coarser classification, whereas SIMCA and SVM produced finer classifications into subgroups. The *E. coli* analysis was carried out in baby spinach using the PLS method (35). For contamination studies, outliers need to be eliminated using feature selection techniques such as PCA. In studies carried out to detect *E. coli* and *Z. rouxii*, PCA used to remove outliers (6, 35). In the case of *E. coli* detection in spinach, Q-residual-Hotelling's T2 plot was used to find out the outliers (35). *E. coli* was detected using PLS-DA supervised learning in the 450–994 nm range, with a 100% accuracy rate in the prediction set (35). The model could not predict unsafe samples in the early stages. Therefore, when the cell concentration reached 6.67 log CFU/mL, it was detected as unsafe during statistical analysis (35).

A study on tomato pathogen analysis achieved a classification accuracy rate of 99.3% on the first and second day

using PCA followed by an SVM classifier (83). Although higher classification rates were achieved as the mixing of samples began, the SVM classification could not handle multiple factors affecting the same. The high classification rates were able to detect *F. oxysporum* and *R. solani* fungi, as well as bacteria such as *B. atrophaeus* and *P. aeruginosa* (83).

The analysis of liquid samples created problems due to transparency (82). NIR essentially works through photons and their response to corresponding compounds, so as a solution becomes clear, it becomes non-homogeneous and therefore changes the spectral response. In this context, it is best to use the SW-NIR technique for analysis due to its higher penetration power and lesser interference from water bands (83). However, the study on microbial contamination in lettuce suggested that the water band corresponding to 970 nm O-H stretching needs to be accounted for (36). The differences in microbial cell count can cause molecular vibration patterns (36). In studies of solid lettuce samples, preprocessing techniques such as standard normal variate (SNV) and multiplicative scatter correction (MSC) were used to remove scatter effects (36). Elimination of end bands and spectral resolution enhancement can be performed in order to bring greater focus to prominent wavelengths (36). Preprocessing is carried out to avoid noise and the inclusion of unwanted information that does not add to prediction accuracy. A ratio of prediction of deviation between 1.5 and 2.0 indicates well-calibrated models (82).

Fungal infection caused by *B. cinerea* and *C. acutatum* in strawberry was analyzed using spectroscopy. The analysis extracted regions of interest (ROI) in the spectral range of 450–2,500 nm with 3500 pixels (81). The data were analyzed using four different classification methods, viz. backpropagation neural network (BPNN), random forest (RF), naïve Bayes (NB), and support vector machine (SVM) (81). The classification accuracy decreases once the fungal activity is reduced, which was observed after 4 days in *C. acutatum*-infected samples. However, this technology has its challenges, viz. problems with reproducibility, recovery, and the negative effects of humidity and temperature (81). The insight into the influencing factors was a new finding, and the results of this study can thus be used to improve techniques through control of the same.

Yeast species tend to cause deterioration of high-sugar products, particularly fruit products. The yeast species *Zygosaccharomyces rouxii*, which displays this behavior, was analyzed by Niu et al. (6) using NIR to perform non-destructive detection. During preprocessing, it was found that either SG smoothing or direct orthogonal signal correction (DOC) can be used (6). Terminologies such as the limit of detection (LOD) and limit of quantification (LOQ) were introduced in this study, and they varied along with various preprocessing and classification algorithms (6). These terms can be treated as the threshold values for performing detection using the model. The model was able to detect *Z. rouxii* according to the existing

TABLE 6 Summary of NIRS applications for the detection of target microbial/fungus contamination in agricultural products.

Agricultural product	Spectral range	Software package	Number of samples	Accuracy	Findings	References
Tomato	550–1,100 nm	Unscrambler v. 10.3, SpectraSuite	45	74–90%	<i>F. oxysporum</i> f. sp. <i>lycopersici</i> , <i>R. solani</i> , <i>B. acillanatrophaea</i> , and <i>P. aeruginosa</i> infections were detected.	(83)
Strawberry	400–1,000 nm	Unscrambler X v. 10.1, SpectralDAQ v. 2.1 for STATISTICA 10	2,700	97%	<i>B. cinerea</i> and <i>Colletotrichum acutatum</i> infections were detected. The BNN model exhibited the highest predictive accuracy.	(81)
Kivi	833–2,500 nm	OPUS v. 5.5, MATLAB 2012a, Libsvm v. 3.20	352	$R^2 = 0.961–0.999$	<i>Z. rouxi</i> , <i>Hanseniaspora uvarum</i> , and <i>C. tropicalis</i> infections were detected. The SVM model was on par with the plate counting method.	(6)
Lettuce	350–1,100 nm	SpectraWiz, Unscrambler X10.3	200	87.1–89.39%	<i>E. coli</i> ATCC infection was detected. SIMCA and SVM outperformed HCA, PCA. <i>E. coli</i> content varied with the chemical compositions, creating non-linear relationships.	(36)
Cabbage	700–1,100 nm	CA Maker, Unscrambler	20 g	$R = 0.47–0.91$, SECV = 0.45–1.17	<i>E. coli</i> , <i>S. typhimurium</i> infections were detected. Shredded leaves were more suitable for detection. Not a directly non-destructive approach.	(82)

BPNN, Backpropagation neural network; SVM, Support vector machine; SIMCA, Soft independent modeling by class analogy; HCA, Hierarchical cluster analysis; PCA, Principal component analysis.

standards, with an accuracy rate >85% for both the calibration and prediction models (6).

Challenges and future prospects

NIR is one of the most researched non-destructive methods in the food processing industry due to its fast-paced analysis, minimal need for sample preparation, cost-effectiveness, and non-destructive nature (84). However, like all technological advancements, NIRS also comes with some disadvantages to be tackled, viz. standard error in the laboratory (SEL) causing standard error in prediction (SEP); variations in orientation (29); and performance issues due to biological variation (8). Apart from these challenges, which are related to spectral measurements, regression tools, machine learning, and ANN can also cause problems such as overfitting (53), low correlation values (making the method a primary sorter only), pattern

recognition problems (39), and a lack of proper knowledge for choosing the best techniques among the plethora available.

Because these problems have been around since the beginning, some solutions have also been found for them. These include testing and optimizing orientation for parameters, like the testing done with SSC determination that determined that keeping the stem-calyx axis vertical, with stem upward, was best due to the presence of SSC in the stem region (29). During model preparation, the conditions need to be perfect to avoid SEL and SEP, because a single error can cause the model to underperform.

Biological variation remains a challenge, especially when testing with another sample of the same variety from the next season. This is primarily because machine learning works on pattern recognition, and every single year, a large number of biological variables change, which alters the parameters of the product. The way to solve this problem is to have a large training set from various seasons, which is, however, practically impossible.

In order to solve this issue, the techniques of the preprocessing and prediction phases need to be modified so as to adapt the pattern; otherwise, every year, the model must be taught from scratch. This is not possible, since perfect conditions would be needed to recreate a model with minimum errors. Neural network-related issues are being tackled using various techniques such as BPNN, GRNN, and PSO, which performed better when it was left individually (39). Although these methods seem to solve the problem to a certain extent, they can increase the computational load. Cloud computing-based models from various parts of the world may become the next solution, since most of the advanced systems are built around such a setup.

However, machine learning aided by ANN seems to be the foremost approach for NIR analysis, since it deals with biological commodities. The extensive variations in agricultural products can be accounted for using neural networks and pattern recognition algorithms. Predetermined algebraic equations and models created on exact mathematics can be expected to fail in this scenario due to their inability to account for complex phenomena underlying the products' life cycles. Whereas, NIRS got its start as a series of offline systems, today's researchers can create online rapid analysis prototypes, demonstrating the current pace of technology.

The research matrix shows that a significant portion of the studies to date have been conducted on apples, and a huge list of agricultural crops are waiting to be explored. Compact handheld systems have been shown to perform satisfactorily with NIRS, spurring hopes for the next era, when these technologies will hopefully be more widely available and more affordable. Many contemporary smartphones come with IR blasters, light detection and ranging (LIDAR) hardware, and many other sensors. Soon, we all might have NIRS scanners with cloud-based models that can also account for biological variations.

Conclusion

Spectroscopic techniques can create unique signatures for chemicals due to the particular behaviors of various molecular levels. This allows the quantitative and qualitative analysis of agricultural crops and the selection of those products with the most desirable characteristics. Spectrometer data are used to perform regression analysis and machine learning, by which the evaluation of agriculture crops is conducted. What began as offline test measurement is now capable of performing rapid online measurements without destroying the samples. Advances in computation have greatly improved the technology without much economic expenditure. Current methods involve the creation of hybrid models, which can address biological variations to a certain extent by taking the best parts of various techniques and combining them. Apart from this advancement, challenges remain on the technical side, i.e., the need to address the variation of products, data accusation, and model

optimization. There is still an urgent need for a compact, non-destructive technology that can characterize a wide variety of horticultural commodities.

Author contributions

RP: conceptualization, methodology, data curation, formal analysis, and writing—review and editing. VP: visualization, writing original draft, and writing—review and editing. AK: visualization, methodology, writing original draft, investigation, writing—review and editing, supervision, and funding acquisition. MM: visualization, methodology, writing original draft, investigation, writing—review and editing, and funding acquisition. AR and MT: resources, writing—review and editing, and funding acquisition. AKh: conceptualization, methodology, validation, resources, and writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work is based upon the work from COST Action 18101 SOURDOMICS – Sourdough biotechnology network toward novel, healthier and sustainable food and bioprocesses (<https://sourdomics.com/>; <https://www.cost.eu/actions/CA18101/>), where the authors (AR, MT, and AKh) are members of the Working Group (1). SOURDOMICS supported by COST (European Cooperation in Science and Technology). COST is a funding agency for research and innovation networks. COST Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers – thus boosting their research, career and innovation. We would also like to acknowledge the funding received by AR from the Romanian National Authority for Scientific Research and Innovation, CNCS-UEFISCDI, project number PN-III-P2-2.1-PED-2019-1723 and PFE 14, within PNCDI III. RP was highly grateful to ICAR-AICRP on PHET for financial assistance.

Acknowledgments

We would like to acknowledge the funding received by AR from the Romanian National Authority for Scientific Research and Innovation, CNCS—UEFISCDI, project number PN-III-P2-2.1-PED-2019-1723 and PFE 14, within PNCDI III.

Conflict of interest

Author MT was employed by Centiv.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Madanayake NH, Hosain A, Adhisorriya NM. Nanobiotechnology for agricultural sustainability, and food and environmental safety. *Q Assurance Safety Crops Foods*. (2021) 13:20–36. doi: 10.15586/qa.v13i1.838
- Yadav A, Kumar N, Upadhyay A, Singh A, Anurag RK, Pandiselvam R. Effect of mango kernel seed starch-based active edible coating functionalized with lemongrass essential oil on the shelf-life of guava fruit. *Q Assurance Safety Crops Foods*. (2022) 14:103–115. doi: 10.15586/qa.v14i3.1094
- Beegum PS, Pandiselvam R, Ramesh SV, Thube SH, Pandian TP, Khanashyam AC, et al. A critical appraisal on the antimicrobial, oral protective, and anti-diabetic functions of coconut and its derivatives. *Q Assurance Safety Crops Foods*. (2022) 14:86–100. doi: 10.15586/qa.v14i2.1040
- Kaaya B, Pandiselvam R, Mohammed M, Dakshayani R, Kothakota A, Ramesh SV, et al. Application of infrared spectroscopy techniques for the assessment of quality and safety in spices: a review. *Appl Spectrosc Rev*. (2020) 55:593–611. doi: 10.1080/05704928.2020.1713801
- Munawar AA, von Hörsten D, Wegener JK, Pawelzik E, Mörlein D. Rapid and non-destructive prediction of mango quality attributes using Fourier transform near infrared spectroscopy and chemometrics. *Eng Agric Environ Food*. (2016) 9:208–15. doi: 10.1016/j.eaef.2015.12.004
- Niu C, Guo H, Wei J, Sajid M, Yuan Y, Yue T. Fourier transform near-infrared spectroscopy and chemometrics to predict *Zygosaccharomyces rouxii* in apple and kiwi fruit juices. *J Food Prot*. (2018) 81:1379–85. doi: 10.4315/0362-028X.JFP-17-512
- Yan H, Xu YC, Siesler HW, Han BX, Zhang GZ. Hand-Held Near-Infrared Spectroscopy for authentication of figendous and quantitative analysis of mulberry fruits. *Front Plant Sci*. (2019) 10:1–15. doi: 10.3389/fpls.2019.01548
- Grandón S, Sanchez-Cortez J, Torres CA. Prediction models for sunscald on apples (*Malus domestica* Borkh.) cv. Granny Smith using Vis-NIR reflectance. *Postharvest Biol Technol*. (2019) 151:36–44. doi: 10.1016/j.postharvbio.2019.01.012
- Huang Y, Lu R, Chen K. Detection of internal defect of apples by a multichannel Vis/NIR spectroscopic system. *Postharvest Biol Technol*. (2020) 161:111065. doi: 10.1016/j.postharvbio.2019.111065
- Wang Z, Künsemeyer R, McGlone A, Burdon J. Potential of Vis-NIR spectroscopy for detection of chilling injury in kiwifruit. *Postharvest Biol Technol*. (2020) 164:111160. doi: 10.1016/j.postharvbio.2020.111160
- Ncama K, Tesfay SZ, Opara UL, Fawole OA, Magwaza LS. Non-destructive prediction of 'Valencia' orange (*Citrus sinensis*) and 'Star Ruby' grapefruit (*Citrus × paradisi* Macfad) internal quality parameters using Vis/NIRS. *Acta Hort*. (2018) 1194:1119–25. doi: 10.17660/ActaHortic.2018.1194.160
- Tan J, Li MF. Rapid and nondestructive identification of Belgian and Netherlands trappist beers by front-face synchronous fluorescence spectroscopy coupled with multiple statistical analysis. *Q Assurance Safety Crops Foods*. (2021) 13:83–92. doi: 10.15586/qa.v13i1.839
- Noyyitak S, Imaabai W, Noknoi W, Kanojue S, Terdwingworakul A, Koberi H. Detection of cracked shell in intact aromatic young coconut using near infrared spectroscopy and acoustic response methods. *J Food Meas Charact*. (2019) 13:1991–5. doi: 10.1007/s11694-019-00119-2
- Coriás V, Blanco J, Aleixos N, Cabero S, Talem P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends Food Sci Tech*. (2019) 85:338–48. doi: 10.1016/j.tifs.2019.01.015
- Purser A, Cozzolino D. How fishy is your fish? Authentication, provenance and traceability in fish and seafood by means of vibrational spectroscopy. *Appl Sci*. (2020) 10:1–13. doi: 10.3390/app10124150
- Pu YF, O'Donnell C, Tobin JT, O'Shea N. Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing. *Int Dairy J*. (2020) 103:104623. doi: 10.1016/j.idairyj.2019.104623
- Dixit Y, Casado-Gavaldá MP, Cama-Moncunill R, Cama-Moncunill X, Markiewicz-Keszycka M, Callen P, et al. Developments and challenges in online NIR spectroscopy for meat processing. *Compr Rev Food Sci Food Saf*. (2017) 16:1172–87. doi: 10.1111/1541-4337.12295
- Lan W, Jaillon B, Leca A, Renard CMGC, Bureau S. A new application of NIR spectroscopy to describe and predict puros quality from the non-destructive apple measurements. *Food Chem*. (2020) 310:125944. doi: 10.1016/j.foodchem.2019.125944
- Abasi S, Minsai S, Jamshidi B, Fathi D, Khoshbaghata MH. Rapid measurement of apple quality parameters using wavelet de-noising transform with Vis/NIR analysis. *Sci Hort*. (2019) 252:7–13. doi: 10.1016/j.scienta.2019.02.085
- Theerajumpol P, Wongsowasakun K, Mummamtee N, Wongsuwan S, Krongchai C, Changruae V, et al. Non-destructive identification and estimation of granulation in 'lai Nam Pung' tangerine fruit using near infrared spectroscopy and chemometrics. *Postharvest Biol Technol*. (2019) 153:13–20. doi: 10.1016/j.postharvbio.2019.03.009
- Čaerha V, Fodor M, Kovács S, Tóth M. Potential of fourier transformed near-infrared (FT-NIR) spectroscopy for rapid analysis of elderberry (*Sambucus nigra* L.) fruits. *Czech J Food Sci*. (2019) 37:21–8. doi: 10.17221/144/2018-CJFS
- Ntarambirwe JFI, Nieuwoudt HH, Perold WJ, Opara UL. Non-destructive measurement of internal quality of apple fruit by a contactless NIR spectrometer with genetic algorithm model optimization. *Sci African*. (2019) 3:e0051. doi: 10.1016/j.sciaf.2019.e0051
- Hu R, Zhang L, Yu Z, Zhai Z, Zhang R. Optimization of soluble solids content prediction models in 'Hami' melons by means of Vis-NIR spectroscopy and chemometric tools. *Infrared Phys Technol*. (2019) 102:102999. doi: 10.1016/j.infrared.2019.102999
- Kanchanomai C, Naphrom D, Ohashi S, Nakano K, Theerajumpol P, Manivara P. Nondestructive determination of quality management in table grapes using Near Infrared Spectroscopy (NIRS) technique. *ETP Int J Food Eng*. (2019) 5:63–7. doi: 10.18178/ijfe.5.1.63-67
- Amuah CLY, Teye E, Lamptey FP, Nyandey K, Opolo-Anaah J, Adjuarming POW. Feasibility study of the use of handheld NIR spectrometer for simultaneous authentication and quantification of quality parameters in intact pineapple fruits. *J Spectrosc*. (2019) 2019:e5975461. doi: 10.1155/2019/5975461
- Kashivagi A, Kaminatti R, Nakayama S, Kishimoto N. Nondestructive prediction of oil yield through direct Near Infrared Spectroscopy measurements of fresh olives. In 2019 IEEE International Symposium on Objectives and Electronic Nose (ISOEN). IEEE (2019). doi: 10.1109/ISOEN.2019.8823467
- Khodabakhshian R, Emadi R, Khojastehpour M, Golzarian MR. A comparative study of reflectance and transmittance modes of Vis/NIR spectroscopy used in determining internal quality attributes in pomegranate fruits. *J Food Meas Charact*. (2019) 13:3130–9. doi: 10.1007/s11694-019-00255-z
- Hasnah Ar N, Purwanti YA, Budiatra IW, Sobri S. Prediction of soluble solid content, vitamin C total acid and firmness in astriniger persimmon (*Diospyros kaki* L.) cv. Rendu using NIR spectroscopy. *JOP Conf Ser Mater Sci Eng*. (2019) 537:e012086. doi: 10.1088/1757-899X/557/1/012086
- Xia Y, Huang W, Fan S, Li J, Chen L. Effect of spectral measurement orientation on online prediction of soluble solids content of apple using Vis/NIR diffuse reflectance. *Infrared Phys Technol*. (2019) 97:467–77. doi: 10.1016/j.infrared.2019.01.012
- Huang Y, Lu R, Chen K. Assessment of tomato soluble solids content and pH by spatially-resolved and conventional Vis/NIR spectroscopy. *J Food Eng*. (2018) 236:19–28. doi: 10.1016/j.jfoodeng.2018.05.008
- Azadshahraki F, Jamshidi B, Rasooli Sharabiani V. Non-destructive determination of vitamin C and lycopene contents of intact cv. Newton tomatoes using NIR spectroscopy. *Yuz El Univ J Agric Sci*. (2018) 28:389–97. doi: 10.29133/yypubd.423458
- Ncama K, Opara UL, Tesfay SZ, Fawole OA, Magwaza LS. Application of Vis/NIR spectroscopy for predicting sweetness and flavour parameters of 'Valencia'

- orange (*Citrus sinensis*) and 'Star Ruby' grapefruit (*Citrus x paradisi* Macfad). *J Food Eng.* (2017) 193:86–94. doi: 10.1016/j.jfoodeng.2016.08.015
33. Lan W, Bureau S, Chen S, Leca A, Renard CMGC. Visible, near- and mid-infrared spectroscopy coupled with an innovative chemometric strategy to control apple purity quality. *Food Cont.* (2021) 120:107546. doi: 10.1016/j.foodcont.2020.107546
34. Sans S, Ferré J, Boquié R, Sabaté J, Casals J, Simó J. Determination of chemical properties in 'calçot' (*Allium cepa* L.) by near infrared spectroscopy and multivariate calibration. *Food Chem.* (2018) 262:178–83. doi: 10.1016/j.foodchem.2018.04.102
35. Rahi S, Mobli H, Jamshidi B, Azizi A, Sharifi M. Visible/near-infrared spectroscopy as a novel technology for nondestructive detection of *Escherichia coli* ATCC 8739 in lettuce samples. (2019) 24–6. doi: 10.33422/worldvet.2019.10.285
36. Rahi S, Mobli H, Jamshidi R, Azizi A, Sharifi M. Different supervised and unsupervised classification approaches based on visible/near infrared spectral analysis for discrimination of microbial contaminated lettuce samples: Case study on *E. coli* ATCC. *Infrared Phys Technol.* (2020) 108:103355. doi: 10.1016/j.infrared.2020.103355
37. Jainwal P, Jha SN, Bharadwaj R. Non-destructive prediction of quality of intact banana using spectroscopy. *Sci Hortic.* (2012) 135:14–22. doi: 10.1016/j.scienta.2011.11.021
38. Kanchanmai C, Ohashi S, Naphorn D, Nemoto W, Maniwaru P, Nakano K. Non-destructive analysis of Japanese table grape qualities using near-infrared spectroscopy. *Hortic Environ Biotechnol.* (2020) 61:725–33. doi: 10.1007/s13580-020-00256-4
39. Wu Y, Li L, Liu L, Liu Y. Nondestructive measurement of internal quality attributes of apple fruit by using NIR spectroscopy. *Multimed Tools Appl.* (2019) 78:4179–95. doi: 10.1007/s11042-017-5388-0
40. Torres I, Sánchez MT, Entrenas JA, Garrido-Varo A, Pérez-Marin D. Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors. *Postharvest Biol Technol.* (2019) 154:21–30. doi: 10.1016/j.postharvbio.2019.04.015
41. Camps C, Gilli C. Prediction of local and global tomato texture and quality by FT-NIR spectroscopy and chemometric. *Eur J Hortic. Sci.* (2017) 82:126–33. doi: 10.11766/eJHS.2017/82.3.2
42. Sharma S, Sarasondhon P. Feasibility on using NIR spectroscopy for the measurement of the textural parameters in mango. *JOP Conf Ser Earth Environ Sci.* (2019) 301:s012064. doi: 10.1088/1755-1315/301/1/012064
43. Mohammadi-Moghaddam T, Razavi SMA, Sargarnia A, Taghizadeh M. Predicting the moisture content and textural characteristics of roasted pistachio kernels using Vis/NIR reflectance spectroscopy and PLSR analysis. *J Food Mass Charact.* (2018) 12:346–55. doi: 10.1007/s11094-017-9646-7
44. Sánchez MT, Entrenas JA, Torres I, Vega M, Pérez-Marin D. Monitoring texture and other quality parameters in spinach plants using NIR spectroscopy. *Comput Electron Agric.* (2018) 155:446–52. doi: 10.1016/j.compag.2018.11.004
45. Feng L, Zhang M, Adhikari B, Guo Z. Nondestructive detection of postharvest quality of cherry tomatoes using a portable NIR spectrometer and chemometric algorithms. *Food Anal Methods.* (2019) 12:914–25. doi: 10.1007/s12161-018-01429-9
46. Berardinelli A, Bonelli A, Tartagni M, Ragni L. Kivifruit flesh firmness determination by a NIR sensitive device and image multivariate data analyses. *Sensors Actuators A Phys.* (2019) 296:265–71. doi: 10.1016/j.sna.2019.07.027
47. Cirilli M, Bellincanto A, Urbani S, Servili M, Esposito S, Mencarelli F, et al. On-field monitoring of fruit ripening evolution and quality parameters in olive mutants using a portable NIR-AOTF device. *Food Chem.* (2016) 199:96–104. doi: 10.1016/j.foodchem.2015.11.129
48. Yu X, Lu H, Wu D. 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.* (2018) 141:39–49. doi: 10.1016/j.postharvbio.2018.02.013
49. Labaky P, Grosmaire L, Ricci J, Wisniewski C, Lutka N, Dahdoud L. Innovative non-destructive sorting technique for juicy stone fruits textural properties of fresh mangos and pears. *Food Bioprod Process.* (2020) 123:188–98. doi: 10.1016/j.fbp.2020.06.013
50. Cortés V, Cuñero S, Blasco J, Aleixos N, Talens P. In-line application of visible and near-infrared diffuse reflectance spectroscopy to identify apple varieties. *Food Bioscience Tech.* (2019) 12:1021–30. doi: 10.1007/s11947-019-02268-0
51. Eizenstecken D, Stürz B, Robatscher P, Lorenz L, Zanella A, Oberhuber M. The potential of near infrared spectroscopy (NIRS) to trace apple origin: Study on different cultivars and orchard elevations. *Postharvest Biol Technol.* (2019) 147:123–31. doi: 10.1016/j.postharvbio.2018.08.019
52. Li X, Huang J, Xiong Y, Zhou J, Tan X, Zhang B. Determination of soluble solid content in multi-origin 'Fuji' apples by using FT-NIR spectroscopy and an origin discriminant strategy. *Comput Electron Agric.* (2018) 155:23–31. doi: 10.1016/j.compag.2018.10.003
53. Zhang Y, Nuck JF, Al Shofee Y, Watkins CB. Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy. *Postharvest Biol Technol.* (2019) 151:111–8. doi: 10.1016/j.postharvbio.2019.01.009
54. Sánchez MT, Torres I, de la Haba M, Chamorro A, Garrido-Varo A, Pérez-Marin D. Rapid, simultaneous, and in situ authentication and quality assessment of intact bell peppers using near-infrared spectroscopy technology. *J Sci Food Agric.* (2019) 99:1615–22. doi: 10.1002/jsfa.9342
55. Jha SN, Jainwal P, Narvaiah K, Kumar R, Sharma R, Gupta M, et al. Authentication of mango varieties using Near-Infrared Spectroscopy. *Agric Res.* (2013) 2:229–35. doi: 10.1007/s40003-013-0068-4
56. Li J, Zhang H, Zhao B, Wang Z, Jiang Y. Determination of SSC in pears by establishing the multi-cultivar models based on visible-NIR spectroscopy. *Infrared Phys Technol.* (2019) 102:103066. doi: 10.1016/j.infrared.2019.10.0066
57. Sa WH, Bakalis S, Sun DW. NIR/MIR spectroscopy in tandem with chemometrics for rapid identification and evaluation of potato variety and doneness degree. In 2019 ASABE Annual International Meeting. (2019). doi: 10.13031/aim.201900120
58. Wu X, Wu R, Sun J, Li M, Du H. Discrimination of apples using near infrared spectroscopy and sorting discriminant analysis. *Int J Food Prop.* (2016) 19:1016–28. doi: 10.1080/10942912.2014.971181
59. Zimmer M, Schneider J. Near-infrared diffuse reflectance spectroscopy for discriminating fruit and vegetable products preserved in glass containers. *Croat J Food Sci Technol.* (2019) 11:104–12. doi: 10.17508/CJFST.2019.11.1.15
60. Li C, Li L, Wu Y, Lu M, Yang Y, Li L. Apple variety identification using Near-Infrared Spectroscopy. *J Spectrosc.* (2018) 2018:1–6. doi: 10.1155/2018/6955197
61. Li J, Chen L, Huang W. Detection of early bruises on peaches (*Amygdalus persica* L.) using hyperspectral imaging coupled with improved watershed segmentation algorithm. *Postharvest Biol Technol.* (2018) 135:104–13. doi: 10.1016/j.postharvbio.2017.09.007
62. Wu G, Wang C. Investigating the effects of simulated transport vibration on tomato tissue damage based on vis/NIR spectroscopy. *Postharvest Biol Technol.* (2014) 98:41–7. doi: 10.1016/j.postharvbio.2014.06.016
63. dos Santos Neto JP, de Assis MWD, Casagrande IP, Cunha Junior LC, de Almeida Teixeira GH. Determination of 'Palmer' mango maturity indices using portable near infrared (VIS-NIR) spectrometer. *Postharvest Biol Technol.* (2017) 130:75–80. doi: 10.1016/j.postharvbio.2017.03.009
64. Alenzi MM, Shaqiq M, Alhadon AA, Albalal EM, Alhamdan AM, Selieman TH, et al. Non-destructive assessment of flesh firmness and dietary antioxidants of greenhouse-grown tomato (*Solanum lycopersicum* L.) at different fruit maturity stages. *Saudi J Biol Sci.* (2020) 27:2839–46. doi: 10.1016/j.sjbs.2020.07.004
65. Chen H, Qiao H, Xu L, Feng Q, Cai K. A Fuzzy optimization strategy for the implementation of RBF-LSSVR model in Vis-NIR analysis of pomelo maturity. *IEEE Trans Ind Informatics.* (2019) 15:5971–9. doi: 10.1109/TII.2019.2933582
66. Daniels AJ, Poblete-Echeverría C, Opara UL, Nieuwoad HH. Measuring internal maturity parameters non-invasively on intact table grape bunches using NIR spectroscopy. *Front Plant Sci.* (2019) 10:1–14. doi: 10.3389/fpls.2019.01517
67. Blakey RJ. Evaluation of avocado fruit maturity with a portable near-infrared spectrometer. *Postharvest Biol Technol.* (2016) 121:101–5. doi: 10.1016/j.postharvbio.2016.06.016
68. Jie D, Zhou W, Wei X. Nondestructive detection of maturity of watermelon by spectral characteristic using NIR diffuse transmittance technique. *Sci Hortic.* (2019) 257:108718. doi: 10.1016/j.scienta.2019.108718
69. Saputro AH, Juanyah SD, Handayani W. Banana (*Musa sp.*) maturity prediction system based on chlorophyll content using visible-NIR imaging. *Int Conf Signals Syst.* (2018) 2018:64–68. doi: 10.1109/ICSS/SYS.2018.8373569
70. Chen HZ, Xu LL, Tang GQ, Song QQ, Feng QX. Rapid detection of surface color of shatian pomelo using Vis-NIR Spectrometry for the identification of maturity. *Food Anal Met.* (2016) 9:192–201. doi: 10.1007/s12161-015-0188-5
71. Ribera-Fonseca A, Noferrini M, Rombolà AD. Non-destructive assessment of highbush blueberry fruit maturity parameters and anthocyanins by using a visible/near infrared (Vis/NIR) spectroscopy device: A preliminary approach. *J Soil Sci Plant Nutr.* (2016) 16:174–86. doi: 10.4067/S0718-95162016005000014
72. Pourdarbani R, Sabzi S, Kalantari D, Karimzadeh R, Ebeyji E, Arefbasi JI. Automatic non-destructive video estimation of maturation levels in Fuji apple (*Malus Malus pumila*) fruit in orchard based on colour (Vis) and spectral (NIR) data. *Biosyst Eng.* (2020) 195:136–51. doi: 10.1016/j.biosystemseng.2020.04.015

73. Kim JG, Park Y, Shin MH, Muner S, Lerud R, Michelsen C, et al. Application of NIR-Spectroscopy to predict the harvesting maturity, fruit ripening and storage ability of Ca-chitosan treated baby kiwifruit. *J Stored Prod Postharvest Res.* (2018) 9:44–53. doi: 10.5897/JSPPR2018.0257
74. Polinar YQ, Yapienco KE, Peralta EK, Agravante JU. Near-infrared spectroscopy for non-destructive prediction of maturity and eating quality of 'carabao' mango (*Mangifera indica* L.) fruit. *Agric Eng Int CIGR J.* (2019) 21:209–19.
75. Skolik P, McAlinh MR, Martin FL. ATR-FTIR spectroscopy non-destructively detects damage-induced sour rot infection in whole tomato fruit. *Plants.* (2019) 249:925–39. doi: 10.1007/s00425-018-3060-1
76. Fan S, Li C, Huang W, Chen L. Detection of blueberry internal bruising over time using NIR hyperspectral reflectance imaging with optimum wavelengths. *Postharvest Biol Technol.* (2017) 134:55–66. doi: 10.1016/j.postharvbio.2017.08.012
77. Lu R. Multispectral imaging for predicting firmness and soluble solids content of apple fruit. *Postharvest Biol Technol.* (2004) 31:147–57. doi: 10.1016/j.postharvbio.2003.06.006
78. Moschetti R, Huff RP, Monarca D, Cecchini M, Massantini R. Near-infrared spectroscopy for detection of hailstorm damage on olive fruit. *Postharvest Biol Technol.* (2016) 120:204–12. doi: 10.1016/j.postharvbio.2016.06.011
79. Shao Y, Xuan G, Hu Z, Gao Z, Liu L. Determination of the bruise degree for cherry using Vis-NIR reflection spectroscopy coupled with multivariate analysis. *PLoS ONE.* (2019) 14:1–13. doi: 10.1371/journal.pone.0222633
80. Du X, Long L X, Yu L Y, Zhou W, Heng, et al. Genetic algorithm optimized non-destructive prediction on property of mechanically injured peaches during postharvest storage by portable visible/shortwave near-infrared spectroscopy. *Sci Hortic.* (2019) 249:240–9. doi: 10.1016/j.scienta.2019.01.057
81. Sielinka A, Baranowski P, Zubik M, Mazurek W, Sosnowska B. Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. *Postharvest Biol Technol.* (2018) 139:115–26. doi: 10.1016/j.postharvbio.2018.01.018
82. Matulaprunsan B, Wong-Aree C, Perchaiya P, Maniwar P, Karlayanurat S, Ohaabi S, et al. Feasibility of determination of foodborne microbe contamination of fresh-cut shredded cabbage using SW-NIR. *Agri Eng.* (2019) 1:246–56. doi: 10.3390/agriengineering1020018
83. Abu-Khalaf N. Sensing tomato's pathogen using Visible/Near infrared (VIS/NIR) spectroscopy and multivariate data analysis (MVDA). *Palest Tech Univ Res J.* (2015) 3:12–22. doi: 10.53671/ptur.v3i1.35
84. Fan G, Zhu J, Du R, Gao L. Determination of soluble solids and firmness of apples by Vis/NIR transmittance. *J Food Eng.* (2009) 93:416–20. doi: 10.1016/j.jfoodeng.2009.02.006

A NIR-spektroszkópia legfrissebb eredményei a kertészeti termények minőségének és biztonságának értékelésében: Átfogó áttekintés

R. Pandiselvam^{1*}, V. Prithviraj², M. R. Manikantan^{1*}, Anjineyulu Kothakota³, Alexandru Vasile Rusu^{4,5}, Monica Trif^{6*} és Amin Mousavi Khaneghah⁷

¹Fiziológiai, biokémiai és betakarítás utáni technológiai osztály, ICAR - Központi Növénytermesztési Kutatóintézet, Kasaragod, Kerala, India

²Élelmiszermérnöki Tanszék, Nemzeti Élelmiszertechológiai, Vállalkozási és Menedzsment Intézet, Sonapat, Haryana, India.

³Agro-feldolgozási és Technológiai Osztály, CSIR-Nemzeti Interdiszciplináris Tudományos és Technológiai Intézet (NIIST), Trivandrum, Kerala, India.

⁴Élettani Intézet, Kolozsvári Agrártudományi és Állatorvosi Egyetem, Kolozsvár, Románia

⁵Az Állattudományi és Biotechnológiai Kar, Kolozsvári Agrártudományi és Állatorvosi Egyetem, Kolozsvár, Románia

⁶Élelmiszeripari Kutatási Osztály, Innovatív Folyamattechnológiai Központ (CENTIV) GmbH, Stuhr, Németország

⁷Gyümölcs- és zöldségtermék-technológiai osztály, Prof. Waclaw Dabrowski Mezőgazdasági és Élelmiszeripari Biotechnológiai Intézet - Állami Kutatóintézet, Varsó, Lengyelország

Az agrártermékek minőségi és mennyiségi értékelését gyakran a hagyományos, azaz roncsolásos módszerek alkalmazásával végzik. A bennük rejlő hátrányok miatt a közeli infravörös spektroszkópiát (NIRS) használó, kemometriával párosított roncsolásmentes módszerek a különféle mezőgazdasági termékek értékelésében is hasznosnak bizonyulnak. A számítógépes teljesítmény, a gépi tanulás, a regressziós modellek, a mesterséges neurális hálózatok (ANN) és más prediktív eszközök fejlődése utat tört magának a NIRS területén, növelve annak potenciálját, hogy a roncsolásos vizsgálatok helyett megvalósítható alternatívát nyújtson. Ezen túlmenően a megfelelő előfeldolgozási technikák és hullámhossz-választási módszerek beépítése vitathatatlanul bizonyította a gyakorlati megvalósíthatóságát. Jelen áttekintés az összegyűjtött spektrális adatok feldolgozására alkalmazott számos számítási módszerre összpontosít, és megvitatja a NIRS lehetséges alkalmazásait a mezőgazdasági termékek minőségének és biztonságának értékelésére. Emellett tárgyalja a technológiával kapcsolatos kihívásokat, valamint a lehetséges jövőbeli perspektívákat is. Következtetésünk szerint a NIRS egy potenciálisan hasznos lehetőség a mezőgazdasági termékek minőségének és biztonságának gyors értékeléséhez.

KULCSSZAVAK

fajtahitelesítés, sérülések kimutatása, érettség, gyors módszer, spektroszkópia, textúra

Bevezetés

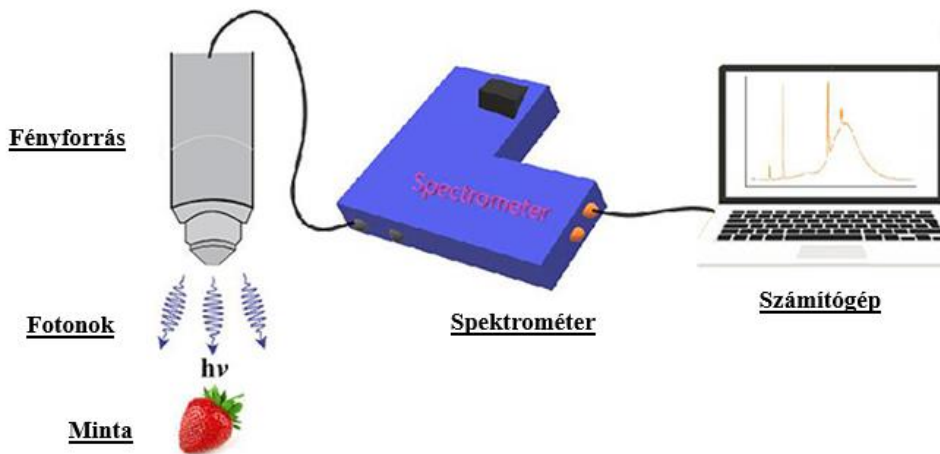
Az agrár alapú feldolgozóipari ágazatok a friss és minimális mértékben feldolgozott termékek előállítására haladnak. A fogyasztói érdeklődés hatására új eljárások és termékek jelennek meg a piacon. A mezőgazdasági termékek jelentik a legfontosabb nyersanyagot számos élelmiszeripari ágazat számára (1, 2). A termékek környezeti viszonyok között rendkívül romlandóak, ezért eltarthatóságuk korlátozott, azonban a hűtött tárolási körülmények között meghosszabbítható. A gyümölcsök és zöldségek nem tárolhatóak hosszú ideig a romlásra, leginkább a mikrobiális és kémiai romlásra való érzékenységük miatt, így a betakarítás utáni ellátási lánc egyes folyamatait, mint például az osztályozás és a válogatás, azonnal a betakarítás után kell elvégezni.

A végtermékek minőségének és biztonságának biztosítása érdekében a feldolgozás előtt és a betakarítás utáni teljes ellátási láncban megfelelő analitikai módszerek alkalmazására van szükség (3). A gyümölcsök és a zöldségek hibáinak kimutatására alkalmazott hagyományos módszerek azonban számos hátránya van: A friss gyümölcs minőségének, a fajta eredetiségének és a termékek sérülésének értékelése kihívást jelent. A friss gyümölcs minőségének és a fajta eredetiségének vizsgálata, valamint a termék sérülésének értékelése kihívást jelent a jó minőségű alapanyagok gazdáktól vagy nagykereskedőktől való begyűjtése során. Sokféle sérülés nem észlelhető kizárólag szemrevételezéssel, ami megnehezíti a kiváló minőségű termékek kiválasztásának folyamatát. A hagyományos vizsgálati módszerek, mint például a nagy tételekből történő mintavételek, időigényesek és igen költségesek (4-7). Ezenkívül a hagyományos vizsgálatok és mintavételek során gyakran jelentős mennyiségű termék megveszendőbe.

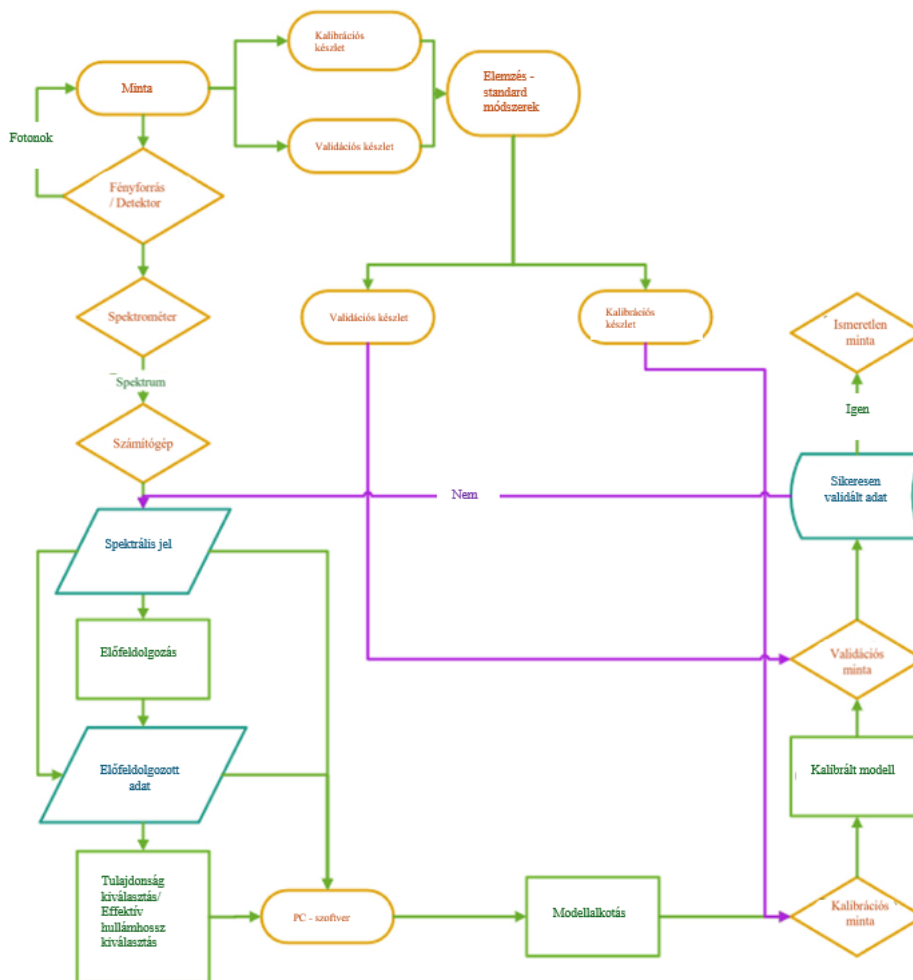
Különleges technikák szükségesek például az almák napfény okozta hegesedésének felismeréséhez és a hibás gyümölcsök kiválogatásához, mivel az ilyen sérülések nem kezelhetők kémiai úton (8). Az almáknak lehetnek olyan belső barnulási károsodásai is, amelyek a felületen nem láthatóak (9), így ismét kizárható a szemmel történő észlelés lehetősége. Ugyanígy a gyümölcsöket gyakran hűtve tárolják az eltarthatósági idő meghosszabbítása érdekében, ez azonban a hűtési sérülések kockázatát hordozza magában, ami a gyümölcshéj színéből vagy a külsejének egyéb jellemzőiből nem állapítható meg (10). A helyi piacok általános gyakorlata ezért az, hogy a gyümölcsök és zöldségek árának meghatározásakor a fizikai tulajdonságaik alapján, mennyiségi és minőségi becsléssel történik az árképzés (11).

Tehát a roncsolásmentes analitikai módszerek létfontosságú szerepet játszanak a hagyományos laboratóriumi módszerek kihívásainak leküzdésében. E tekintetben a közeli infravörös spektroszkópia (NIRS), prediktív algoritmusokkal ötvözve, a legalkalmasabb a termékminőség értékelésére és a sérülések felismerésére, valamint a fajták azonosítására (12).

A NIR spektroszkópiai rendszer főként egy egymással összekapcsolt fényforrásból, spektrométerből és számítógépből áll, ahogyan az 1. ábrán látható (ez az ábra a lehető legegyszerűbben ábrázolja, és nem számol egy sokkal praktikusabb NIRS-rendszerrel). A fényforrás az infravörös tartományban fotonokat bocsát ki, amelyek a mintával találkoznak, és további kölcsönhatásaik rezgéseket vagy kötésnyúlást (13) hoznak létre a minta belsejében lévő molekulákban. Az ilyen rezgések a molekula tulajdonságaitól és a vonatkozó kémiai kötésektől függő spektrumot hoznak létre. Ezek a spektrumprofilok a spektrum meghatározott részein érvényesülnek. Például az olyan molekulák, mint a klorofill, a 500-750 nm-es tartományban rezegnek (8, 9), míg a vízben lévő O-H kötések a 970-1 150 nm-es tartományban (9). A spektrumokat speciális előfeldolgozó technikákkal kell analizálni a zaj, valamint a nem kívánt és felesleges adatok kiszűrése céljából. A későbbiek során további előfeldolgozási technikákat alkalmaznak az effektív hullámhossz megtalálására, amely a spektrum meghatározott részeit oly módon foglalja magába, hogy a termék precíz osztályozására alkalmas legyen, ahogyan azt a 2. ábra is mutatja. A megfelelően felépített modell képes a terméket kis hibával vagy akár hibátlanul osztályozni. A NIR spektroszkópiai rendszer tervezésénél ügyelni kell arra, hogy a standard módszerek alkalmazásával, a lehető legkisebb időkésleltetéssel és nagy pontossággal tökéletes analízist végezzenek. A standard hiba a laboratóriumban (SEL) az előrejelzés standard hibájához (SEP) vezethet, amelyet a kvantitatív előrejelzés során megfelelő módszerek alkalmazásával a minimálisra kell csökkenteni.



8. ábra: Spektrális adatok gyűjtésére szolgáló egység



9. ábra: A NIRS modellezési folyamatot bemutató folyamatábra

Áttekintették a NIR-spektroszkópia felhasználási lehetőségeit a gabonafélék (14) és a tenger gyümölcseinek (15) feldolgozásakor, a tejtermékekkel kapcsolatos különböző alkalmazások

esetén (16), valamint a hús és a fűszerek minőségének elemzésénél (4, 17). Ennek ellenére kiemelkedően fontos megérteni a NIRS különböző mezőgazdasági termékekkel kapcsolatos különféle alkalmazásait, mivel ezek az alkalmazástechnikai lehetőségek megnyitják az utat a kiváló minőségű késztermék előállításához ezekből a nyersanyagokból. E tekintetben a mostani áttekintés a mezőgazdasági termékeknél alkalmazott különböző NIR-spektroszkópiai mérési technikákat és optimalizálási stratégiákat tárgyalja.

A spektroszkópiai technikák alkalmazása

Minőségi paraméterek

Az élelmiszerek és azok megbízhatósága azon termékek minőségétől függ, amelyekből készülnek. A NIR-spektroszkópia az egyik legjobb módszer ezen termékek elsődleges jellemzőinek, azaz az összes oldható szilárd anyag (TSS), az oldható szilárd anyag tartalom (SSC), a titrálható savtartalom (TA) és a pH előrejelzésére, spektrális jelek és algoritmusok segítségével. A NIRS különböző alkalmazásait a mezőgazdasági termékek minőségének meghatározására az 1. táblázat mutatja be.

1. táblázat: A NIRS felhasználása különböző mezőgazdasági termékek minőségi szempontjainak meghatározására

Termény	Spektrális tartomány	Szoftver	Minták száma	Pontosság	Megállapítások	Hivatkozás
Alma és alma püré	800-2500 nm	OPUS v. 5.0, XLSTA T, MATLA B v. 7.5, R v. 3.5.2	240	Az alma osztályozási pontossága = 82% és a püréké = 88%	A legjobb előrejelzést a viszkozitás, a sejtfal-tartalom, a szárazanyag, az SSC és a titrálható savtartalom mutatta. ($R^2 > 0.8$). Az almapüré végső állapotát az almákból jósolták.	(18)
Alma	350-2500 nm	MATLAB 2014a, Unscrambler v. 10.5x	120	SSC: ($R^2=0,87$; RMSEP=0,55) pH: ($R^2= 0,72$, RMSEP=0,009)	A hangtompítás a wavelet transzformációval az előkezelés előtt hatásosnak bizonyult.	(19)

Mandarin	700-1100 nm	MATLA B v. 7.0	275	94,0% SSOM-ot használva	Az SSOM egy nem lineáris osztályozó, amelyet sikeresen használnak MC, SSC, TA és szemcsésedés felismerésére.	(20)
Bodza	800-2500 nm	OPUS v. 7.2, Unscrambler 10.4, Statistica v. 7.2	11 mennyiségű gyümölcs, mixelve	97,06%	A TSS és az SSC jól korrelált a spektrális adatokkal, és a mintafelismerés lehetséges volt.	(21)
Alma	800-2500 nm	OPUS v. 7.2	214	TA: (R ² =68,17%, RMSEP=0,12) TSS/TA (R ² =82,62%, RMSEP=0,43) TSS: (R ² =90,93%, RMSEP=0,61)	A spektrális intenzitások a használt spektroszkóptól függően változtak. A GA jelentősen javította az előrejelzés minőségét.	(22)
Hami dinnye	550-950 nm	Unscrambler v. 9.7	120	RMSEP=0,95-0,99	Az MC-UVE-SPA-MLR kombinált előfeldolgozási és nemlineáris előrejelző algoritmus esetében az SSC-meghatározás adta a legjobb eredményt.	(23)
Szőlő	800-1100 nm	Unscrambler v. 10.5	120	TA: (R ² =0,716, RMSEP= 0,103) pH: (R ² =0,547, RMSEP= 0,395) SSC: (R ² =0,971, RMSEP= 0,522)	A különböző előkezelések kapcsán jelentős eltéréseket figyeltek meg. A kalibrációs eredmények javítása nem mindig javítja a validálási eredményeket.	(24)

Ananász	740-1070 nm	MATLA B v.9.5.0	90	85%	A TSS jó előrejelzési pontossága RMSEC = 0,95 és RMSEP = 0,84.	(25)
Olajbogyó	2307-2348 nm	Ucal	100 (12 fajta)	$R^2=0,964$ és 959	Az olajtartalom előrejelzése az olaj kinyerése előtt.	(26)
Gránátalma	400-1100 nm	ParLeS v. 3.1, AvaSoft 7	100	TSS: ($R=0,95$, RMSEC=0522)	A reflexiós mód jobb volt, mint a transzmissziós mód a TSS és a pH meghatározásához.	(27)
Kakukkfű	1000-2500 nm	NIRwar e v. 1.2	147	$R > 0,75$, $RPD < 1,5$	Az SSC előrejelzések a legjobb eredményt MSC előkezeléssel érték el.	(28)
Alma	550-950 nm	Unscrambler v. 10.1 MATLAB N 2016a, Spectra Suite, Visual Studio 2010	180	$r_p=0,842$, RMSEP=0,453	A legjobb tájolásnak az bizonyult, ha a törzs és a kocsány tengelyét függőlegesen, a szárat felfelé tartják. A PLS és LS-SVM modelleket a kompenzációs modellek létrehozására, a CARS és SPA modelleket pedig a hatékony hullámhosszok kiválasztására használták.	(29)
Szőlő	450-2500 nm	Vision software , GenStat	120	TSS: ($R^2=0,896$, RMSEP=0,308) TA: ($R^2=0,835$, RMSEP=0,066) TSS/TA:	A cukor és a sav közötti egyensúlyt mennyiségi paraméterként vették figyelembe, amelyet az ízérzékeléssel	(11)

				($R^2=0,812$, RMSEP=0,451)	hoztak összefüggésbe.	
Paradicsom	400-1100 nm 900-1700 nm	MATLA B R2016	600	pH: ($r_p=0,819$) SSC: ($r_p=0,800$)	A térbeli felbontású spektroszkópia hatékonyan bizonyult a különböző szerkezetű és kémiai összetételű mezőgazdasági növények esetében.	(30)
Paradicsom	930-1650 nm	Spectra Wiz, ParLeS v. 3.1	120	Likopin: ($r_{cv}=0,840$, RMSECV=2,25 6) C-vitamin: ($r_{cv}=0,818$, RMSECV=1,08 7)	Az MSC az első származékkal kombinálva képes volt megjósolni a likopin és a C- vitamin tartalmát a PLS-modell segítségével. Az itt alkalmazott NIR- eljárás nem biztosít nagy pontosságot.	(31)
Narancs és szőlő	450-2500 nm	Vision TM v. 3.5.0.0, GenStat	120 grapefru it 120 narancs	TSS: ($R^2=0,927$, RMSEP= 0,283) TA: ($R^2=0,929$, RMSEP=0,017) TSS/TA: ($R^2=0,958$, RMSEP=0,2) TSS: ($R^2=0,896$, RMSEP=0,308) TA: ($R^2=0,835$, RMSEP= 0,066) TSS/TA: ($R^2=0,812$, RMSEP=0,451)	A narancs és a szőlő édes és ízjellemzőit vizsgálták. A BrimA (Brix csökkentve savakkal) érzékszervi paramétert értékelték.	(32)

SSC, oldható szilárdanyag-tartalom; RMSEP, az előrejelzés átlagos négyzetes hibája; SSOM, felügyelt önorganizáló térkép; MC, nedvességtartalom; TA, titrálható savtartalom; TSS, összes oldható szilárd anyag; GA, genetikai algoritmus; RPD, maradék előrejelzési eltérés; MSV, multiplikatív szórás-korrekció; MC-UVE, Monti Carlo-uninformatív változóelimináció; SPA-MLR, szukcesszív projekciós algoritmus-többszörös lineáris

regresszió; CARS, kompetitív adaptív újrásúlyozott mintavételezés; RMSECV, a keresztellenőrzés átlagos négyzetes hibája.

A Lan és társai által az alma minőségéről készített tanulmány az almából készült püré jellemzőit értékelte. Az alma spektrális részleteit a viszkozitás, a sejtfa-tartalom, a szárazanyag, az SSC és a püré termék TA előrejelzésére használták. Az említett paraméterek 0,8-nál nagyobb R^2 értékeket mutattak, ami a módszernek a pontosságát jelzi (18). A vizsgálat spektrális méréseit a 800-2 500 nm-es tartományban végezték egy 18 különböző pozíciójú automatikus mintavevő tárcsa segítségével. A hat optikai szál (9) vagy kézi pozicionálást alkalmazó módszer azonban valóságosabb és pontosabb. Ennek oka lehet a mintavevő tárcsa beállításának állandósága. Az SSC-tartalom meghatározása 0,92-es R^2 értéket eredményezett a püré termék homogenitása miatt (18).

Ezek az algoritmusok akkor teljesítenek a legjobban, ha a termék homogén, nem pedig inkonzisztens. Ehhez hasonló vizsgálatokat végeztek calçot hagymán és almapürén, és pontos előrejelzéseket kaptak az oldható szárazanyagra, a glükózra, az almasavra és a szárazanyagra vonatkozóan is (33, 34). Az említett vizsgálatok azt bizonyítják, hogy a pürék minőségének értékelése sikeresen elvégezhető a spektrumanalízis segítségével.

Ezeknél a körülményeknél a részleges legkisebb négyzetek (PLS) modellek használata a TSS előrejelzésével $R^2=0,95$ értéket eredményezett, ami jobb, mint minden más paraméter (27). Ugyanakkor a reológiai állandók és a színértékek alulteljesítettek. Ez a reológiai változások nem lineáris jellegéből vagy a spektrális sávok átfedéséből adódhat (33). A püré jellemzőinek előrejelzése az ép almák minőségének értékelésével kihívást jelentő és félig kvantitatív, ipari eljárásra alkalmas volt (33). A NIRS a kémiai összetétel és a fotonválasz alapján működik; így az alapanyagok feldolgozása és a nyers termények értéknövelt késztermékké alakítása jelentősen megváltoztatja a termék kémiai tulajdonságait. Emiatt lehet, hogy a feldolgozott termékek minőségi jellemzőinek előrejelzésében több rejtett változót és kisebb pontosságot kapunk.

Khodabakhshian és munkatársai belső minőségi vizsgálatokat végeztek gránátalmán a 400-1100 nm-es tartományban mind transzmissziós, mind reflexiós módszerek alkalmazásával. Vizsgálatukban a modell teljesítményét a megjósolt reziduális hiba négyzetek összegének (PRESS) módszerével értékelték, amely a regressziós elemzésben keresztvalidálási technikaként szolgált. A szokásos előkezeléseket, különösen a standard normálváltozó (SNV_)

és a többszörös szóráskorrekciót (MSC) elvégezték, így figyelembe vették a gránátalmafajták között fennálló morfológiai változásokat (27).

Gyakorlati eseti tervekben mindig figyelembe kell venni a morfológiai változásokat. Ezért, bár az SNV és az MSC előfeldolgozása használható, nem biztos, hogy ez minden helyzetben megfelelő. A reflexiós és transzmissziós spektrumok általános tulajdonságai hasonlóak voltak, és csúcsokat képeztek a 750 és 970 nm-es tartományban, bár a transzmissziós spektrumoknál több zaj volt megfigyelhető (27). A hullámhosszal kapott spektrumok szabálytalan kiugrásokat ábrázoltak az adatok teljes egészén, ami a zajproblémákra utal. A spektrális jellemzőkre vonatkozó megkülönböztető csúcsok a NIRS-elemzés előnyben részesített módszere. Az előfeldolgozási technikák érdekes módon hasonlóak maradtak a legtöbb minőségi tulajdonság esetében, függetlenül a terménytől, ami az elemzési pontban rejlő hatékonysági lehetőségeket mutatja. Mind a transzmissziós, mind a reflexiós módszer hasonló teljesítményt nyújtott, bár az utóbbi a jobb átviteli képessége (21, 35, 36) és a robosztusabb rendszere miatt némileg felülkerekedett. A vékony héjú gyümölcsök, mint például a gránátalma, jobb eredményt mutattak a reflexiós módban, ami arra utal, hogy ez a módszer használható vastag falú gyümölcsök, például kókuszdió, bétel dió és kakaó esetében is.

Egy Hu és munkatársai által az SSC meghatározásáról végzett átfogó vizsgálatban különböző mérési konfigurációkat, változóválasztó algoritmusokat és osztályozási modelleket vizsgáltak a Hami dinnye esetében. Ezzel összefüggésben a calyx modell teljesített a legjobban, ami a calyx régió magasabb SSC-tartalmának tudható be (23). Az eredményt tovább lehet interpolálni, hogy minden előfeldolgozás után a legjobb előrejelzési eredményeket kapjuk. A méréseket az attribútum legkiemelkedőbb pontján kell elvégezni.

Egy almán végzett hasonló vizsgálatban többféle irányt vettek figyelembe, és ennek eredményeképpen a függőleges szár-kocsány irányt választották optimális irányynak, amelyben a szár felfelé áll (29). Ez arra enged következtetni, hogy az, hogy egy tájolás jobban teljesít-e, mint a többi, a tényező előtérbe kerülésének köszönhető az adott tájolási területen. A változóválasztás és a predikciós algoritmusok különböző kombinációi azt mutatják, hogy a Monte Carlo algoritmus (MC-UVE-SPA = Monte Carlo–uninformative variable elimination–successive projections algorithm) az összes előrejelző algoritmussal a legjobb eredményeket érte el az SSC meghatározásához ($R_p > 0,8$) (23). Xia és munkatársai azt tapasztalták, hogy a CARS-SPA-PLS teljesített a legjobban az alma SSC előrejelzésében, alacsony, $< 0,573$ Brix° átlagos négyzetes előrejelzési hibát (RMSEP) elérve. Arra is rámutattak, hogy az effektív

hullámhossz globális modellből történő meghatározása segíthet az irányultság hatásának bizonyos határig történő csökkentésében (29).

A hatékony hullámhosszok kiválasztása, amelyek jól megfelelnek a mintákat osztályozó régióknak, csökkentheti a berendezés költségeit, ahogyan azt egy banánminőség-értékelésről szóló tanulmány is mutatja, amelyben a lehető legalacsonyabb hullámhosszú ablakot javasolták (37). Minden mezőgazdasági termék hatékony hullámhossza eltérő, ezért a spektrumanalízis során standardizálni kell. A banán és az alma esetében az effektív hullámhossz változásának ez a különbözősége a szín és az általános összetétel eltéréseiből adódhat. Így levonhatjuk azt a következtetést, hogy az optimális teljesítmény érdekében az online mérőrendszer tervezésekor figyelembe kell venni a tájolást, az effektív hullámhosszt, a használt termékeket és a berendezés kialakítását. Különböző tanulmányok képesek voltak az SSC-tartalom előrejelzésére $R_p > 0,9$, és jobb eredményeket értek el, mint más paraméterek, például a TA és a pH esetében (24). Hasonlóképpen, az ananász TSS-tartalmát különböző algoritmusok kombinálásával hasonlították össze (25). Az utóbbi tanulmány, amely egy kézi spektrométert használt, 85%-os pontosságot ért el, 0,95 RMSEC és 0,84 RMSEP értékkel (25). Ez a technológia a legjobb gyakorlati eredményeket kínálja a megvalósíthatóság és a jövőbeli kilátások szempontjából, még kézi készülék használata esetén is.

Kanchanomai és munkatársai (24) az SSC, a pH és a TA segítségével vizsgálták a szőlő minőségének lehetséges gyors értékelési technikáit. Az SSC előrejelzés 0,97-es R_p értéket ért el, míg a TA ($R_p = 0,71$) és a pH ($R_p = 0,54$) viszonylag kevésbé volt pontos. A szilárdságot és a magnélküliséget is vizsgálták, azonban ezek a tényezők is alacsony előrejelzési pontossággal rendelkeztek. Ebből a kutatók azt a következtetést vonták le, hogy a 800-1100 nm-es NIR-tartomány alkalmas lehet a szőlő belső minőségének előrejelzésére. Hasonló módon a csemegeszőlőt is vizsgálták különböző előfeldolgozási technikákkal és a 400-1000 nm-es NIR-tartományban (27, 38). A megfelelő előfeldolgozási technika 0,6-ról 0,8052-re és annál magasabbra javította az R_{Pred} értékét, amikor Savitzky-Golay második deriválási modelljét (SG2) alkalmazták. Ez arra enged következtetni, hogy a csemegeszőlő esetében az SG2 előfeldolgozás alkalmazásával különböző paraméterek, úgymint a pH, a TSS és a keménység előrejelezhető. Az elért jobb eredmények a megfelelő hullámhossz-tartomány és a csemegeszőlőre alkalmazott előfeldolgozási technika alkalmazásával igazolhatóak.

Az alma SSC- és összes savtartalmának (TAC) meghatározásához a visszaterjedéses neurális hálózat (BPNN), az általános regressziós neurális hálózat (GRNN) és a részecske raj optimalizálás (PSO) kombinációját használták (39). Ez a vizsgálat a korábbi technikáknál

említett hibrid mesterséges neurális hálózat (ANN) modelljének kifejlesztésére összpontosított; amelyre az ANN-ben rejlő korlátok áthidalása miatt volt szükség. Az eredmények azt mutatták, hogy az SSC előrejelzése során a hibrid modell és a BPNN modell közel azonos RMSEC és RMSEP értékeket mutat ($<0,6$, illetve $<0,7$), míg a GRNN nem nyújtott megbízható teljesítményt ($>0,6$, illetve $>0,9$ értékek). Azonban a TAC meghatározása során a GRNN és a hibrid modellek alacsony RMSEC és RMSEP értékekkel rendelkeztek ($<0,1$), ugyanakkor a BPNN sem működött jól ($>0,2$). Ez alátámasztja, hogy a hibrid modellek adaptálása valóban növelte az algoritmusok sokoldalúságát a különböző minőségi jellemzők meghatározásában, és nem a paraméterekkel kellett változtatni azokat.

Az online meghatározási módszerek tekintetében különböző rendszereket vizsgáltak meg alma, mangó és banán esetében. A termék helyes igazitása, amint azt az előző megállapítás is említette, az online rendszer tervezésének egyik fő kihívásának bizonyult. E probléma megoldására törekedve vizsgálták a CARS-SPA-PLS modell használatát az SGS után, és azt tapasztalták, hogy az alma esetében hatékony (29), $r_p > 0,8$ értékkel. Az almák online előrejelzésére az MCARS-SPA-PLS rendszerhez hasonló, nagyobb pontosságú rendszert javasoltak. Ezen algoritmusok képesek a biológiai változékonyságot kezelni és az előrejelzés során a tájolási problémákat kezelni. Az online rendszereket a tökfélék (40) és a gyümölcslevek esetében az irányultsági kihívásoktól függetlenül rendkívül jól alkalmazhatónak találták. Valószínűleg ez a gyümölcsök inkonzisztenciájának és az értékteremtő termékek homogenitásának köszönhető.

Gyors minőségvizsgálati eljárásokra van szükség az olyan gyorsan romlandó termékek esetében, mint a banán. Erre egy online szállítórendszert terveztek, amely figyelembe vette a TSS-t, a pH-t, a szárazanyagot (DM) és a sav-brix arányt (ABR) (37). A validációs szakaszban 0,81, 0,78, 0,78 és 0,87 értékű megfelelő R^2 adatokat kaptak, amelyek rendkívül pontos előrejelzési eredményeket mutattak. A vizsgálatban szelektív előfeldolgozást is végeztek mind a pH, mind az ABR esetében, ami javította az előrejelzési eredményeket. Az alma esetében az online felismerés pontossága magasabb volt, mivel az algoritmusok kombinációja "sikeresen kezelte a termék tájolásának nehézségeit". Ennélfogva hasonló kombinációs algoritmusokat kell adaptálni az online rendszerekhez, tekintettel az ilyen mérések pontos tájolással történő elvégzésének kihívásaira.

Texturális tulajdonságok

A mezőgazdasági termények minőségét, érzékenységét és megjelenését befolyásoló egyik legfontosabb tényező a textúra. A friss termények textúrájának mérésére leggyakrabban használt mutatók a keménység és a penetrométerrel mért eredmények (41, 42). Néhány tanulmány a paraméterek szélesebb körét elemezte, nevezetesen a szakítóerőt, a keménységet, a látható rugalmassági modulust, a nyomóenergiát (43), a kezdeti szilárdságot, az átlagos szilárdságot, a héj szakítóerejét, a szakítótávolságot, a penetrációs energiát és a penetrációs erőt (42). Ezek a paraméterek megkülönböztethetik a fajtákat (41), mivel a különböző fajok és a különböző biológiai feltételek egyedi tartományokkal rendelkeznek. Azt is érdemes megjegyezni, hogy egyes tanulmányok úgy döntöttek, hogy a nagyobb méretű tárgyak szilárdságának és penetrációs mélységének értékelésére inkább univerzális tesztelőgépet (UTM) használtak, mint texturális analizátort.

A mezőgazdasági termények texturális tulajdonságainak mérésére szolgáló NIRS különböző alkalmazásainak összefoglalása a 2. táblázatban látható. Sánchez és munkatársai (44) a spenót szerkezetét elemezték egy mikro elektromechanikus rendszerhez (MEMS) csatlakoztatott kézi NIR-készülékkel. A kutatók egy egyszerű kézi készüléket használtak a NIR-méréshez, ezzel demonstrálva a kompakt, praktikus technológiák hasznosságát a bonyolultabb, magas költségű műszerekkel szemben. A NIR-spektroszkópia CH-, OH- és NH-kötésekre és azok abszorpciós mintázatára adott válasza csúcsokat és hullámokat (41) eredményez a spektrális jelben, ami az elemzés alapját képezi. Megállapítható volt, hogy a kézi spektrométer használata közel 1-hez közeli R^2 értéket ért el. A kísérlethez használt nyersanyagok Spanyolország különböző gazdaságaiból származtak. Az ehhez hasonló (azaz nagyobb mennyiségű fajtát és biológiai variációt összesítő) gyakorlatok jobb kalibrációt és adatfeldolgozást tesznek lehetővé, ami viszont hozzájárul a robusztusabb előrejelző rendszerek kialakításához, és végső soron jobb eredményeket eredményez.

2. táblázat: A NIRS felhasználása a mezőgazdasági termékek texturális tulajdonságainak meghatározására

Termény	Spektrális tartomány	Szoftver	Minták száma	Pontosság	Megállapítások	Hivatkozás
Mangó	800-2500 nm	OPUS v. 7.0.129	85	$R = 0.7-0.75$	A penetrációs paramétereket, a szilárdságot és a	(42)

					szakítóerőt előrejelezték.	
Spenót	1600-2400 nm	WinISI II v. 1.50	149	$R^2_{cv} = 0.74$	Meghatározták a fogyasztói elfogadottsági paramétereket, a szárazanyagot és a texturális tulajdonságokat.	(44)
Paradicsom	800-2500 nm	OPUS	30	$R^2 = 0.7-0.97$ RMSECV = 0.4% Brix	Az SSC-tartalmat PLS-elemzéssel jelezték előre. A texturális előrejelzés pontos volt.	(41)
Paradicsom	950-1650 nm	SPSS 13.0, Permut Matrix v. 1.9.3, Unscra mbler v. 9.7, MATLA B 2017	90	$R_p^2 = 0.85-$ 0.966 RMSEP < 0.5	A feldolgozatlan spektrumok felülmúlták az előfeldolgozottakat. Az ELM a legjobb előrejelzést adta a PLS és az SVM felett.	(45)
Kiwi	800-900 nm	ImageJ, RGS- AvaCam v. 3.7.0, MATLA B R2016b, Unscra mbler v. 9.7	116	$R^2 = 0.68-0.77$	A szilárdságot a hasonló szürke képpontok befolyásolták. A PLS valamivel jobb eredményeket ad, mint az ANN.	(46)
Olajbogyó	1100-2300 nm	SNAP 2.03, Unscra mbler	100	$R^2 = 0.8-0.99$	Egy nagy adathalmaz több tényezőt is figyelembe vehet, és így pontos előrejelzést tud	(47)

		v. 9.7, SigmaPI ot v. 10.0			elérni. Egy gyors és olcsó technikát alkalmaztak.	
Körte	380-1030 nm	Spectral Image System, Unscrambler v. 9.7, ENVI v. 5.0, MATLAB v. 8.1	135	$R_p^2 = 0.89-0.92$	A valós idejű alkalmazáshoz > 1,5 m/s szkennelési sebesség szükséges. SAE-FNN segítségével mélytanulást végeztek.	(48)
Pisztácia	200-1100 nm	AvaSoft 7, ParLeS v. 3.1	81	$R^2 = 0.754-0.91$ RMSEP = 0.253-26.049	Az SNV modell jobban teljesített, mint az MSC. A texturális paraméterek közül a törőerőt jelezték előre a legpontosabban.	(43)

SSC, oldható szilárdanyag-tartalom; RMSEP, az előrejelzés átlagos négyzetes hibája; RMSECV, a keresztellenőrzés átlagos négyzetes hibája; PLS, részleges legkisebb négyzetek; ELM, extrém tanulási gépek; SVM, támogató vektor gépek; ANN, mesterséges neurális hálózat; SAE, halmozott automatikus kódoló; FNN, teljes mértékben összefüggő neurális hálózat; MSC, többszörös szórás-korrekció.

Más tanulmányok eredményei, a paradicsom textúrájának meghatározására szolgáló módszerek vizsgálata, amelyben a hasonló érettségi csoportokba tartozó paradicsomok a szórásos diagramon különálló klasztereket adtak, alátámasztják ezt az állítást (41). A jól elkülönülő klaszterek nem veszik figyelembe a mintában jelen lévő különböző érettségi állapotokat és fajtákat; ennél fogva a mintavétel nagyságát növelni kell egy jobban teljesítő modell létrehozása érdekében. A szerkezetvizsgálatokat nem textúraelemző készülékkel, hanem univerzális vizsgálóberendezéssel végezték (44, 47). Az élelmiszermatrixot szűrőerő, zúzóképeség, keménység, szakítópont (44) és szilárdság (47) szempontjából vizsgálták. A levelek átszűréséhez szükséges maximális erő ($r^2cv = 0,67$; $RPDcv = 1,72$), keménység ($r^2cv = 0,62$; $RPDcv = 1,62$), szilárdság ($r^2cv = 0,69$, $RPDcv = 1,79$) és az egyes levelek áttöréséhez szükséges szonda kitérése mutatta ($r^2cv = 0,62$, $RPDcv = 1,61$) (44). Hasonlóképpen, egy olajbogyón végzett keménységmérésnél magas, 0,997-es R^2 értéket mértek, ami arra utal, hogy

ez a paraméter kiváló minőségű előrejelzéseket tesz lehetővé (47). Az R^2 érték megfelelő előfeldolgozással jelentősen javul, amint azt a pisztáciamagokon végzett vizsgálat is szemlélteti, amely a nyers adatokat és a különböző előfeldolgozott eredményeket hasonlította össze (43).

A textúraelemzéshez alkalmazandó előfeldolgozási módszer kiválasztása a vizsgált paramétertől és a méréshez használt műszerektől függ. A Mohammadi-Moghaddam és társai (43) által végzett kutatás különböző előfeldolgozások után jobban teljesítő PLS-modelleket tudott kialakítani a texturális elemzéshez. Meg kell jegyezni, hogy az előfeldolgozás nem mindig szükséges, és hogy a körte keménységének meghatározásakor (PLS-regresszióval) előfeldolgozás nélkül is jobb eredményeket értek el (48). A különböző előkezelések közvetlen összehasonlítása a legjobb megoldás meghatározása érdekében nem lehetséges, mivel ez számos tényezőtől függ.

Az olajbogyóban a keménység az olajtartalom és a szárazanyag közötti kapcsolat miatt korrelál az érettséggel és az olajtartalommal (47). Mivel ezen tényezők mechanikai és szerkezeti kapcsolatban állnak egymással, a NIRS segítségével történő vizsgálat roncsolásmentes mérési módszert kínál a szájéret meghatározására. Annak ellenére, hogy a keménység előrejelzése kissé bonyolult, egy olajbogyón végzett vizsgálat során a PLS-regresszió segítségével 1,248, 1,449, 1,758, 1,917, 1,990 és 2,238 nm-es csúcsterületeken magas korrelációt és alacsony hibarányt biztosítottak (47). Hasonló vizsgálat cherry paradicsom vizsgálatával $R^2 = 0,966$ -os keménység-előrejelzést eredményezett (45). A vizsgálatban használt extrém tanulógép algoritmus a rejtett szintek optimalizálásával működik a keresztvalidálási lépésben, ezzel gyakorlati problémákat is megoldva (45).

Camps és Gilli (41) a kompressziós és a penetrométeres vizsgálatokat együtt végezte el paradicsomon, és a kompressziós vizsgálat jobb előrejelzési teljesítményt mutatott ($R^2 = 0,85-0,97$). A megfigyelést alátámasztja egy lédús csonthéjas gyümölcsökön végzett vizsgálat is, amely szerint a gyümölcs érésével párhuzamosan a kompressziós teszt is jelentőséggel bírt (49). Egy paradicsomon elvégzett textúraelemzés kimutatta, hogy a kompressziós paraméterek esetében körülbelül 7-10 rejtett változó létezik, míg a penetrométeres paraméterek esetében 3 és 8 között van (41), valamint hogy az adatok csoportosítása lehet az oka a jobb R^2 eredményeknek. Bár azonban a klaszterezett adatok javíthatják az R^2 értékeket, az RMSE értékek nem különbözhetnek jelentősen.

A mangó textúrája és minősége közötti összefüggésről szóló tanulmányok azt sugallják, hogy az R^2 érték alapján a paramétert vagy nagyfokú válogatásra, vagy pedig alapos értékelésre kell alkalmazni. Ez alapján egy tanulmány megállapította, hogy a gyümölcshús behatolási erejének, a héj felszakadási erejének és a gyümölcshús penetrációs energiájának paraméterei durva osztályozásra használhatóak (42). A szerzők a texturális paraméterek és más tényezők, például az érettség, a tárolási idő és a feldolgozási szakasz közötti kapcsolatot is tárgyalták. A feldolgozási műveletek, mint például a pörkölés, szintén a nedvességtartalom csökkenését okozhatják, ami nagyobb reflexiós értéket eredményez (43); hasonlóan a pisztáciamagoknál is megfigyeltek eltérést a csúcs amplitúdójában a kezelés hatására. Camps és Gilli (41) három paradicsomfajtát vizsgált 90 mintával, és ugyanezen paraméterek tekintetében túlszárnyalta a jelen vizsgálatot, bár a mintanagyság 80 volt egyetlen fajtán. Az eredmények megerősítik mind a paradicsom rendkívül változatos jellemzőit, mind az e jellemzőkkel kapcsolatos kutatási hiányosságokat. Yu és munkatársai (48) egy mély tanulási modellt javasoltak egy halmozott automatikus kódolóval (SAE) és egy teljes mértékben összekapcsolt neurális hálózattal (FNN), ahol a SAE-t használták a jellemzők kinyerésére, amelyeket FNN bemenetként adtak meg. Egy körtefélékkel végzett vizsgálatban a PLS-regressziót hasonlították össze egy mélytanulási algoritmussal, és a SAE-FNN esetében szignifikánsan jobb eredményeket kaptak (48). A mélytanulás 0,9-es tartományban jósolta az R^2 értéket, míg a PLS-regresszió csak 0,8 közeli értékeket tudott elérni (48). A kiwi gyümölcsök keménységének előrejelzésével kapcsolatos vizsgálatok azt mutatták, hogy a PLS ($R^2=0,77$) átlagosan jobban teljesített, mint az ANN ($R^2=0,72$) (46).

A fajták és a kultúrnövény azonosítása és hitelessége

A hiteles, nyomon követhető adatok az egész ellátási láncban alapvető fontosságúak, különösen az olyan technológiák alkalmazása esetén, mint például a blokk lánc módszerek. Manapság a fogyasztók a hitelességre összpontosítanak; követelik a nyomon követhetőséget és a szigorú minőségellenőrzést a gyártás során. Az elmúlt években a NIRS-t az osztályozási algoritmusokkal együtt integrálták az ellátási és termelési láncokba, hogy megfeleljen ezeknek az igényeknek.

Az alma (50-53), a paprika (54), a mangó (55), a körte (56), a burgonya (57) és az eperfa (7) különböző genotípusainak megkülönböztetésére alkalmas modellek kifejlesztését a NIR spektrum segítségével a 3. táblázat foglalja össze. Gyakorlatilag ezeknek a terményeknek a szétválogatását egy gyorsmozgású szállítószalagon kellene elvégezni, de egy spektrumgyűjtő rendszer tervezése ilyen körülmények között problémás. Mivel ugyanazon fajtájú

mezőgazdasági termékek is eltérő felületi tulajdonságokkal rendelkezhetnek, több, akár 250 leolvasott spektrummintát (7, 55, 59) kell beszerezni. A minta méretét és a szükséges mérések számát a minta heterogenitása és összetettsége határozza meg, például az, hogy a minta darabolt-e (59). A minta heterogenitását olyan technikákkal lehet figyelembe venni, mint például három vagy négy különböző típusú mérés elvégzése; a 120 és 90 fokos elforgatás (54, 58); valamint a gyárilag készített terménytartók használata (59, 61) vagy a spektrométeren belüli elhelyezés. A szabványos NIR-modellek a gyümölcs teljes felületének letapogatását igénylik diódásoros műszerek elrendezésével (44), a körülöttük lévő tér integrálását az egyébként elvesző információ visszanyerése érdekében (52), valamint olyan módszereket is, amelyek a gyümölcs és a mérőfej közötti távolságot a mérőfej méretétől függetlenül állandó szinten tartják (50).

3. táblázat: A különböző mezőgazdasági termények NIRS-sel történő fajtameghatározásának összefoglalója

Termény	Spektrális tartomány	Szoftver	Minták száma	Pontosság	Megállapítások	Hivatkozás
Paprika	1600-2400 nm	Win ISI II v. 1.5, MATLA B 2015a	394	88,28-91,37%	Az SSC és a szárazanyag felhasználásával végzett előszűrés sikeres volt. A SEP és SEL fontosságát megvitatták.	(54)
Eperfa	909-1649 nm	PLS Toolbox v. 6.21, MATLA B R2009	468	84,1%	A <i>Dendrobium officinale</i> Kimura et Migo (DOK) megkülönböztetésre került a <i>Dendrobium devonianum</i> Pax-tól (DDP).	(7)
Alma	1000-2500 nm	MATLA B 7.11, Antaris II System	180	74,44%	A PCA, a PCA+LDA, az SDA és a DPLS közül az SDA jobb teljesítményt mutatott a jellemzők kinyerésében.	(58)

Alma	1000-2500 nm	Fiber Optic Solids cell, NIRWare Unscrambler	410	77,9%	Az almák osztályozása különböző talajtípusok szerint.	(51)
Mandarin, vörös káposzta, savanyú uborka, kelkáposzta és almaszós	1100 nm és 2100 nm	PAS LABS v. 1.2, SIMCA v. 14.1	15	99%	A NIRS-előrejelzés az üvegekben tartósított termékek esetében volt lehetséges. Az OPLS-DA felülmúlta a PCA-t és a PLS-DAT.	(59)
Burgonya	964,13-1645,01 nm és 2502,50-16666,67 nm	Spectral Cube, OPUS v. 7.2, PLS-toolbox v. 8.6, Unscrambler v. 10.1, MATLAB R2017b	240	RP = 0,954 RMSEP = 0,421	A PLSR-modellt használták az átsütöttségi fok meghatározására és a fajta előrejelzésére.	(57)
Alma	300-1100 nm	ModelBuilder, R Statistical software	640	R ² érték 0,90 és 0,92, illetve az RMSE 0,67%	A fajták egyedi modelljei jobban teljesítettek, mint a kombinált modell.	(53)
Mangó	1200-2200 nm	Unscrambler	1310	Alphonso és Banganapalli (99,07%, 99,58%), Dasherri és	A pontosabb osztályozást egy megkülönböztetett pontozási ábrázolástette lehetővé.	(55)

				Malda (98,37%, 94%)		
Alma	400-1021 nm	Ocean View, MATLA B R2014b	300	SPA-SVM 85,83% SPA-ELM 95%	A BPNN, SVM és ELM modellek közül az ELM teljesített jobban. Az SPA-val és ELM-mel kombinált jellemzőkiválasztás jobb eredményeket hozott, mint a PCA.	(60)
Körte	350-1800 nm 350-1000 nm 1000-1800 nm	Unscra mbler v. 9.7	110	R ² 0,90–0,92 RMSEP 0,23–0,30	A CARS segítségével jobb volt a jellemzők kiválasztása, mint az MC-UVE és az SPA segítségével. A CARS-MLR és a CARS-PLS pontosan meghatározta az SSC-t.	(56)
Alma	1000-2500 nm	MATLA B R2014a	208	98,1%	A földrajzi régió jelentős hatással volt az SSC-re. A CARS jellemzőválasztás és a PLS-DA jó előrejelzési pontossággal rendelkezett.	

SSC, oldható szilárd anyag tartalom; SEP, az előrejelzés standard hibája; SEL, standard hiba a laboratóriumban; PCA, fő komponens analízis; LDA, lineáris diszkriminancia analízis; SDA, halmozott zajmentesítő automatikus kódolás; DPLS, dinamikus részleges legkisebb négyzetek; OPLS, ortogonális részleges legkisebb négyzetek; PLS-DA, részleges legkisebb négyzetek diszkriminancia analízis; PLSR, parciális legkisebb négyzet regresszió; BPNN, visszaterjedési neurális hálózat; SVM, támogató vektor gépek; ELM, extrém tanulási gépek; SPA, szukcesszív vetítési algoritmus; CARS, kompetitív adaptív újrasúlyozott mintavételezés; MC-UVE, Monti Carlo-uninformatív változóelimináció; SPA MLR, szukcesszív projekciós algoritmus-többszörös lineáris regresszió.

A kutatók többsége a rendelkezésre álló spektrométer-rendszert alkalmazza rögtön, ezért az osztályozási módszerek nagyobb hangsúlyt kapnak. Szélesebb választékú fotométerek és a nagyobb mintanagyság jelentősen javíthatják a modellt (54). A bővített modellek hasznosak ahhoz, hogy a spektrumban olyan egyedi csúcsokat kapjunk, amelyekkel megkülönböztethetjük

a fajtákat. Mivel a hőmérséklet és a megvilágítás ingadozása fontos szerepet játszik a NIR-modell létrehozásában, gondoskodni kell arról, hogy a környezeti feltételek minden mérésnél ugyanazok legyenek (53).

A fajták előrejelzésére a bemeneti adatokat előfeldolgozzák olyan módszerekkel, mint a többszörös szórás-korrektúra (MSC) (51, 55, 60), EMSC (50, 62), standard normálváltozó (SNV) (7, 51, 56), trendszűrő módszer (51, 55), normalizálás (58, 59), Savitzky-Golay (7, 50, 51, 61) és Norris gap (55). Ezek független referenciatechnikák a spektrumokban lévő irreleváns információk nem kívánt hatásainak kiküszöbölésére (52). E módszerek nélkül a zaj a valódi kalibrációs adatokkal együtt kerülne megtanulásra, ami túlillesztést eredményezne (53). A különféle eljárások alkalmazása a leírás által megkövetelt korrekciókon alapul, és a kutató feladata eldönteni, hogy mely javítások a leginkább elfogadhatóak. Az előfeldolgozás szükséges lépés a fajtafelismerés során, mivel a mérendő termény feldolgozatlan állapotban van, és emiatt a zajszint is magas lehet. Ugyanakkor szem előtt kell tartani, hogy a fentiekhez hasonló finomítási technikák alkalmazása a nyers adatok feldolgozásához értékes információk elvesztését eredményezheti.

A gyümölcs- és zöldségspektrumok vizsgálata az üvegedényekben lévő terményekkel az előfeldolgozás során nem hozott jobb eredményeket, mint a feldolgozatlan adatok, legalábbis abban a helyzetben, amikor a referenciáfüggő ortogonális részleges legkisebb négyzetek (OPLS) modell integrált OSC-szűrőt tartalmazott (59). A szűrők hatékonyságát egy válaszpermutációs készlet és teljesítménymátrix segítségével értékelték, azaz érzékenység, specifikusság, hatékonyság, fals negatív, fals pozitív és valódi negatív értékek alapján (59).

A spektrumok közvetlen összehasonlítása a keresztezések és átfedések miatt nem mindig lehetséges. Li és munkatársai (60) 300 almamintát mértek meg, majd 400 és 1021 nm közötti spektrumokat figyeltek meg, amelyek mindegyike nagymértékben hasonlított egymáshoz. A megfigyelt spektrumok különböző attribútumokból álltak, ezért később olyan dimenziócsökkentő technikákat alkalmaztak az osztályozáshoz, mint a PCA és a score plot. A dimenziócsökkentésre a nem releváns és felesleges spektrális változók kiküszöbölése miatt volt szükség (56). A fajtaazonosítással összefüggésben a PCA az almatermesztésre vonatkozó adatok 98%-át meg tudta őrizni (51), így ideális jellemzőkinyerési módszer.

Jha és munkatársai (55) az almafajták osztályozását vizsgálták, amelynek során 99,07%-os és 99,58%-os pontosságot értek el az Alphonso és a Banganapalli mangó fajták tekintetében. A score plot varianciája jellemzően a PC1, PC2 és PC3-tól kezdve csökken, emiatt a legtöbb

kutató az első kettőt (51) vagy az első hármat (50) használja. Az azonosítás folyamatában rejlő változatosságnak olyan score plotra van szüksége, ahol a hasonló eredményeket egy területre csoportosítják vagy klaszterezik. A PC2 és PC3 PC1-től független választása azonban hatékonyabb csoportosítással jobb modelleket eredményezhet (54). A változó kiválasztását genetikai algoritmus (GA), szukcesszív projekciós algoritmus (SPA), Monte Carlo-uninformatív változóelimináció (MC-UVE) és kompetitív adaptív újrasúlyozott mintavételezés (CARS) segítségével kell elvégezni, és ezeket a kiválasztott változókat egy későbbi szakaszban PLS-sel vagy többszörös lineáris regresszióval (MLR) kell kombinálni az előrejelzéshez (7, 54, 61). E technikák összehasonlító értékeléséből kiderült, hogy a CARS a legjobb teljesítményt nyújtja, és 0,9 közeli R_p -értékeket, valamint 0,5 közeli RMSEP-értékeket eredményez (7, 52, 56). A mintafelismerés olyan gépi tanulási módszerekkel is elérhető, mint a visszatérjedéses neurális hálózatok (BPNN), a támogató vektor gépek (SVM) és az extrém tanulási gépek (ELM). Ezek a módszerek alkalmasak a gyümölcsök fajtájának és földrajzi eredetének azonosítására a spektrum alapján. A tényezők és csoportok száma jelentős szerepet játszik a modell teljesítményében. Ha csökkentjük a lehetséges előrejelzések számát, akkor a modell pontossága nő. Egy különböző gyümölcsökből származó almákról készült tanulmányban például a modellek csak a völgy és a hegyvidék közötti osztályozásban teljesítettek jól - azaz egy kétváltozós modellben (51).

Az egyedi modellek sokkal pontosabbak, mint a többváltozós modellek (53), azonban a gyakorlatban többváltozós modelleket kell létrehozni. Amennyiben az egyedi modelleket tudatosan alkalmazzák a fajták előrejelzéséhez, az adatok túlillesztése és a nem megfelelő külső validálás jelenthetik a lehetséges buktatókat (53). A modellfejlesztéshez szükség van megfelelő gyakorló- és tesztkészletekre, és az az ideális, ha a megkülönböztetésre szánt fajták mindegyikéhez ugyanannyi gyakorló- és tesztkészlet tartozik. Amennyiben az egyes elemek száma nem azonos a gyakorló és tesztelő halmazok között, akkor kialakulhat az osztályok közötti kiegyensúlyozatlanság, amely a modellen belüli torzításokhoz vezethet. A kiegyensúlyozatlanságot többnyire a mintavételi eljárások csökkentésével és meghaladásával kezelik, habár más módszerek is lehetségesek; Zimmer és Schneider (59) például egy megfelelően kezelt modellt értékelt ki az osztályok kiegyensúlyozatlanságának feldolgozása nélkül. A modelleket roncsolásos módszerekből és laboratóriumi kiértékelésből származó adatokkal képzik le. Ezért a modell előrejelzési standard hibája (SEP) nagymértékben függ a laboratóriumi standard hibától (SEL). Sánchez és munkatársai (54) azt javasolták, hogy a SEL értékének ötszörösénél nagyobb vagy azzal egyenlő SEP értékek egy elfogadhatatlan minőségű

modellt mutatnak. Ebből az okból kifolyólag egy kvantitatív előrejelzésnél a jobb megértés érdekében jobb a mérések megismételhetőségét és reprodukálhatóságát megbecsülni. A kutatások többsége lényeges fenntartással bír, mivel hiányzik a külső értékelés (53), azonban az ilyen jellegű kiértékelésre szükség van egy olyan modell esetében, amelyet különböző területekről származó termesztett fajták és kultúrák előrejelzésére használnak.

A fajta és a kultúrnövény azonosításánál a modelleket a minőségi adatok osztályozására használják, és ez automatikusan lehetővé teszi a mennyiségi osztályozást is (54), tehát a különböző termőhelyek számos olyan belső tulajdonsággal bírnak, amelyek csak az adott fajtára jellemzőek (60). Eisenstecken és munkatársai (51), valamint Li és munkatársai (61) kísérleteket végeztek a különböző tengerszint feletti magasságon termesztett fajták meghatározására, ahol a mennyiségi változók (azaz a szénhidrát, a fruktóz, a szacharóz, a glükóz, a szorbit és a citromsav) kiszámítható eltéréseket mutattak a gyümölcs termesztési magasságától függően. Ezen ismeretek egy több származási helyről származó regressziós modellel kombinálva olyan minőségi paraméterek előrejelzésére alkalmasak, mint a pH, a keménység, a savasság és a nedvesség (52). Ugyanakkor az egy adott szezonban felállított modellek nem alkalmasak a következő szezonra a környezetben bekövetkező természeti eltérések miatt, ami viszont eltéréseket eredményez a termés tulajdonságaiban is (53).

Az érettség felismerése

A mezőgazdasági termények esetében az érettség kimutatása a NIR egyik fő alkalmazási területe. A különböző biológiai tényezők eltérései miatt számos terménynek nincsenek standard érettségi mutatói (63). Bár nincs mód a gyümölcs érettségének közvetlen előrejelzésére, az olyan tényezők, mint a szín, a szilárdság, a TSS, a TA és a szárazanyag gyakran összefüggésben vannak az érettséggel. Az említett paraméterek közül a szín és a színért felelős pigmentkomponenseket használják leggyakrabban az érettség megállapításához. Ez annak köszönhető, hogy a gyümölcsök érettségét hagyományosan a szín szabad szemmel történő vizsgálatával állapították meg. A paradicsom lipofil antioxidánsainak és a paradicsom keménységének egyik NIR-vizsgálata az érettséghez közeledve sajátos tendenciákat mutatott a spektrumvonalakban (64). Annak ellenére, hogy ez a tanulmány nem kapcsolódik közvetlenül az érettség előrejelzéséhez, az eredmények arra utalnak, hogy a NIR-ben rejlik potenciál az érettség kimutatására.

Az érettség felismerésére bevezetett legújabb technológiák közé tartoznak a fuzzy logikai algoritmusok. Chen és munkatársai (65) a pomelo érettségét egy a legkisebb négyzetekből álló

támogatási vektor regressziós (LSSVR) modelljével és a PCA algoritmus segítségével értékelték a "fuzzy-optimalizált" NIR adatok segítségével. A hagyományos PCA-hoz képest a fuzzy átalakított PCA gátolja a kezdeti adatok zaját és kiemeli a PC-t (65). Az LSSVR modellt az 1. egyenlet mutatja be.

$$Y = \omega \cdot \phi(x) + b \quad (1)$$

ahol ω = a regressziós súlymátrix; $\phi(x)$ = kernelfüggvény, amely radiális bázisfüggvény (RBF);
 b = küszöbérték.

A tanulmány a mintákat három kategóriába sorolta, nevezetesen a kalibráció, a validálás és a tesztelés kategóriájába. A vizsgálati minták nem kapcsolódnak a kalibráláshoz vagy a validáláshoz; így ezek segítségével ellenőrizhető, hogy a modell reprezentatív-e (65). Az érettségi paramétert - amely a szín (L, a, b) értékeket mutatja - tehát a modell segítségével kapjuk meg.

A 4. táblázat a NIRS különböző alkalmazásait mutatja be az érettség kimutatásához. Egy, a szőlő érettségének PLS-modellezéssel történő megállapítására vonatkozó tanulmány olyan változó tényezőket vett figyelembe, mint a fajta, a termőhely, az érettségi szint és az évszak (66). Egy olyan tanulmányhoz képest, amely az osztályozáshoz, validáláshoz és teszteléshez csak egyetlen forgalmazótól származó fajtákat vett figyelembe, az előbbi megközelítés praktikusabb és az ipar szempontjából relevánsabb. A 2016-os fajtán alapuló modell létrehozása és a 2017-es fajtával való validálása (66) bizonyította a modell hosszú távú robusztusságát. A lehető legjobb előrejelzési eredmények elérése az egyes vizsgált paraméterek esetében különböző előfeldolgozási technikák alkalmazását tette szükségessé; néhány példa: MSW+MSC a TSS esetében és SNV a TA esetében (66).

4. táblázat: A különböző mezőgazdasági termények érettségének kimutatása NIRS segítségével

Termény	Spektrális tartomány	Szoftve r	Minták száma	Pontosság	Megállapítások	Hivatkozás
Szőlő	800-2500 nm	OPUS v. 7.2	267+71	$R_p^2 = 0,28-0,77$ RMSEP = 0,14-7,83	A TSS és a TA paraméterek képesek megkülönböztetni a szőlőfajtákat. A fogyasztói igények és	(66)

					az érzékszervi paraméterek korrelálnak egymással.	
Avokádó	940-1798 nm	Latentix v. 2.12	>10000	$R^2 = 0,732$ RMSEP = 1.83	Egy hordozható NIR készülék teljesítményét értékelték. Külső validálást végeztek különböző évszakokban.	(67)
Mangó	306-1140 nm	Unscrambler v. 10.3	149	$R^2_{cv} = 0,84-0,87$ RMSE _{cv} = 139%	A PLS-előrejelzés jól működött az SSC és a DM esetében. A szilárdság kalibrálása gyenge volt.	(63)
Görögdi- nye	802-805 nm	OOIBase32, Unscrambler v. 9.7, MATLAB v. 7.10 R2010b	200	76.7–85.1%	RPP és NDIP technikákat alkalmaztak. Túlértséget állapítottak meg.	(68)
Paradicsom	285-1200 nm	Model Builder v. 1.1.0.10 5, SAS v. 9.2		$R^2 = 0,67-0,86$ RMSCV = 0,64–1,082	A lipofil antioxidánsok mérési eredményei csak részben voltak helyesek, és roncsolásos mérési technikákat alkalmaztak. Az online mérés megfelelő fejlesztésekkel megvalósítható.	(64)
Banán	400-1000 nm		45	$R^2 = 0,89$ RMSE = 0,000598	A klorofilltartalmat előrejelezték. A héj	(69)

					színe az érettséggel függött össze.	
Pomelo	400-2500 nm		168	$R_T = 0,893-0,912$ $RMSE_T = 0,87-7,28$	Fuzzy logikai alkalmazás LSSVR-ben RBF kernellel. Online minőségmeghatározás Vis-NIR segítségével lehetséges.	(65)
Pomelo	380-2520 nm	Quality Spec Pro		$R_p = 0,913-0,997$ $RMSE_p = 0,59-5,82$	A MOPLEC előfeldolgozást és az MWPLS-t robusztus modell kialakításához kombinálták. Az L^* , a^* és b^* színparaméterek közel álltak a LabMeas értékekhez.	(70)
Kékáfonya	400-2500 nm	Sisvar 3.0, IBM SPSS Statistics 19	300	$R = 0,714-0,970$	Az IAD-változások és az antocián szintek összefüggést mutattak.	(71)
Alma	200-1100 nm	Spectra wiz	170	93,27–99,62%	Az alma színe és érettsége összefüggést mutatott. Az előrejelzést ANN-SA segítségével végezték el, és sikeres volt.	(72)
Kiwi	729-975 nm	SAS v. 8.2, F-750 Model Builder	100	$R^2 = 0,73$ $R = 0,48-0,74$	A szárazanyag és az SSC értékét a NIRS segítségével lehet megbecsülni. A kiwi bogyómölcs eheto minőségét az éretlen állapotból kiindulva jósolták meg.	(73)

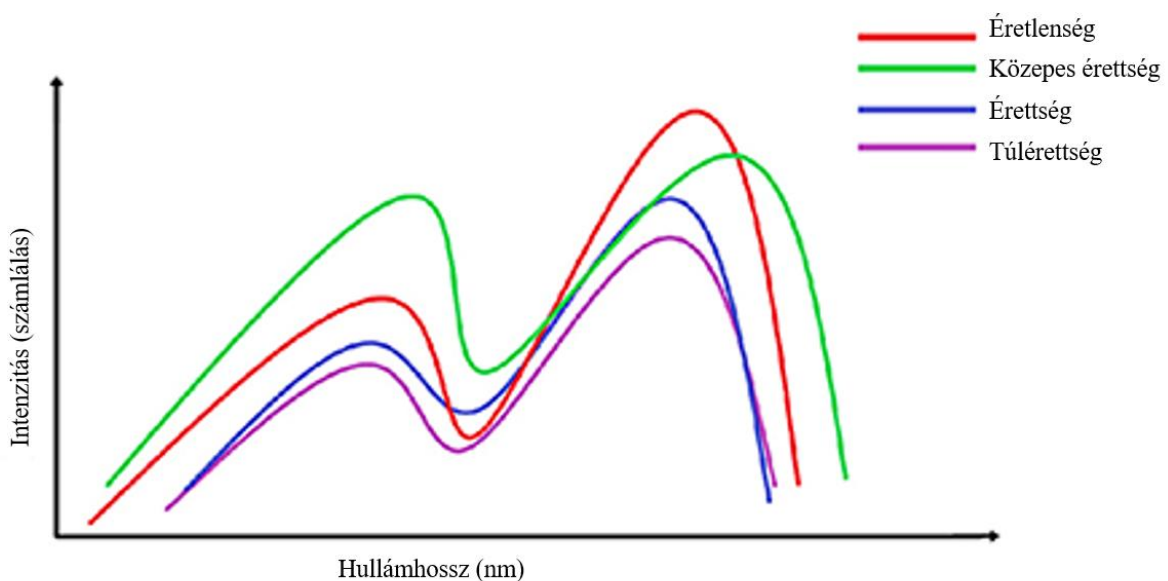
Mangó	700-990 nm	Spectras uite, XLSTA T 2014.1, SAS v. 9.4	1200	70–72%	Az érettség a DM és a TSS NIR-mérések segítségével határozható meg.	(74)
Ananász	740 - 1070 nm	MATLA B v. 9.5.0	90	100%	Az organikus és a szokványos termesztést tökéletesen meghatározták.	(25)

TSS, Összes oldható szilárd anyag; TA, Titrálható savtartalom; RMSEP, Az előrejelzés átlagos négyzetes hibája; RMSECV, A keresztellenőrzés átlagos négyzetes hibája; PLS, Részleges legkisebb négyzetek; SSC, Oldható szilárdanyag-tartalom; DM, szárazanyag; RPP, a csúcs1 és a csúcs2 közötti intenzitás aránya; NDIP, a csúcs normalizált különbség intenzitása; LSSVR, Legkisebb négyzetek támogatási vektor regressziója; RBF, Radiális ó bázisfüggvény; MOPLEC, Módosított optikai úthosszbecslés és korrekció; MWPLS, Mozgó ablakos részleges legkisebb négyzetek; ANN-SA, Mesterséges neurális hálózat szimulált lágyítási algoritmus.

A kiwi és a mangó érettségét a szárazanyag-tartalom (DM) és a szilárd oldhatóanyag-tartalom (SSC) mérésével jelezték előre (63, 73). Annak ellenére, hogy az SSC használható az érettség kimutatására a mezokarpiumban lévő, cukorra átalakult keményítő mérésével, a mangó esetében ez a módszer nem javasolt (63). Az avokádogyümölcsök egy vizsgálata az érettséget a nedvességtartalom (MC) meghatározásával jelezte előre, azonban ez a módszer a gyenge spektrumok miatt nem volt megfelelően alkalmazható (67). Kezdetben a vizsgálat a héj gyenge penetrációja miatt rosszul működött, ám a héj eltávolítása után a kísérlet eredményessége sokat javult. Ez arra enged következtetni, hogy a NIR nem használható minden célra. A gyümölcshússal és gyümölcshús nélkül végzett kiértékeléseknél ugyanannyi rejtett változót határoztak meg, ám az előbbinél az R^2 -érték 0,6 volt, ami később 0,8-ra javult (67). Ezen problémák tükrében az avokádóhoz hasonló tulajdonságokkal rendelkező gyümölcsöknél a kapcsolódó kihívásokkal szükséges foglalkozni, továbbá megfelelő módszereket kell kidolgozni.

A kémiai vegyületek viselkedésének változékonysága miatt a hőmérséklet közvetlen hatással van a spektrális adatokra. Ennélfogva a hőmérsékletet a mérés előtt stabilizálni kell, ahogy azt a kiwi spektrumai esetében is tették (73). Kim és munkatársai (73) szintén értékelték a Ca-kitozán bevonat alkalmazását a kiwi gyümölcsökön és annak az érettségre gyakorolt hatását. Az érési folyamat ütemének csökkenése miatt a NIRS alkalmazása növelni tudta az SSC-

tartalmat (73). A modell képes volt az érettség előrejelzésére, az R^2 érték 0,73 volt (73), amely elfogadható eredmény. Érdekes módon ezek az eredmények azt mutatják, hogy a bevonatok alkalmazása megváltoztathatja a spektrális viselkedést a fotonok és az új kémiai vegyületek közötti kölcsönhatás miatt. Ugyanakkor az algoritmusok és a feldolgozási technikák alkalmazásának köszönhetően az azonos vegyületekkel bevont gyümölcsök mindegyike csak más változók, például az SSC és a DM tekintetében mutat spektrális különbségeket, feltéve, hogy a bevonat felvitele egyenletes (3. ábra).



10. ábra: A különböző érettségi fokú termékek spektrális mintázata

Saputro és munkatársai (69) banánérettségi vizsgálatait a klorofilltartalmat az érettség meghatározásával hozzák összefüggésbe. A PCA, PCR és SVM kombinációja 0,000598%-os RMSE értéket eredményezett, amely kedvező teljesítményre utal. A klorofill sávok a 680-700 nm-es tartományban helyezkednek el, amit NIRS segítségével határoztak meg (69). A mangóban az SSC meghatározása a 699-999 nm-es tartományban, a DM pedig a 699-981 nm-es tartományban történt, ami a keresztellenőrzés során $R^2 = 0,874-0,87$ értéket eredményezett (63).

Ribera-Fonseca és munkatársai (71) egy nem invazív eszköz lehetőségeit vizsgálták a gyümölcs betakarítási időpontjának előrejelzésére az érési paraméterek figyelembevételével. A készülék az abszorbancia-különbség indexét (IAD) méri, amely magas korrelációt mutat a gyümölcs minőségi paramétereivel (71). A megközelítés nem tudott hagyományos algoritmusokat használni, ezért csak az IAD-értékekre támaszkodik. Polinar és munkatársai (74) a mangó

érettségi vizsgálatai során olyan osztályozási modellt alkottak, amely képes az érett és éretlen gyümölcsök előrejelzésére. A virágzást követő napok (DAFI) alapján határozták meg a küszöbértéket úgy, hogy a lebegők aránya $<25\%$. Az érett stádium előrejelzése ezen technikák segítségével képes volt a mesterségesen érlelt, íztelen gyümölcsök felismerésére (74). Ezen újítások felhasználhatóak az automatizált osztályozás létrehozására, a jobb minőségű gyümölcsök azonosítására és a piaci csalások megelőzésére.

A görögdinnye érettségének előrejelzését a csúcsok arányának (RP) módszerével végezték. Az RP akkor alkalmazható, ha két csúcs intenzitása a gyümölcs érésével ellentétesen változik. A görögdinnye esetében az 1. és a 2. csúcs intenzitásának arányának kiszámítása után a csúcsok normalizált intenzitáskülönbségét (NDIP) alkalmazták a gyümölcs érettségi stádiumának meghatározására (68). Az SPA szukcesszív projekciós algoritmust (SPA) használták a jellemzők kinyerésére, az osztályozáshoz pedig a legkisebb négyzetű támogató vektor gépet (LS-SVM) (68). Az 1. és 2. csúcs közötti intenzitás arányának (RPP) és az egyes érettségi fokozatok közötti ábrázolása alkalmas volt a gyümölcsök különböző érettségi fokozatok szerinti osztályozására (68). Így az RPP önmagában megkülönböztette a különböző érettségi stádiumokat, anélkül, hogy PCA vagy LDA vizsgálatot végeztek volna a legnagyobb eltérésekkel rendelkező plotokhoz. A helyes osztályozási arány (CCR) a korrigált RPP értékek esetében volt a legmagasabb (88,1%), az NDIP (85,1%) esetében; míg a legalacsonyabb az LS-SVM (76,7%) esetében (68). A CCR magas szintje célszerű, ugyanakkor a hibaparamétereket is értékelni kell, amiről ebben a kutatásban nem volt szó.

Több tanulmány az érettség előrejelzésére PLS-t használó SSC- és DM-modellekre támaszkodik. A spektrumanalízis következő szintre emelése különböző olyan alkalmazások használatát igényli, mint például a robotikában használt vagy videofilmek feldolgozására alkalmas rendszerek. A Fuji almák kapcsán végzett egyik vizsgálatban NIR spektrális elemzést alkalmaztak videóképekre az érés kimutatására (72). A mennyiségi adatok és a korreláció használata helyett inkább szakértői paneleket választottak az érettség meghatározására. A tanulmányban használt adatfeldolgozó számítógép a 2020-as szabványokhoz képest „belépő szintű” specifikációkkal rendelkezett, ami az elemzés megfizethetőségének bizonyítéka, szemben a „high-end” specifikációkkal. Az osztályozási modell a szimulált lágyítással (SA) kombinált ANN volt. Az SA különböző vektorokat küld az ANN-nek, és így találja meg a legmegfelelőbb ANN-struktúrát (72). A módszer az 535-560 nm, a 835-855 nm, illetve a 950-975 nm hullámhosszúságú tartományokat használva kiválóan teljesített, 90% feletti pontossággal (72).

Chen és munkatársai (70) a Shatian pomelo érettségét a szín és az érettségi szintek korrelációjával állapították meg. Hagyományos előfeldolgozási módszerek helyett a az optikai úthossz változását korrigáló többszörös hatáskorrekciót (MOPLEC) alkalmazták (70). Ez a módszer jobban teljesített a Savitzky-Golay-simításhoz képest, és javította a jel-zaj arányt (70). Az előfeldolgozás után mozgóablakos részleges legkisebb négyzetek alkalmazása (MWPLS) következett, amely a PLS-modellezéshez magas jelű hullámhossz-kombinációt talált (70). Az MWPLS a szomszédos hullámhosszok helyét és számát változtatja, hogy megtalálja az optimális kombinációt a PLS-modellezéshez (70). A tanulmány összehasonlította a nyers adatokat, az SG-feldolgozott és a MOPLEC-feldolgozott adatokat, és megállapította, hogy az R_v a MOPLEC-feldolgozott adatoknál volt a legmagasabb, és az SG-feldolgozott adatoknál azonos vagy jobb értékeket mutatott (70). Az új módszerek általánosságban jobban teljesítettek a pomelo érettségének színből történő előrejelzésében. Ezért ezeket a technikákat más gyümölcsök esetében is meg kell vizsgálni annak megállapítása érdekében, hogy a szín előrejelzése számukra is tovább fejleszhető-e.

Sérülések detektálása

A friss és minimális mértékben feldolgozott élelmiszerek iránt napjainkban nagy az érdeklődés, ezért egyre nagyobb a tétje annak, hogy az árukat minimális sérüléssel szállítsák. A megfelelően válogatott és kevésbé sérült élelmiszerek a fogyasztók elégedettségét is növelik. Ugyanakkor a mezőgazdasági termények terén nagy a változatosság: Az olyan gyorsan romló áruk, mint a paradicsom, 48 órán belül (75) is mutathatják a zúzódások vagy sérülések hatásait, míg az alma esetében akár több mint 50 nap is eltelhet (8), mire a külső felületen a sérülések bármilyen jele megjelenik. A betakarított terményeket haladéktalanul értékelni kell a sérülések szempontjából, hogy az árut átvehessék, majd értéknövelő termékeket kapjanak. Az értékelési időszak rövideje miatt számos élelmiszergyártó és -kutató olyan roncsolásmentes technikákat alkalmaz, mint például a NIR-spektroszkópia, amelyekkel a károsodások a termékre gyakorolt hatások nélkül kimutathatóak (5. táblázat). Az áfonya belső zúzódása például már 30 perccel a mechanikai hatást követően kimutatható, az R^2 értéke 0,7 (76). Mivel meg tudjuk becsülni a minőségromlás időtartamát és a károsodás mértékét, a spektrális jelek jellemzése a sérülések vagy eltérések felismerése érdekében javíthatja a termények hasznosítását (75).

5. táblázat: A különböző mezőgazdasági termények sérülésének kimutatására szolgáló NIRS-alkalmazások összefoglalása

Termény	Spektrális tartomány	Szoftver	Minták száma	Pontosság	Megállapítások	Hivatkozás
Őszibarack	400–2500 nm 325–1100 nm 930–2548 nm	ENVI v. 4.6, Isuzu Optics	200	Zúzódások felismerése = 96,5, Hangminta felismerése = 97,5	Az SW-NIR alkalmasabbnak bizonyult, mint az LW-NIR. Továbbfejlesztett vízgyűjtő szegmens algoritmust használtak.	(61)
Kékáfonya	950–1650 nm	LabVIEW, MATLAB 2016b, ENVI v. 4.7	320	70,8%–100%	Az előrejelzés pontossága az idő múlásával nőtt. A kocsány oldalsó zúzódását nehéz kimutatni.	(76)
Alma	400–1100 nm	MATLAB R2017a, PLS Toolbox v. 8.6	550	Az előrejelzés korrelációja = 0,87, SEP = 5,8N	A visszaterjedési neurális hálózat a szórási profil bemeneti arányaival kombinálva megjósolta a gyümölcsök keménységét és az SSC-t.	(77)
Olajbogyó	1100–2300 nm	R v. 3.1.3 SNAP! V. 2.04	744	96%	A QDA jobbnak bizonyult az LDA-nál.	(78)
Alma	550–1650 nm	MATLAB R2017a, PLS Toolbox v. 8.6	430	92,9%	A különböző pozícióknak a spektrumok felvételére gyakorolt hatását vizsgálták. Az ekvatoriális tájolás	(9)

					bizonyult a legmegfelelőbbnek.	
Kiwi	400–1100 nm	MATLAB R2018a, PLS toolbox v. 8.7	129		A vízzel áztatott szöveteknél az SNV-eljárás után kisebb a sérülések kimutatásának lehetősége.	(10)
Kókuszdió	900–2500 nm	OPUS v. 6.5, Unscrambler v. 9.8	202	94,03%	A NIR-modell jobban teljesített a kókuszdióhéjon keletkezett repedések kimutatásában, mint az akusztikus módszer.	(13)
Alma	350–1100 nm	Unscrambler v. 10.5.1, MATLAB B R2017b, PLS Toolbox	393	$R^2=0,59$	A napsugárzás a klorofilltartalom változását okozza.	(8)
Cseresznye	350–2500 nm	RS3, MATLAB B 2011a, ENVI v. 4.6, ViewSpec Pro v. 6.2.0, Origin8 SR0, Unscrambler X10.1	300	93,3%	A szín, a keménység és az SSC összhangban volt a Vis-NIR fényvisszaverő képességgel. Az LS-SVM az SPA-val kombinálva kimutatta a mintán belüli sérüléseket.	(79)
Őszibarack	300–1150 nm	OMNIC v. 8.2, R	840	$R_p = 0,71-0,92$; $RMSEP = 0,17-20,34$	Az élettani mutatók és az abszorpciós spektrumok közötti	(80)

		Studio, SPSS v. 22.0, OriginPro 2017			korreláció jó. A GA- PLS teljesített a legjobban.	
Paradicsom	5555,56– 11111,11 nm	MATLAB 2016		83-97%	A PCA-LDA azonosította a <i>G. candidum</i> fertőzést	(75)

SW-NIR, rövidhullámú közeli infravörös; LW-NIR, hosszuhullámú közeli infravörös; SSC, oldható szilárdanyag-tartalom; QDA, négyzetes diszkriminancia analízis; LDA, lineáris diszkriminancia analízis; SNV, standard normálváltozó; LS-SVM, legkisebb négyzetű támogató vektor gép; GA-PLS, genetikus algoritmus részleges legkisebb négyzet; PCA-LDA, főkomponens analízis és lineáris diszkriminancia analízis.

A sérülés előrejelzése a külső héj, az úgynevezett epidermisz spektrális vizsgálatával történik. Mivel az epidermisz a gyümölcs szerkezetének „határfelülete”, bármilyen kedvezőtlen körülmény olyan kis mértékű változásokat okozhat (75), amelyek az 500-900 nm-es (azaz szabad szemmel nem látható) tartományban nem feltétlenül láthatóak (79). Az ilyen változások azonban NIR-spektroszkópiával kimutathatók. Néhány változás, például az oxidatív barnulás azonban a jelentős időbeni csúszás miatt emberi szemmel is kimutatható (76).

A spektrumfeldolgozáshoz szükséges bemeneti adatok gyűjtése tekintetében többféle beállítást vizsgáltak, például 90 fokos (10) vagy 120 fokos (13) elforgatással történő mérést, az epidermisz érintését (8), 360 fokos mérést (9), valamint több különböző ponton történő mérést és az értékek átlagolását (75). A különböző tájolásban tartott almák összehasonlítása során a mérésekhez az almák ekvatoriális irányban történő tájolása mutatta a legjobb lehetőséget a hibák felismerésére (9). Ez valószínűleg azért van így, mert ez az elhelyezkedés nagyobb mennyiségű görbült felületet tesz szabaddá, így érzékenyebb méréseket eredményez. A kiwi esetében azonban az egyenlítői irányban történő igazítás adta a legrosszabb eredményeket a hűtési sérülések kimutatásában; ebben az összefüggésben a bibeszár forradásáig futó igazítással jobb eredmények születtek, valószínűleg azért, mert a hűtési tünetek a gyümölcs végén kezdődnek (10). Áfonya hiperspektrális elemzése során megfigyelték, hogy a kocsányvéget tévesen minősítették zúzódásnak; ezeket az adatokat ki kellett zárni a hamis pozitív arány csökkentése és a 0,1-ről 0,13-ra csökkentett RMSEP csökkentése érdekében (76). Ezért a gyümölcsök jellemzőinek mérésére szolgáló NIR-rendszerek tervezésénél figyelembe kell venni mind az észlelni kívánt hibát, mind az elemzéshez használt algoritmust.

A 6 szás egyedi mérőrendszer a hagyományos egypontos mérőrendszerénél (9) jobban teljesített, mivel a pontosság 0,8%-ról 1,7%-ra nőtt, amennyiben átlagos spektrumokat használtak. A méréshez nem létezik egyetlen, legmegfelelőbb módszer, mivel mindegyik méréshez különböző módszereket alkalmaznak az utófeldolgozás során. Az újrendezés jelentős mennyiségű időt vehet igénybe, ami egy kereskedelmi online rendszer kifejlesztéséhez szükséges, ezért nem megfelelő. A tanulás során ügyelni kell a kiugró értékek elkerülésére, melyek túlllesztési problémákat okozva befolyásolhatják az osztályozás pontosságát. A kiugró értékek a sérülések felismerésével összefüggésben a súlyosan sérült árukra vonatkoznak. Az ilyen esetek a PCA és az LDA algoritmusok alulteljesítését okozzák. A naptól megbarnult almák elemzése ezt az eljárást követte, hogy elkerülje az igen nagy mértékben sérült almákat, és így növelje az osztályozás pontosságát (8). A gyakorlatban ez azért lehetséges, mert a komoly sérülések szabad szemmel is láthatóak.

A kókuszdió repedt héjának felismerésére irányuló kísérlet során megállapították, hogy a PCA-diagram önmagában is használható a kiugró értékek megtalálására azokban az esetekben, amikor a kókuszdió héját nem lehet megfigyelni a burok belsejében (13). Ha a PCA analízis csak a tényleges hullámhosszokkal és a kiugró értékek kiküszöbölésével futtatjuk, akkor relevánsabb eredményeket kapunk. Az őszibarackon lévő horzsolások PCA elemzése - 781, 816, 840, 945 és 1000 nm-es effektív hullámhosszúsággal - PC4-et eredményezett magas kontraszthatású horzsolásos területtel (60). Az optimális hullámhossz kiválasztásakor a szukcesszív projekciós algoritmus (SPA) egy iterációs eljárást követ, minden egyes lépéssel eggyel több hullámhosszat ad hozzá, ami csökkenti a teljes spektrum kezelésének bonyolultságát (79). A cseresznye sérülésének elemzése során az SPA 3 változóval 96,6%-os pontosságot ért el, míg a teljes spektrum 2 változóval 96%-os pontosságot ért el kisebb számítási igénybevétel mellett (79). Az alacsonyabb számítási terhelhetőség, azaz a rövidebb elemzési idő alapvető fontosságú az egymásba épített rendszer tervezéséhez. A sérülések elemzése során azt tapasztalták, hogy a kontrollminták kismértékű sérülése is részlegesen befolyásolja az osztályozási modellt (76). Egy további fenntartás az, hogy a termények független, különböző régióból vagy szezonból származó készletét kell használni. A területi és időbeli változékonyság hatásait a szokásos osztályozási algoritmusok a nagy biológiai változékonyság miatt még mindig nem veszik figyelembe. Ahogy azt a különböző évekből származó almák napperzselődésének kimutatásánál láttuk, az előrejelzés RMSEP-je háromszorosa volt a kalibráció értékének (8). Az olyan osztályozási modellekkel, mint a legkisebb négyzetű támogató vektorgépekkel (LS-SVM), bizonyos mértékig figyelembe vehető

a biológiai variabilitás, a hőmérséklet-ingadozás és a mérési elhelyezkedés okozta változatosság (76).

A hibafelismerés osztályozása általában szubjektíven, azaz nem kvantitatív módon történik, ami azt jelenti, hogy emberi hiba előfordulhat (10). Amennyiben a spektrumokat kemometriai vagy kvalitatív méréshez használják, a spektrális mérés után azonnal értékelni kell őket, hogy csökkentsék az időtartam növekedése okozta hibákat (80).

A felismerési régiók feltérképezése az ép és sérült gyümölcsök spektrális visszaverődési görbéinek összevetésével történik. Mindkét minta esetében a reflexió és a hullámhossz közötti függvényt ábrázolják. A károsodás okozta változások miatt a diagram egyes részein lényeges eltérések lesznek az egészséges és a sérült termények között, és ezeket az egymástól elkülönülő sávokat használjuk az osztályozáshoz. Az osztályozási szempontokat tekintve az egészséges gyümölcsök homogénebbek, és ezért könnyebben osztályozhatóak, mint a sérült gyümölcsök (61). A spektrális előjelek egyenletes változása szükséges a pontos elválasztáshoz (52). Ez a módszer csökkenti a felesleges információkat és a feldolgozáshoz szükséges időt is. Az SW-NIR és a látható NIR esetében a hullámpontok helyett hullámsávokat választanak ki a vizsgálat bonyolultságának mérséklése érdekében (80).

Az őszibarackok sérülésének elemzése során megállapították, hogy az SW-NIR (7801-1000 nm) hatékonyabb, mint az LW-NIR (1000-2500 nm), mivel a sérült területen nagyobb a kontrasztkülönbség (60). A HIS analízise során hullámpontokat és releváns területeket (ROI) választanak ki az elemzéshez (61, 76), és ezekből a sérülések térbeli eloszlását is meg lehet állapítani. A kutatások során az epidermisz és a spektrális görbe jelentős hatásának korrelációját vizsgálva megfigyelhető, hogy a pigmentek, például a karotinoidok és a klorofill sávok ígéretes eltérési területeket szolgáltatnak, nevezetesen a klorofill tartomány esetében 670 nm körül (8, 10, 52), az antocianin tartomány esetében 550 nm (79), a karotinoid tartomány esetében pedig 740 nm (8). Előfordulhat, hogy az alacsonyabb klorofilltartalmú gyümölcsök hibái nem mutatnak jelentős eltéréseket a klorofillsávban, mint például az alma belső barnulása esetében (60). Az egyes kémiai kötések kiértékelése és a hullámpontok meghatározása lényegesen egyszerűbb, mint a TSS vagy az SSC jellemzése, amelyek egyenként több vegyületet tartalmaznak (80), és ezért a spektrális reflexiós diagramon való azonosításuk hosszadalmas. Az SSC hullámhossztartománya a cseresznye esetében 900-2500 nm volt, ami egy nagy hullámsáv, amely további elemzést igényel (79).

Az áfonya sérülésének egyik tanulmánya a darabolt gyümölcs pixelek elemzését követte a sérülés mértékének értékeléséhez (76), ami pontosabbnak bizonyult, mint a szubjektív emberi becslés. A vízben áztatott szövetek optikai tulajdonságainak vizsgálatára olyan kiegészítő technikák is elvégezhetők, mint az abszorbancia-különbség index (IAD) (10). Megállapították, hogy a víz abszorpciós csúcsai a 970 nm és az 1200 nm tartományban vannak. Egy sejtmembrán-szakadást okozó sérülés kétségtelenül jelentősen eltér ebben a régióban (76).

Az eddigi módszerek a PCA és a PLS módszereket követték, amelyek az összes változót előrevetítik a dimenziók csökkentése, illetve a szétválaszthatóság növelése érdekében. A PCA elemzés során az első néhány PC-t választják ki az osztályozáshoz a tulajdonságokban bővelkedő ábrák miatt. Mindazonáltal egy őszibarackon végzett sérüléselemzés során megállapították, hogy a PC4 a legjobb sérülés-megkülönböztetést biztosította, a hatékony hullámhosszok megtartásával (61). Ennélfogva a PCA elvégzésekor meg kell figyelni az eredményül kapott ábrát, hogy ellenőrizzük, hogy az megfelel-e a célnak. Az eredményesség további növelése érdekében nemlineáris optimalizálást kell végezni, például a genetikus algoritmus (GA) segítségével, ahol a jellemzők kiválasztása a természetes szelekciós eljárást követi (80). Az egyik online rendszerről készült tanulmány szerint a sávarányos módszer valós idejű és költséghatékony módszerként használható, 90% feletti pontossági rátával (76).

Mikrobiális/gombás szennyeződés kimutatása

A kertészeti termények hajlamosak a mikrobiális fertőzöttségre. Ebben a tanulmányban a mikrobiális és gombás szennyeződés korai felismerését vizsgálták a NIRS segítségével. A *Botrytis cinerea* és a *Colletotrichum acutatum* esetében a korai, 24 órán belüli roncsolásmentes felismerés megakadályozhatja a keresztfertőzést és a gazdasági veszteséget (81). Matulaprungsan és munkatársai (82) *E. coli* és *S. typhimurium* által okozott káposztafertőzéssel kapcsolatos vizsgálatai mesterséges körülményeket teremtettek, és különböző időtartamokra oltottak be mintákat aprított káposztaféléből, hogy feltárják a NIR-analízisben rejlő lehetőségeket (6. táblázat). A spektrumanalízis során ez a módszer a káposzta sejteinek kiáramlása és a baktériumok növekedése miatt spektrális eltolódást mutatott (82). Ez az eltolódó viselkedés az SG-derivált módszerrel (82), majd paradicsomos kórokozó elemzéssel (83) kezelték. A spektrumanalízis kimutatta, hogy a felaprított minták jobban elkülönültek az oltási idők között, mint az összedarált minták (82), ezért az elemzéshez az előbbi módszert választották. A felaprított káposzta tehát magasabb r -értéket ($r = 0,91-0,95$) eredményezett, amely megbízható előrejelzésre utal (82). Az *E. coli* kimutatása salátában szintén 100%-os pontosságot eredményezett a validációs készletben NIR-spektroszkópia és részleges legkisebb

négyzetek diszkrimináns analízis (PLS-DA) alkalmazásával (36). Az *E. coli* különböző koncentrációit öt különböző technikával, nevezetesen a PLS-DA, SVM, PCA, hierarchikus klaszteranalízis (HCA) és lágy független modellezés osztályanalízis segítségével (SIMCA) tudták osztályozni. A PCA és a HCA a durvább osztályozás felé tendált, míg a SIMCA és az SVM finomabb alcsoportokba sorolást eredményezett. Az *E. coli* vizsgálatot a bébispenótban PLS-módszerrel végezték (35). A kontaminációs vizsgálatok esetében a kiugró értékeket olyan jellemzőválasztási technikák segítségével kell kiküszöbölni, mint a PCA. Az *E. coli* és a *Z. rouxii* kimutatására végzett vizsgálatokban a PCA-t használták a kiugró értékek kiszűrésére (6, 35). A spenótban lévő *E. coli* kimutatása esetén a Q-residual-Hotelling T2-diagramot alkalmazták a kiugró értékek megtalálására (35). A PLS-DA felügyelt tanulóval az *E. coli*-t a 450-994 nm-es tartományban detektálták a predikciós készletben, 100%-os pontossággal (35). A modell nem volt képes előrejelezni a kezdeti szakaszokban a nem biztonságos mintákat. Ezért amikor a sejtkoncentráció elérte a 6,67 log TKE/ml értéket, akkor a statisztikai elemzés során nem biztonságosnak minősítették (35).

6. táblázat: A különböző mezőgazdasági terményekben lévő mikrobiális/gombás szennyeződések célzott kimutatására szolgáló NIRS-alkalmazások összefoglalása

Termény	Spektrális tartomány	Szoftver	Minták száma	Pontosság	Megállapítások	Hivatkozás
Paradicsom	400–2500 nm 325–1100 nm 550-1100 nm	Unscrambler v. 10.3, Spectra Suite	45	74-90%	<i>F. oxysporum f. sp. lycopersici</i> , <i>R. solani</i> , <i>Bacillus atrophaeus</i> és <i>P. aeruginosa</i> fertőzéseket mutattak ki.	(83)
Földieper	400-1000 nm	Unscrambler X v. 10.1, Spectral DAQ v. 2.1 for STATIS TICA 10	2700	97%	<i>B. cinerea</i> és <i>Collatotrichum acutatum</i> fertőzéseket mutattak ki. A BNN modell mutatta a legnagyobb előrejelzési pontosságot.	(81)

Kiwi	833–2500 nm	OPUS v. 5.5, MATLA B 2012a, Libsvm v. 3.20	352	$R^2 = 0.961–$ 0.999	<i>Z. rouxi</i> , <i>Hanseniaspora</i> <i>uvarum</i> és <i>C.</i> <i>tropicalis</i> fertőzéseket mutattak ki. Az SVM-modell meggyezett a telepszámlálási módszerrel.	(6)
Saláta	350–1100 nm	Spectra Wiz, Unscra mbler X10.3	200	87,1-83,39%	<i>E. coli</i> ATCC fertőzést mutattak ki. A SIMCA és az SVM felülmúlta a HCA-t, a PCA-t. Az <i>E. coli</i> - tartalom a kémiai összetételekkel változott, nem lineáris összefüggést mutatva.	(36)
Káposzta	700–1100 nm	CA Maker, Unscra mbler	20 g	$R = 0,47–0,91;$ $SECV =$ $0,45–1,17$	<i>E. coli</i> , <i>S.</i> <i>typhimurium</i> fertőzéseket mutattak ki. A felaprított levelek alkalmasabbak voltak a kimutatásra. Nem közvetlenül roncsolásmentes megközelítés.	(82)

BPNN, visszaterjedéses neurális hálózat; SVM, támogató vektorgép; SIMCA, lágy független modellezés osztályanalízissel; HCA, hierarchikus klaszteranalízis; PCA, főkomponens-analízis.

Egy kutatás a paradicsom kórokozóinak vizsgálatával 99,3%-os osztályozási pontosságot ért el az első és a második napon PCA, amelyet SVM osztályozó követett (83). Jóllehet a minták keveredésének megkezdésével magasabb osztályozási arányt értek el, az SVM besorolás nem tudott megbirkózni az azt befolyásoló több tényezővel. A nagy osztályozási arányok képesek voltak a *F. oxysporum* és a *R. solani* gombák, valamint az olyan baktériumok, mint a *B. atrophaeus* és a *P. aeruginosa* kimutatására (83).

A folyadékminták elemzése az áttetszőség miatt problémákat okoz (82). Mivel a NIR alapvetően fotonok és a megfelelő vegyületekre adott válasz révén működik, így ahogy egy oldat áttetszővé válik, nem lesz homogén, így megváltozik a spektrális válasz. E tekintetben az analízishez a SW-NIR technikát célszerű használni, mert nagyobb a penetrációs ereje és kisebb a vízsávok okozta interferencia (83). A saláták mikrobiális szennyeződéséről szóló tanulmány azonban azt mutatta, hogy a 970 nm-es O-H megnyúlásnak megfelelő vízsávot is figyelembe kell venni (36). A mikroorganizmusok sejtszámának különbségei molekuláris rezgésmintázatokat eredményezhetnek (36). A szilárd halmazállapotú salátamintákkal végzett vizsgálatok során olyan előfeldolgozási technikákat alkalmaztak a szórás hatások kiszűrése érdekében, mint a standard normálváltozó (SNV) és a többszörös szórás-korrektúra (MSC) (36). A szélső sávok kiküszöbölése és a spektrális felbontás növelése is elvégezhető ahhoz, hogy a kiemelkedő hullámhosszokra nagyobb hangsúlyt lehessen helyezni (36). A zaj és az előrejelzési pontosságot nem javító, nem kívánt információk beépítésének elkerülése érdekében előfeldolgozást végeznek. Az eltérés előrejelzési aránya 1,5 és 2,0 között van, ami jól kalibrált modelleket mutat (82).

Spektroszkópia segítségével elemezték a *B. cinerea* és a *C. acutatum* által okozott gombafertőzést földieperben. Az analízis során a 450-2 500 nm-es spektrális tartományban 3500 képpontot tartalmazó, releváns régiókat (ROI) vontak ki (81). Az adatok elemzése négy különböző osztályozási módszerrel történt, nevezetesen a visszaterjedési neurális hálózat (BPNN), a véletlen erdő (RF), a Naiv Bayes-féle osztályozó (NB) és a támogató vektor gépek (SVM) segítségével (81). Az osztályozás pontossága csökken, amint a gomba aktivitása lecsökken, amelyet a *C. acutatum*-mal fertőzött mintáknál 4 nap után figyeltek meg. Azonban ennek a technológiának is vannak kihívásai, nevezetesen a reprodukálhatósággal, a visszanyeréssel, valamint a páratartalom és a hőmérséklet negatív hatásaival kapcsolatos problémák (81). A befolyásoló tényezőkre való rálátást új eredménynek tekinthetjük, így a vizsgálat eredményei felhasználhatóak a módszerek javítására ezen tényezők ellenőrzése révén.

Az élesztőfajok gyakran okoznak romlást a magas cukortartalmú termékekben, különös tekintettel a gyümölcsstermékekre. A *Zygosaccharomyces rouxii* élesztőfajt, amely ezt a viselkedést mutatja, Niu és munkatársai (6) elemezték NIR segítségével a roncsolásmentes kimutatás érdekében. Előfeldolgozás során megállapították, hogy vagy SG simítás, vagy közvetlen ortogonális jelkorrekció (DOC) használható (6). A tanulmányban olyan kifejezéseket is bevezettek, mint a kimutatási határ (LOD) és a mennyiségi meghatározás határa (LOQ), amelyek a különböző előfeldolgozási és osztályozási algoritmusokkal együtt változtak (6). E

fogalmak a modell segítségével történő kimutatás végrehajtásához szükséges küszöbértékeként kezelhetőek. A modell képes volt a *Z. rouxii* kimutatására a meglévő standardok szerint, mind a kalibrációs, mind az előrejelző modellek esetében 85% feletti pontossággal (6).

Kihívások és jövőbeni lehetőségek

Az élelmiszer-feldolgozó iparban a NIR az egyik leggyakrabban alkalmazott roncsolásmentes módszer, mivel az elemzés gyorsan elvégezhető, a minta előkészítése minimális, költséghatékony és roncsolásmentes (84). Mint minden technológiai fejlesztés, a NIRS is rendelkezik azonban néhány hátrányos tulajdonsággal, amelyekkel számolni kell, például a laboratóriumi standard hiba (SEL), amely az előrejelzés standard hibáját okozza (SEP); a tájolási eltérések (29); és a biológiai variáció miatti teljesítményproblémák (8). A spektrális mérésekkel kapcsolatos kihívások mellett a regressziós eszközök, a gépi tanulás és az ANN is okozhatnak olyan problémákat, mint a túlillesztés (53), az alacsony korrelációs értékek (ami a módszert csak elsődleges osztályozóvá teheti), a mintafelismerési problémák (39), valamint a rendelkezésre álló rengeteg technika közül a legjobb technikák kiválasztásához szükséges megfelelő ismeretek hiánya.

Ezek a problémák már a kezdetektől fogva fennállnak, így már találtak rájuk megoldásokat is. Ez magába foglalja a paraméterek irányultságának tesztelését és optimalizálását, például az SSC meghatározásával végzett teszteléseket. Eszerint az SSC jelenléte miatt a szár és a kocsány függőleges helyzetének, a szár felfelé irányuló tengelyének megtartása volt a legjobb (29). A modell elkészítése során a feltételeknek ideálisnak kell lenniük a SEL és az SEP elkerülése érdekében, mivel egyetlen hiba is a modell teljesítményének romlását okozhatja.

A biológia változatosság továbbra is kihívást jelent, különösen akkor, ha a vizsgálatot a következő szezonban ugyanazon fajtából származó másik mintával végezzük. Ez főként azért van így, mert a gépi tanulás mintafelismerésen alapul, és minden egyes évben számos biológiai variáns változik, ami módosítja a termék paramétereit is. A probléma megoldásának eszköze a különböző szezonokból származó széles tanulási készlet, ami azonban gyakorlatilag lehetetlen.

A probléma megoldása érdekében az előfeldolgozási és előrejelzési fázisok módszereit úgy kell fejleszteni, hogy a mintához igazodjanak; máskülönben a modellt minden évben a teljesen előlről kell újratanítani. Ez azonban nem lehetséges, mivel tökéletes körülményekre lenne szükség a minimális hibával rendelkező modell újbóli létrehozásához. A neurális hálózatokkal kapcsolatos problémákat különböző technikákkal kezelik, például a BPNN, a GRNN és a PSO

segítségével, melyek jobban teljesítenek, ha egyenként kezelik őket (39). Noha ezek a módszerek bizonyos mértékig úgy tűnik, hogy megoldják a problémát, növelhetik a számítási ráfordítást. A világ különböző részeiről származó, felhőalapú modellek jelenthetik a megoldást, mivel a legtöbb fejlett rendszer ilyen beállításra épül.

Az ANN által támogatott gépi tanulás tűnik azonban a NIR-elemzés legmegfelelőbb módszerének, mivel itt elsősorban biológiai nyersanyagokról van szó. A mezőgazdasági termékek nagyszámú változékonyságát neurális hálózatok és mintafelismerő algoritmusok segítségével lehet kezelni. Az előre meghatározott algebrai egyenletek és az egzakt matematikán alapuló modellek feltehetően nem fognak működni ebben az esetben, hiszen nem képesek figyelembe venni a termények életciklusának alapjául szolgáló összetett jelenségeket. Míg a NIRS offline rendszerek sorozataként indult, a jelenlegi technológia jelenlegi ütemét demonstrálva a kutatók már online gyorsselemezési prototípusokat is képesek létrehozni.

A kutatás mátrixa azt mutatja, hogy az eddigi vizsgálatok jelentős részét almán végezték, és a mezőgazdasági termények hatalmas listája vár felfedezésre. Bebizonyosodott, hogy a kompakt kézi rendszerek kielégítően teljesítenek a NIRS-sel, és ez reményeket ébreszt a következő korszakra vonatkozóan, mikor ezek a technológiák remélhetőleg szélesebb körben elérhetővé és megfizethetőbbé válnak. Sok mai okostelefon rendelkezik infravörös jeltovábbítóval, fényérzékelő és távolságmérő (LIDAR) hardverrel és számos más érzékelővel. Rövidesen talán mindannyiunknak lesz NIRS-szkennerünk felhőalapú modellekkal, amelyek a biológiai különbségeket is figyelembe tudják venni.

Következtetés

A különböző molekuláris szintek egyedi tulajdonságai miatt a spektroszkópai módszerek egyedülálló jeleket tudnak létrehozni a vegyületekről. Ennek köszönhetően a mezőgazdasági termények mennyiségi és minőségi elemzése és a legmegfelelőbb tulajdonságokkal rendelkező termények kiválogatására lehetőség adódik. A spektrométeres adatokat regressziós elemzés és gépi tanulás végrehajtására alkalmazzák, amelyek segítségével a mezőgazdasági termények minősítése elvégezhető. A kezdetben offline tesztelésnek indult rendszer ma már képes a minták tönkretétele nélkül gyors online mérések végrehajtására. A számítástechnika fejlődése jelentősen javította a technológiát, különösebb gazdasági ráfordítás nélkül. A jelenlegi módszerek hibrid modellek létrehozásával rendelkeznek, amelyek a különböző technikák legjobb elemeinek felhasználásával és kombinálásával képesek bizonyos mértékig kezelni a

biológiai eltéréseket. Ettől a fejlődéstől eltekintve a technikai oldalon továbbra is vannak kihívások, azaz a termékek variációjának kezelése, az adatok feltárása és a modell optimalizálása. Továbbra is nagy szükség van egy olyan kompakt, roncsolásmentes technológiára, amely a legkülönbözőbb kertészeti termények jellemzésére alkalmas.

Hivatkozások

1. Madanayake NH, Hossain A, Adassooriya NM. Nanobiotechnology for agricultural sustainability, and food and environmental safety. *Q Assurance Safety Crops Foods*. (2021) 13:20–36. doi: 10.15586/qas.v13i1.838

[CrossRef Full Text](#) | [Google Scholar](#)

2. Yadav A, Kumar N, Upadhyay A, Singh A, Anurag RK, Pandiselvam R. Effect of mango kernel seed starch-based active edible coating functionalized with lemongrass essential oil on the shelf-life of guava fruit. *Q Assurance Safety Crops Foods*. (2022) 14:103–115. doi: 10.15586/qas.v14i3.1094

[CrossRef Full Text](#) | [Google Scholar](#)

3. Beegum PS, Pandiselvam R, Ramesh SV, Thube SH, Pandian TP, Khanashyam AC, et al. A critical appraisal on the antimicrobial, oral protective, and anti-diabetic functions of coconut and its derivatives. *Q Assurance Safety Crops Foods*. (2022) 14:86–100. doi: 10.15586/qas.v14i2.1040

[CrossRef Full Text](#) | [Google Scholar](#)

4. Kaavya R, Pandiselvam R, Mohammed M, Dakshayani R, Kothakota A, Ramesh SV, et al. Application of infrared spectroscopy techniques for the assessment of quality and safety in spices: a review. *Appl Spectrosc Rev*. (2020) 55:593–611. doi: 10.1080/05704928.2020.1713801

[CrossRef Full Text](#) | [Google Scholar](#)

5. Munawar AA, von Hörsten D, Wegener JK, Pawelzik E, Mörlein D. Rapid and non-destructive prediction of mango quality attributes using Fourier transform near infrared spectroscopy and chemometrics. *Eng Agric Environ Food*. (2016) 9:208–15. doi: 10.1016/j.eaef.2015.12.004

[CrossRef Full Text](#) | [Google Scholar](#)

6. Niu C, Guo H, Wei J, Sajid M, Yuan Y, Yue T. Fourier transform near-Infrared spectroscopy and chemometrics to predict *zygosacchomyces rouxii* in apple and kiwi fruit juices. *J Food Prot*. (2018) 81:1379–85. doi: 10.4315/0362-028X.JFP-17-512

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

7. Yan H, Xu YC, Siesler HW, Han BX, Zhang GZ. Hand-Held Near-Infrared Spectroscopy for authentication of fengdous and quantitative analysis of mulberry fruits. *Front Plant Sci.* (2019) 10:1–15. doi: 10.3389/fpls.2019.01548

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

8. Grandón S, Sanchez-Contreras J, Torres CA. Prediction models for sunscald on apples (*Malus domestica* Borkh.) cv. Granny Smith using Vis-NIR reflectance. *Postharvest Biol Technol.* (2019) 151:36–44. doi: 10.1016/j.postharvbio.2019.01.012

[CrossRef Full Text](#) | [Google Scholar](#)

9. Huang Y, Lu R, Chen K. Detection of internal defect of apples by a multichannel Vis/NIR spectroscopic system. *Postharvest Biol Technol.* (2020) 161:111065. doi: 10.1016/j.postharvbio.2019.111065

[CrossRef Full Text](#) | [Google Scholar](#)

10. Wang Z, Künnemeyer R, McGlone A, Burdon J. Potential of Vis-NIR spectroscopy for detection of chilling injury in kiwifruit. *Postharvest Biol Technol.* (2020) 164:111160. doi: 10.1016/j.postharvbio.2020.111160

[CrossRef Full Text](#) | [Google Scholar](#)

11. Ncama K, Tesfay SZ, Opara UL, Fawole OA, Magwaza LS. Non-destructive prediction of 'Valencia' orange (*Citrus sinensis*) and 'Star Ruby' grapefruit (*Citrus × paradisi* Macfad) internal quality parameters using Vis/NIRS. *Acta Hort.* (2018) 1194:1119–25. doi: 10.17660/ActaHortic.2018.1194.160

[CrossRef Full Text](#) | [Google Scholar](#)

12. Tan J, Li MF. Rapid and nondestructive identification of Belgian and Netherlandish trappist beers by front-face synchronous fluorescence spectroscopy coupled with multiple statistical analysis. *Q Assurance Safety Crops Foods.* (2021) 13:83–92. doi: 10.15586/qas.v13i1.839

[CrossRef Full Text](#) | [Google Scholar](#)

13. Noypitak S, Insabai W, Noknoi W, Karoojee S, Terdwongworakul A, Kobori H. Detection of cracked shell in intact aromatic young coconut using near infrared spectroscopy and acoustic

response methods. *J Food Meas Charact.* (2019) 13:1991–9. doi: 10.1007/s11694-019-00119-2

[CrossRef Full Text](#) | [Google Scholar](#)

14. Cortés V, Blasco J, Aleixos N, Cubero S, Talens P. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends Food Sci Tech.* (2019) 85:138–48. doi: 10.1016/j.tifs.2019.01.015

[CrossRef Full Text](#) | [Google Scholar](#)

15. Power A, Cozzolino D. How fishy is your fish? Authentication, provenance and traceability in fish and seafood by means of vibrational spectroscopy. *Appl Sci.* (2020) 10:1–13. doi: 10.3390/app10124150

[CrossRef Full Text](#) | [Google Scholar](#)

16. Pu YY, O'Donnell C, Tobin JT, O'Shea N. Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing. *Int. Dairy J.* (2020) 103:104623. doi: 10.1016/j.idairyj.2019.104623

[CrossRef Full Text](#) | [Google Scholar](#)

17. Dixit Y, Casado-Gavaldà MP, Cama-Moncunill R, Cama-Moncunill X, Markiewicz-Keszycka M, Cullen PJ, et al. Developments and challenges in online NIR spectroscopy for meat processing. *Compr Rev Food Sci Food Saf.* (2017) 16:1172–87. doi: 10.1111/1541-4337.12295

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

18. Lan W, Jaillais B, Leca A, Renard CMGC, Bureau S. A new application of NIR spectroscopy to describe and predict purees quality from the non-destructive apple measurements. *Food Chem.* (2020) 310:125944. doi: 10.1016/j.foodchem.2019.125944

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

19. Abasi S, Minaei S, Jamshidi B, Fathi D, Khoshtaghaza MH. Rapid measurement of apple quality parameters using wavelet de-noising transform with Vis/NIR analysis. *Sci Hortic.* (2019) 252:7–13. doi: 10.1016/j.scienta.2019.02.085

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

20. Theanjumpol P, Wongzeewasakun K, Muenmanee N, Wongsaipun S, Krongchai C, Changrue V, et al. Non-destructive identification and estimation of granulation in 'sai Num Pung' tangerine fruit using near infrared spectroscopy and chemometrics. *Postharvest Biol Technol.* (2019) 153:13–20. doi: 10.1016/j.postharvbio.2019.03.009

[CrossRef Full Text](#) | [Google Scholar](#)

21. Csorba V, Fodor M, Kovács S, Tóth M. Potential of fourier transformed near-infrared (FT-NIR) spectroscopy for rapid analysis of elderberry (*Sambucus nigra* L.) fruits. *Czech J Food Sci.* (2019) 37:21–8. doi: 10.17221/144/2018-CJFS

[CrossRef Full Text](#) | [Google Scholar](#)

22. Nturambirwe JFI, Nieuwoudt HH, Perold WJ, Opara UL. Non-destructive measurement of internal quality of apple fruit by a contactless NIR spectrometer with genetic algorithm model optimization. *Sci African.* (2019) 3:e0051. doi: 10.1016/j.sciaf.2019.e00051

[CrossRef Full Text](#) | [Google Scholar](#)

23. Hu R, Zhang L, Yu Z, Zhai Z, Zhang R. Optimization of soluble solids content prediction models in 'Hami' melons by means of Vis-NIR spectroscopy and chemometric tools. *Infrared Phys Technol.* (2019) 102:102999. doi: 10.1016/j.infrared.2019.102999

[CrossRef Full Text](#) | [Google Scholar](#)

24. Kanchanomai C, Naphrom D, Ohashi S, Nakano K, Theanjumpol P, Maniwar P. Nondestructive determination of quality management in table grapes using Near Infrared Spectroscopy (NIRS) technique. *ETP Int J Food Eng.* (2019) 5:63–7. doi: 10.18178/ijfe.5.1.63-67

[CrossRef Full Text](#) | [Google Scholar](#)

25. Amuah CLY, Teye E, Lamptey FP, Nyandey K, Opoku-Ansah J, Adueming POW. Feasibility study of the use of handheld NIR spectrometer for simultaneous authentication and quantification of quality parameters in intact pineapple fruits. *J Spectrosc.* (2019) 2019:e5975461. doi: 10.1155/2019/5975461

[CrossRef Full Text](#) | [Google Scholar](#)

26. Kashiwagi A, Kominami R, Nakayama S, Kishimoto N. Nondestructive prediction of oil yield through direct Near Infrared Spectroscopy measurements of fresh olives. in *2019 IEEE*

International Symposium on Olfaction and Electronic Nose (ISOEN). IEEE (2019). doi: 10.1109/ISOEN.2019.8823467

[CrossRef Full Text](#) | [Google Scholar](#)

27. Khodabakhshian R, Emadi B, Khojastehpour M, Golzarian MR. A comparative study of reflectance and transmittance modes of Vis/NIR spectroscopy used in determining internal quality attributes in pomegranate fruits. *J Food Meas Charact.* (2019) 13:3130–9. doi: 10.1007/s11694-019-00235-z

[CrossRef Full Text](#) | [Google Scholar](#)

28. Hasnah Ar N, Purwanto YA, Budiastra IW, Sobir S. Prediction of soluble solid content, vitamin C total acid and firmness in astringent persimmon (*Diospyros kaki* L.) cv. Rendeu using NIR spectroscopy. *IOP Conf Ser Mater Sci Eng.* (2019) 557:e012086. doi: 10.1088/1757-899X/557/1/012086

[CrossRef Full Text](#) | [Google Scholar](#)

29. Xia Y, Huang W, Fan S, Li J, Chen L. Effect of spectral measurement orientation on online prediction of soluble solids content of apple using Vis/NIR diffuse reflectance. *Infrared Phys Technol.* (2019) 97:467–77. doi: 10.1016/j.infrared.2019.01.012

[CrossRef Full Text](#) | [Google Scholar](#)

30. Huang Y, Lu R, Chen K. Assessment of tomato soluble solids content and pH by spatially-resolved and conventional Vis/NIR spectroscopy. *J Food Eng.* (2018) 236:19–28. doi: 10.1016/j.jfoodeng.2018.05.008

[CrossRef Full Text](#) | [Google Scholar](#)

31. Azadshahraki F, Jamshidi B, Rasooli Sharabiani V. Non-destructive determination of vitamin C and lycopene contents of intact cv. Newton tomatoes using NIR spectroscopy. *Yuz Yil Univ J Agric Sci.* (2018) 28:389–97. doi: 10.29133/yyutbd.423458

[CrossRef Full Text](#) | [Google Scholar](#)

32. Ncama K, Opara UL, Tesfay SZ, Fawole OA, Magwaza LS. Application of Vis/NIR spectroscopy for predicting sweetness and flavour parameters of 'Valencia' orange (*Citrus sinensis*) and 'Star Ruby' grapefruit (*Citrus x paradisi* Macfad). *J Food Eng.* (2017) 193:86–94. doi: 10.1016/j.jfoodeng.2016.08.015

[CrossRef Full Text](#) | [Google Scholar](#)

33. Lan W, Bureau S, Chen S, Leca A, Renard CMGC. Visible, near- and mid-infrared spectroscopy coupled with an innovative chemometric strategy to control apple puree quality. *Food Cont.* (2021) 120:107546. doi: 10.1016/j.foodcont.2020.107546

[CrossRef Full Text](#) | [Google Scholar](#)

34. Sans S, Ferré J, Boqué R, Sabaté J, Casals J, Simó J. Determination of chemical properties in 'calçot' (*Allium cepa* L.) by near infrared spectroscopy and multivariate calibration. *Food Chem.* (2018) 262:178–83. doi: 10.1016/j.foodchem.2018.04.102

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

35. Rahi S, Mobli H, Jamshidi B, Azizi A, Sharifi M. Visible/near-infrared spectroscopy as a novel technology for nondestructive detection of *Escherichia coli* ATCC 8739 in lettuce samples. (2019) 24–6. doi: 10.33422/worldcet.2019.10.285

[CrossRef Full Text](#) | [Google Scholar](#)

36. Rahi S, Mobli H, Jamshidi B, Azizi A, Sharifi M. Different supervised and unsupervised classification approaches based on visible/near infrared spectral analysis for discrimination of microbial contaminated lettuce samples: Case study on *E. coli* ATCC. *Infrared Phys Technol.* (2020) 108:103355. doi: 10.1016/j.infrared.2020.103355

[CrossRef Full Text](#) | [Google Scholar](#)

37. Jaiswal P, Jha SN, Bharadwaj R. Non-destructive prediction of quality of intact banana using spectroscopy. *Sci Hortic.* (2012) 135:14–22. doi: 10.1016/j.scienta.2011.11.021

[CrossRef Full Text](#) | [Google Scholar](#)

38. Kanchanomai C, Ohashi S, Naphrom D, Nemoto W, Maniwaru P, Nakano K. Non-destructive analysis of Japanese table grape qualities using near-infrared spectroscopy. *Hortic Environ Biotechnol.* (2020) 61:725–33. doi: 10.1007/s13580-020-00256-4

[CrossRef Full Text](#) | [Google Scholar](#)

39. Wu Y, Li L, Liu L, Liu Y. Nondestructive measurement of internal quality attributes of apple fruit by using NIR spectroscopy. *Multimed Tools Appl.* (2019) 78:4179–95. doi: 10.1007/s11042-017-5388-0

[CrossRef Full Text](#) | [Google Scholar](#)

40. Torres I, Sánchez MT, Entrenas JA, Garrido-Varo A, Pérez-Marín D. Monitoring quality and safety assessment of summer squashes along the food supply chain using near infrared sensors. *Postharvest Biol Technol.* (2019) 154:21–30. doi: 10.1016/j.postharvbio.2019.04.015

[CrossRef Full Text](#) | [Google Scholar](#)

41. Camps C, Gilli C. Prediction of local and global tomato texture and quality by FT-NIR spectroscopy and chemometric. *Eur J Hortic. Sci.* (2017) 82:126–33. doi: 10.17660/eJHS.2017/82.3.2

[CrossRef Full Text](#) | [Google Scholar](#)

42. Sharma S, Sirisomboon P. Feasibility on using NIR spectroscopy for the measurement of the textural parameters in mango. *IOP Conf Ser Earth Environ Sci.* (2019) 301:e012064. doi: 10.1088/1755-1315/301/1/012064

[CrossRef Full Text](#) | [Google Scholar](#)

43. Mohammadi-Moghaddam T, Razavi SMA, Sazgarnia A, Taghizadeh M. Predicting the moisture content and textural characteristics of roasted pistachio kernels using Vis/NIR reflectance spectroscopy and PLSR analysis. *J Food Meas Charact.* (2018) 12:346–55. doi: 10.1007/s11694-017-9646-7

[CrossRef Full Text](#) | [Google Scholar](#)

44. Sánchez MT, Entrenas JA, Torres I, Vega M, Pérez-Marín D. Monitoring texture and other quality parameters in spinach plants using NIR spectroscopy. *Comput Electron Agric.* (2018) 155:446–52. doi: 10.1016/j.compag.2018.11.004

[CrossRef Full Text](#) | [Google Scholar](#)

45. Feng L, Zhang M, Adhikari B, Guo Z. Nondestructive detection of postharvest quality of cherry tomatoes using a portable NIR spectrometer and chemometric algorithms. *Food Anal Methods.* (2019) 12:914–25. doi: 10.1007/s12161-018-01429-9

[CrossRef Full Text](#) | [Google Scholar](#)

46. Berardinelli A, Benelli A, Tartagni M, Ragni L. Kiwifruit flesh firmness determination by a NIR sensitive device and image multivariate data analyses. *Sensors Actuators A Phys.* (2019) 296:265–71. doi: 10.1016/j.sna.2019.07.027

[CrossRef Full Text](#) | [Google Scholar](#)

47. Cirilli M, Bellincontro A, Urbani S, Servili M, Esposito S, Mencarelli F, et al. On-field monitoring of fruit ripening evolution and quality parameters in olive mutants using a portable NIR-AOTF device. *Food Chem.* (2016) 199:96–104. doi: 10.1016/j.foodchem.2015.11.129

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

48. Yu X, Lu H, Wu D. 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.* (2018) 141:39–49. doi: 10.1016/j.postharvbio.2018.02.013

[CrossRef Full Text](#) | [Google Scholar](#)

49. Labaky P, Grosmaire L, Ricci J, Wisniewski C, Louka N, Dahdouh L. Innovative non-destructive sorting technique for juicy stone fruits: textural properties of fresh mangos and purees. *Food Bioprod Process.* (2020) 123:188–98. doi: 10.1016/j.fbp.2020.06.013

[CrossRef Full Text](#) | [Google Scholar](#)

50. Cortés V, Cubero S, Blasco J, Aleixos N, Talens P. In-line application of visible and near-Infrared diffuse reflectance spectroscopy to identify apple varieties. *Food Bioprocess Tech.* (2019) 12:1021–30. doi: 10.1007/s11947-019-02268-0

[CrossRef Full Text](#) | [Google Scholar](#)

51. Eisenstecken D, Stürz B, Robatscher P, Lozano L, Zanella A, Oberhuber M. The potential of near infrared spectroscopy (NIRS) to trace apple origin: Study on different cultivars and orchard elevations. *Postharvest Biol Technol.* (2019) 147:123–31. doi: 10.1016/j.postharvbio.2018.08.019

[CrossRef Full Text](#) | [Google Scholar](#)

52. Li X, Huang J, Xiong Y, Zhou J, Tan X, Zhang B. Determination of soluble solid content in multi-origin 'Fuji' apples by using FT-NIR spectroscopy and an origin discriminant strategy. *Comput Electron Agric.* (2018) 155:23–31. doi: 10.1016/j.compag.2018.10.003

[CrossRef Full Text](#) | [Google Scholar](#)

53. Zhang Y, Nock JF, Al Shoffe Y, Watkins CB. Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy. *Postharvest Biol Technol.* (2019) 151:111–8. doi: 10.1016/j.postharvbio.2019.01.009

[CrossRef Full Text](#) | [Google Scholar](#)

54. Sánchez MT, Torres I, de la Haba MJ, Chamorro A, Garrido-Varo A, Pérez-Marín D. Rapid, simultaneous, and *in situ* authentication and quality assessment of intact bell peppers using near-infrared spectroscopy technology. *J Sci Food Agric.* (2019) 99:1613–22. doi: 10.1002/jsfa.9342

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

55. Jha SN, Jaiswal P, Narsaiah K, Kumar R, Sharma R, Gupta M, et al. Authentication of mango varieties using Near-Infrared Spectroscopy. *Agric Res.* (2013) 2:229–35. doi: 10.1007/s40003-013-0068-4

[CrossRef Full Text](#) | [Google Scholar](#)

56. Li J, Zhang H, Zhan B, Wang Z, Jiang Y. Determination of SSC in pears by establishing the multi-cultivar models based on visible-NIR spectroscopy. *Infrared Phys Technol.* (2019) 102:103066. doi: 10.1016/j.infrared.2019.103066

[CrossRef Full Text](#) | [Google Scholar](#)

57. Su WH, Bakalis S, Sun DW. NIR/MIR spectroscopy in tandem with chemometrics for rapid identification and evaluation of potato variety and doneness degree. in *2019 ASABE Annual International Meeting.* (2019). doi: 10.13031/aim.201900120

[CrossRef Full Text](#) | [Google Scholar](#)

58. Wu X, Wu B, Sun J, Li M, Du H. Discrimination of apples using near infrared spectroscopy and sorting discriminant analysis. *Int J Food Prop.* (2016) 19:1016–28. doi: 10.1080/10942912.2014.971181

[CrossRef Full Text](#) | [Google Scholar](#)

59. Zimmer M, Schneider J. Near-infrared diffuse reflectance spectroscopy for discriminating fruit and vegetable products preserved in glass containers. *Croat J Food Sci Technol.* (2019) 11:104–12. doi: 10.17508/CJFST.2019.11.1.15

[CrossRef Full Text](#) | [Google Scholar](#)

60. Li C, Li L, Wu Y, Lu M, Yang Y, Li L. Apple variety identification using Near-Infrared Spectroscopy. *J Spectrosc.* (2018) 2018:1–6. doi: 10.1155/2018/6935197

[CrossRef Full Text](#) | [Google Scholar](#)

61. Li J, Chen L, Huang W. Detection of early bruises on peaches (*Amygdalus persica* L.) using hyperspectral imaging coupled with improved watershed segmentation algorithm. *Postharvest Biol Technol.* (2018) 135:104–13. doi: 10.1016/j.postharvbio.2017.09.007

[CrossRef Full Text](#) | [Google Scholar](#)

62. Wu G, Wang C. Investigating the effects of simulated transport vibration on tomato tissue damage based on vis/NIR spectroscopy. *Postharvest Biol Technol.* (2014) 98:41–7. doi: 10.1016/j.postharvbio.2014.06.016

[CrossRef Full Text](#) | [Google Scholar](#)

63. dos Santos Neto JP, de Assis MWD, Casagrande IP, Cunha Júnior LC, de Almeida Teixeira GH. Determination of 'Palmer' mango maturity indices using portable near infrared (VIS-NIR) spectrometer. *Postharvest Biol Technol.* (2017) 130:75–80. doi: 10.1016/j.postharvbio.2017.03.009

[CrossRef Full Text](#) | [Google Scholar](#)

64. Alenazi MM, Shafiq M, Alsadon AA, Alhelal IM, Alhamdan AM, Solieman THI, et al. Non-destructive assessment of flesh firmness and dietary antioxidants of greenhouse-grown tomato (*Solanum lycopersicum* L.) at different fruit maturity stages. *Saudi J Biol Sci.* (2020) 27:2839–46. doi: 10.1016/j.sjbs.2020.07.004

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

65. Chen H, Qiao H, Xu L, Feng Q, Cai K. A Fuzzy optimization strategy for the implementation of RBF LSSVR model in Vis-NIR analysis of pomelo maturity. *IEEE Trans Ind Informatics.* (2019) 15:5971–9. doi: 10.1109/TII.2019.2933582

[CrossRef Full Text](#) | [Google Scholar](#)

66. Daniels AJ, Poblete-Echeverría C, Opara UL, Nieuwoudt HH. Measuring internal maturity parameters contactless on intact table grape bunches using NIR spectroscopy. *Front Plant Sci.* (2019) 10:1–14. doi: 10.3389/fpls.2019.01517

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

67. Blakey RJ. Evaluation of avocado fruit maturity with a portable near-infrared spectrometer. *Postharvest Biol Technol.* (2016) 121:101–5. doi: 10.1016/j.postharvbio.2016.06.016

[CrossRef Full Text](#) | [Google Scholar](#)

68. Jie D, Zhou W, Wei X. Nondestructive detection of maturity of watermelon by spectral characteristic using NIR diffuse transmittance technique. *Sci Hortic.* (2019) 257:108718. doi: 10.1016/j.scienta.2019.108718

[CrossRef Full Text](#) | [Google Scholar](#)

69. Saputro AH, Juansyah SD, Handayani W. Banana (*Musa sp.*) maturity prediction system based on chlorophyll content using visible-NIR imaging. *Int Conf Signals Syst.* (2018) 2018:64–68. doi: 10.1109/ICSIGSYS.2018.8373569

[CrossRef Full Text](#) | [Google Scholar](#)

70. Chen HZ, Xu LL, Tang GQ, Song QQ, Feng QX. Rapid detection of surface color of shatian pomelo using Vis-NIR Spectrometry for the identification of maturity. *Food Anal Met.* (2016) 9:192–201. doi: 10.1007/s12161-015-0188-5

[CrossRef Full Text](#) | [Google Scholar](#)

71. Ribera-Fonseca A, Noferini M, Rombolá AD. Non-destructive assessment of highbush blueberry fruit maturity parameters and anthocyanins by using a visible/ near infrared (Vis/NIR) spectroscopy device: A preliminary approach. *J Soil Sci Plant Nutr.* (2016) 16:174–86. doi: 10.4067/S0718-95162016005000014

[CrossRef Full Text](#) | [Google Scholar](#)

72. Pourdarbani R, Sabzi S, Kalantari D, Karimzadeh R, Ilbeygi E, Arribas JI. Automatic non-destructive video estimation of maturation levels in Fuji apple (*Malus Malus pumila*) fruit in orchard based on colour (Vis) and spectral (NIR) data. *Biosyst Eng.* (2020) 195:136–51. doi: 10.1016/j.biosystemseng.2020.04.015

[CrossRef Full Text](#) | [Google Scholar](#)

73. Kim JG, Park Y, Shin MH, Muneer S, Lerud R, Michelson C, et al. Application of NIR-Spectroscopy to predict the harvesting maturity, fruit ripening and storage ability of Ca-chitosan

treated baby kiwifruit. *J Stored Prod Postharvest Res.* (2018) 9:44–53. doi: 10.5897/JSPPR2018.0257

[CrossRef Full Text](#) | [Google Scholar](#)

74. Polinar YQ, Yaptenco KF, Peralta EK, Agravante JU. Near-infrared spectroscopy for non-destructive prediction of maturity and eating quality of ‘carabao’ mango (*Mangifera indica* L.) fruit. *Agric Eng Int CIGR J.* (2019) 21:209–19.

[Google Scholar](#)

75. Skolik P, McAinsh MR, Martin FL. ATR-FTIR spectroscopy non-destructively detects damage-induced sour rot infection in whole tomato fruit. *Planta.* (2019) 249:925–39. doi: 10.1007/s00425-018-3060-1

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

76. Fan S, Li C, Huang W, Chen L. Detection of blueberry internal bruising over time using NIR hyperspectral reflectance imaging with optimum wavelengths. *Postharvest Biol Technol.* (2017) 134:55–66. doi: 10.1016/j.postharvbio.2017.08.012

[CrossRef Full Text](#) | [Google Scholar](#)

77. Lu R. Multispectral imaging for predicting firmness and soluble solids content of apple fruit. *Postharvest Biol Technol.* (2004) 31:147–57. doi: 10.1016/j.postharvbio.2003.08.006

[CrossRef Full Text](#) | [Google Scholar](#)

78. Moscetti R, Haff RP, Monarca D, Cecchini M, Massantini R. Near-infrared spectroscopy for detection of hailstorm damage on olive fruit. *Postharvest Biol Technol.* (2016) 120:204–12. doi: 10.1016/j.postharvbio.2016.06.011

[CrossRef Full Text](#) | [Google Scholar](#)

79. Shao Y, Xuan G, Hu Z, Gao Z, Liu L. Determination of the bruise degree for cherry using Vis-NIR reflection spectroscopy coupled with multivariate analysis. *PLoS ONE.* (2019) 14:1–13. doi: 10.1371/journal.pone.0222633

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

80. Du X, Long, L X, Yu, L Y, Zhou W. Hong, et al. Genetic algorithm optimized non-destructive prediction on property of mechanically injured peaches during postharvest storage

by portable visible/shortwave near-infrared spectroscopy. *Sci Hort.* (2019) 249:240–9. doi: 10.1016/j.scienta.2019.01.057

[CrossRef Full Text](#) | [Google Scholar](#)

81. Siedliska A, Baranowski P, Zubik M, Mazurek W, Sosnowska B. Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. *Postharvest Biol Technol.* (2018) 139:115–26. doi: 10.1016/j.postharvbio.2018.01.018

[CrossRef Full Text](#) | [Google Scholar](#)

82. Matulaprungsan B, Wongs-Aree C, Penchaiya P, Maniwara P, Kanlayanarat S, Ohashi S, et al. Feasibility of determination of foodborne microbe contamination of fresh-cut shredded cabbage using SW-NIR. *Agri Engi.* (2019) 1:246–56. doi: 10.3390/agriengineering1020018

[CrossRef Full Text](#) | [Google Scholar](#)

83. Abu-Khalaf N. Sensing tomato's pathogen using Visible/Near infrared (VIS/NIR) spectroscopy and multivariate data analysis (MVDA). *Palest Tech Univ Res J.* (2015) 3:12–22. doi: 10.53671/pturj.v3i1.35

[CrossRef Full Text](#) | [Google Scholar](#)

84. Fan G, Zha J, Du R, Gao L. Determination of soluble solids and firmness of apples by Vis/NIR transmittance. *J Food Eng.* (2009) 93:416–20. doi: 10.1016/j.jfoodeng.2009.02.006

[CrossRef Full Text](#) | [Google Scholar](#)

Magyar nyelvű összefoglaló

A spektroszkópiai adatok teljes mértékben alkalmasak a biológiai rendszerek modellezésére, többek között a kertészeti termények minőségi paramétereinek előrejelzésére. A széles hullámhosszspektrum adatai nem praktikusak, magas dimenziószámúak. Így rendkívül összetett, nehezen érthető modelleket kapunk. Mindemellett a széles spektrumon található zajok jelei is megtévesztőek lehetnek. A változószelekciós módszerek alkalmazása az optimális modell elérésének egyik hatékony módja. Egy érintésmentes spektrométer előnyeit kihasználva a 800-2500 nm közötti tartományban a közeli infravörös spektrális adatokat három almafajta: a "Golden Delicious", a "Granny Smith" és a "Royal Gala" sérüléseinek osztályozásához használták. Hat kiemelkedő gépi tanulási osztályozó algoritmust és két különböző kiválasztási módszert használtak a sérült és nem sérült gyümölcsök megkülönböztetésére a legrelevánsabb hullámhosszok segítségével. A kiválasztott hullámhosszok a 900 nm, az 1300 nm, az 1500 nm és az 1900 nm tartományokban csoportosultak. A teljes spektrumon alapuló modellekkel hibahatáron belül megegyező osztályozási pontosságot lehetett elérni a 4 fő hullámsávból kiválasztott legfeljebb 50 hullámhossz használatával. A legjobban lecsökkentett osztályozási modellek az LR és SVM gépi tanulási technikákon alapultak, amelyek a fajtától függően 0,7 és 0,9 közötti pontossági értékeket eredményeztek.

Az agrár alapú feldolgozóipari ágazatok a friss és minimális mértékben feldolgozott termékek előállítására haladnak. A termékek környezeti viszonyok között rendkívül romlandóak, ezért eltarthatóságuk korlátozott, azonban a hűtött tárolási körülmények meghosszabbítják az eltarthatóságot. A gyümölcsök és zöldségek osztályozását és válogatását azonnal a betakarítás után kell elvégezni a mikrobiális és kémiai romlásra való érzékenyséjük miatt. A különböző molekuláris szintek egyedi tulajdonságai miatt a spektroszkópiai módszerek egyedülálló jeleket tudnak létrehozni a vegyületekről. Így a mezőgazdasági termények mennyiségi és minőségi elemzésére lehetőség nyílik. A spektrométeres adatokat regressziós elemzés és gépi tanulás segítségével dolgozzák fel. Ma már lehetséges a minták roncsolásmentes, gyors online mérése. A számítástechnika fejlődése jelentősen javította a technológiát, különösebb gazdasági költség nélkül. A jelenlegi módszerek olyan hibrid modellek létrehozását foglalják magukban, amelyek a különböző technikák legjobb elemeinek felhasználásával és kombinálásával bizonyos mértékig képesek kezelni a biológiai változatosságot.

Magyar nyelvű összefoglaló idegen nyelvű fordítása

The data from spectroscopy are fully suitable for modelling biological systems, including the prediction of quality parameters of horticultural crops. The broad wavelength spectrum data are impractical, with a high number of dimensions. This leads to extremely complex models that are difficult to understand. In addition, the signal from noise in the wide spectrum can be misleading. The application of variable selection methods is an efficient way to achieve an optimal model. By taking advantage of a non-contact spectrometer, near-infrared spectral data in the 800-2500 nm range were used to classify damage in three apple varieties: 'Golden Delicious', 'Granny Smith' and 'Royal Gala'. They used six superior machine learning classification algorithms and two different selection methods to discriminate between damaged and non-damaged fruit using the most relevant wavelengths. The chosen wavelengths were clustered in the 900 nm, 1300 nm, 1500 nm and 1900 nm ranges. Classification accuracies within the same error range as the full spectrum models were achieved using up to 50 wavelengths selected from the 4 main wavebands. The classification models with the best reduction were based on LR and SVM machine learning techniques, which resulted in accuracy values between 0.7 and 0.9 depending on the species.

Agriculture-based manufacturing sectors are moving towards the production of fresh and minimally processed products. The products are highly degradable under environmental conditions and therefore have a limited shelf life, but refrigerated storage conditions increase shelf life. Classification and sorting of fruit and vegetables should be carried out immediately post harvest due to their sensitivity to microbial and chemical deterioration. Because of the unique properties of different molecular levels, spectroscopic methods can generate unique signals about compounds. This opens up the possibility of quantitative and qualitative analysis of agricultural crops. Spectrometric data are processed using regression analysis and machine learning. It is now possible to measure samples non-destructively and rapidly online. The development of computer technology has significantly improved the technology without any significant economic cost. The current methods involve the creation of hybrid models that can handle biological variability to some extent by using and combining the best elements of different techniques.

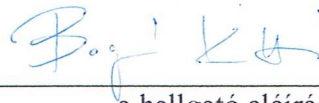
Szerzői/Eredetiség nyilatkozat

Szerzői nyilatkozat

Alulírott Bognár Kitti Annamária az Agrár- és természettudományi szakfordító szakirányú továbbképzés hallgatója kijelentem, hogy a „Feature Reduction for the Classification of Bruise Damage to Apple Fruit Using a Contactless FT-NIR Spectroscopy with Machine Learning - Az almatermésű gyümölcsök sérüléseinek osztályozására szolgáló tulajdonságok csökkentése az érintésmentes FT-NIR spektroszkópia és a gépi tanulás segítségével - Recent advancements in NIR spectroscopy for assessing the quality and safety of horticultural products: A comprehensive review - A NIR-spektroszkópia legfrissebb eredményei a kertészeti termények minőségének és biztonságának értékelésében: Átfogó áttekintés „ című képezőfordítás a saját munkám eredménye. Azon részeket, melyeket más szerzők munkájából vettem át, egyértelműen megjelöltem, s az irodalomjegyzékben szerepeltettem.

Ha a fenti nyilatkozattal valótlan állítottam, tudomásul veszem, hogy a Záróvizsga-bizottság a záróvizsgából kizár és záróvizsgát csak új dolgozat készítése után tehetek.

Budapest, 2024.04.27.



a hallgató aláírása

Konzulensi nyilatkozat

A dolgozat készítőjének konzulense nyilatkozom arról, hogy a Záródolgozatot/Szakdolgozatot/Diplomadolgozatot/Képesítőfordítást áttekintettem, a hallgatót az irodalmi források korrekt kezelésének követelményeiről, jogi és etikai szabályairól tájékoztattam.

A Záródolgozatot/Szakdolgozatot/Diplomadolgozatot záróvizsgán történő védésre javaslom / nem javaslom*.

A dolgozat állam- vagy szolgálati titkot tartalmaz:

igen nem*

Kelt: 2024. év április hó 19. nap


Konzulens