

Article

Vis/NIR Based Flexible Non-Destructive Sensing for Almonds

Tao Sun ^{1,†} , Han Wu ^{2,†}, Wei Liu ³, Ruina Yang ¹, Huimin Zhang ¹, Ju Lu ⁴, Jian Shen ⁴, Ruihua Zhang ² and Xinqing Xiao ^{2,*} 

¹ Institute of Quality Standard and Testing Technology for Agricultural Products, Xinjiang Academy of Agricultural Sciences, Urumqi 830091, China; suntao26@xaas.ac.cn (T.S.); 19199193018@163.com (R.Y.); zhang_hm618@163.com (H.Z.)

² College of Engineering, China Agricultural University, Beijing 100083, China; mayorbwh@163.com (H.W.); bs20243070753@cau.edu.cn (R.Z.)

³ National Institute of Metrology, Beijing 100029, China; liuwei@nim.ac.cn

⁴ Wuxi Institute of Inspection, Testing and Certification, Wuxi 214000, China; diane1980@163.com (J.L.); 18951506375@189.cn (J.S.)

* Correspondence: xxqjd@cau.edu.cn

† These authors contributed equally to this work.

Abstract

A flexible visible/near-infrared (Vis/NIR) sensing system (FVNS) was developed for the non-destructive assessment of almond composition. Almonds from four distinct varieties were measured under non-contact conditions, and the acquired spectra were preprocessed using Savitzky–Golay (S–G) smoothing and standard normal variate (SNV). Based on the spectral data captured by the FVNS, random forest (RF) regression models were established to quantify protein and fat contents. The optimized RF models achieved prediction coefficients of determination (R^2_p) of 0.91 for protein and 0.86 for fat, with corresponding residual predictive deviation (RPD) values of 3.32 and 2.67, respectively. These results demonstrate that the FVNS possesses reliable quantitative capability and can accurately capture compositional variations in almonds while maintaining low cost, portability, and real-time wireless operation.

Keywords: Vis/NIR spectroscopy; flexible sensors; almonds; non-destructive testing; random forest

1. Introduction

Almonds are among the most significant tree nuts globally, with world production levels reaching approximately 1.5 million metric tons (kernel weight) in recent years, dominated primarily by orchards in California, USA, and the Mediterranean region [1]. Driven by a growing consumer shift toward plant-based diets and healthy snacking, global almond consumption has experienced a steady surge, necessitating more stringent and efficient quality control measures across the international supply chain [2]. As a nutrient-dense superfood, almonds are rich in proteins, unsaturated fatty acids, vitamins, and minerals that contribute to cardiovascular health and metabolic regulation [3–6].

In China, the rapid development of the almond industry in Xinjiang has led to increased demand for robust quality evaluation and compositional analysis [7,8]. The qualitative characteristics of almonds are determined by multiple physicochemical parameters, including protein content, fatty acid composition, moisture content, and tissue hardness, representing a complex physiological and chemical process. Among these factors, protein



Academic Editor: Aivars Ābolņiņš

Received: 25 January 2026

Revised: 24 February 2026

Accepted: 26 February 2026

Published: 26 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

and fatty acid contents are the key indicators for evaluating almond quality and nutritional value [9].

Traditionally, the determination of protein and lipid contents in almonds relies on standard wet-chemical analyses, such as the Kjeldahl method for nitrogen and Soxhlet extraction for fats [4,10–12]. While these methods provide high accuracy and serve as the ‘gold standard’ for nutritional profiling, they are inherently destructive, labor-intensive, and involve hazardous chemical solvents, making them unsuitable for rapid, large-scale screening in modern industrial supply chains. To improve these conventional physico-chemical testing methods, early researchers proposed quality evaluation techniques based on mechanical and acoustic responses. These methods infer internal structure and quality characteristics indirectly by measuring parameters such as hardness, elastic modulus, or impact vibration signals. Although such techniques reduce sample damage to some extent, they still require physical contact or external force application and therefore cannot achieve truly non-destructive testing in the full sense [13–16].

In recent years, the rapid and non-destructive detection of almond quality parameters has become a research hotspot. Researchers have proposed various emerging detection approaches, including acoustic analysis, computed tomography (CT), machine vision, and electrical signal-based methods, all of which have demonstrated promising application potential to varying degrees. Nevertheless, in practical use, these techniques still suffer from limitations such as relatively slow detection speed, high equipment cost, and insufficient real-time capability, making it difficult to meet the nut industry’s demand for efficient, accurate, and continuous quality monitoring [17–21].

In the field of agricultural produce quality inspection, visible/near-infrared (Vis/NIR) spectroscopy and hyperspectral imaging (HSI) technologies are increasingly becoming research hotspots due to their rapidity, non-destructiveness, reagent-free operation, and potential for online deployment [22–24]. Because almonds possess a complex chemical composition and microstructural features prone to strong wavelength-dependent scattering, spectral band overlap, and baseline drift, the raw spectral signals are often accompanied by noise and coupling effects. Therefore, before modeling, preprocessing strategies such as standard normal variate (SNV), multiplicative scatter correction (MSC), Savitzky–Golay (S–G) smoothing, and its first- and second-order derivatives have become routine steps to enhance the stability of spectrum-to-quality mappings [25]. In the model-building phase, beyond traditional partial least squares regression (PLSR), machine learning methods such as support vector machine (SVM), random forest (RF), and one-class classification models (e.g., DD-SIMCA) have also been widely applied in quality assessment tasks. For example, NIR spectroscopy combined with machine-learning classifiers such as SVM has been successfully applied to discriminate almonds from different geographical origins, typically achieving classification accuracies above 90% [26]. Short-wave infrared hyperspectral imaging (SWIR-HSI) together with DD-SIMCA and PLSR has also been applied for qualitative and quantitative analysis of peanut-adulterated almond powders, yielding 100% sensitivity and specificity with quantification errors below 1% [27]. In addition, hyperspectral imaging combined with PLS-DA has been shown to accurately identify 5–20% bitter almond admixture in commercial batches [28]. Reviews have further indicated that Vis/NIR spectroscopy and its imaging extension, HSI, can provide non-destructive, high-throughput, and low-cost quality control throughout the nut supply chain [29].

While Vis/NIR spectroscopy is increasingly becoming a research hotspot for non-destructive testing, most current handheld devices and compact multispectral systems predominantly rely on traditional rigid optical probes. These sensors often struggle with “in-bag” sensing, as the irregular orientation of almond kernels and the gap between packaging and the sensor lead to inconsistent contact geometry and significant light-path errors. The

central hypothesis of this research is that a flexible sensing interface can provide superior optical coupling with the curved and multi-layered surfaces of almond groups compared to such rigid designs. By achieving more consistent contact geometry, the flexible Vis/NIR sensor (FVNS) is expected to significantly mitigate surface scattering errors and light-path inconsistencies, thereby enhancing the predictive accuracy and robustness of quantitative models for protein and fat contents. Furthermore, unlike conventional benchtop systems that cost thousands of dollars, our wireless FVNS prototype achieves a total hardware cost of only \$150–\$200. This work thus positions the FVNS as a cost-effective, real-time solution specifically engineered for quality monitoring within the industrial nut supply chain.

In this study, a flexible and low-cost FVNS was developed for the rapid and non-destructive assessment of almond quality. The FVNS integrates dual spectral modules covering both visible and near-infrared regions, allowing the simultaneous acquisition of multi-channel reflection signals to capture the multidimensional spectral responses of almond samples. Designed primarily for quantifying protein and fat contents, the FVNS transmits the measured spectra wirelessly to a host computer for real-time acquisition and analysis. To minimize the effects of surface scattering and instrument noise, the spectral data from the FVNS were preprocessed using SNV transformation and first-derivative Savitzky–Golay smoothing, effectively removing baseline drift and enhancing subtle absorption features. A random forest (RF) regression model was then established to construct the quantitative relationship between the spectral information and the chemical composition of almonds, with model training performed offline under a cross-validation strategy to reduce overfitting. Experimental results demonstrated that the FVNS combined with the RF model achieved high prediction accuracy for both protein and fat contents, validating its feasibility and robustness for nut quality monitoring. This work aims to provide a solid theoretical and technical foundation for the development of intelligent, real-time, and low-cost quality assessment systems based on the FVNS.

2. Materials and Methods

2.1. Sample Preparation

A total of 2000 almonds originating from commercial orchards in Shache County, Xinjiang, were collected in late August 2025. The trees were cultivated under standard intensive agronomic management, including a planting density of 415 trees/ha, drip irrigation, and a balanced application of organic and mineral fertilizers. The collection covered four commercial cultivars: Nongpalei, Shache No. 4, Wanfen No. 18, and Shache No. 69. For clarity, these four cultivars are referred to as Varieties 1–4 throughout the manuscript. To ensure representativeness and uniformity, each cultivar contributed 500 almonds. Harvesting was performed at full maturity using mechanical tree shakers. All subsequent experimental procedures and data collection were conducted in October 2025. The samples were organized into 400 groups, with each group consisting of five almonds from the same cultivar, and individually packaged in resealable transparent polyethylene bags commonly used for laboratory sample handling. All almonds were dried, deshelled, and free of visible mechanical damage. To ensure experimental consistency, the selected almond samples were of comparable size. Following packaging, the samples were stored in a temperature-controlled laboratory environment at $20 \pm 2^\circ\text{C}$, protected from light and maintained in low-humidity conditions for 48 h to reach moisture equilibrium before the experimental assays were conducted.

2.2. Design and Configuration of the FVNS

The FVNS was specifically engineered for the purposes of this study as an integrated optoelectronic platform to facilitate high-precision spectral acquisition from almond sam-

ples. While the detection principle aligns with standard diffuse reflectance spectroscopy, the FVNS incorporates specific hardware optimizations to mitigate signal fluctuations caused by the high reflectivity and pronounced surface roughness of almonds. The core sensing architecture comprises 12 distinct optical channels, achieved by pairing two six-channel photodiode arrays (AS762X, AMS, Plano, TX, USA) with specialized Gaussian-type interference filters. This configuration provides a comprehensive spectral detection window spanning 400–900 nm. Illumination is provided by a coordinated LED array (Lumileds, San Jose, CA, USA), where the light-emitting and receiving components are positioned in a parallel orientation to ensure a homogeneous light field during non-contact measurements.

The physical backbone of the FVNS consists of a flexible, high-performance PI/Cu composite film, featuring a 10 µm conductive copper layer on a 40 µm polyimide substrate (Shenzhen Junlang Electromechanical Equipment Co., Ltd., Shenzhen, China). Precision circuit patterns were defined using a dual-pass ultraviolet nanosecond laser-etching process (Beijing Kaitian Technology Co., Ltd., Beijing, China). Key electronic components, including the sensing arrays and an AT25SF041-SSHD-B storage chip (Adesto Technologies, Santa Clara, CA, USA), were integrated onto the flexible substrate via precision soldering at approximately 220 °C.

To ensure robust operation in agricultural environments, the active sensor surface was encapsulated with a protective polydimethylsiloxane (PDMS; Dow Corning, Midland, MI, USA) layer, followed by a thermal curing cycle at 65 °C for 2.5 h. The control and communication subsystem is driven by an ESP32 microcontroller unit (MCU; Good Display Electronics, Dalian, China), which handles local data processing and facilitates Wi-Fi-based wireless transmission. For localized user feedback, a flexible electrophoretic display (GDEW029I6FD; Good Display Electronics, Dalian, China) provides real-time data visualization, while the ONENET cloud platform (CMIOT, Chongqing, China) serves as the primary gateway for remote data management. Notably, the total hardware material cost for this integrated FVNS prototype is approximately \$150–\$200, representing a significant cost reduction compared to conventional benchtop spectroscopy systems. Detailed optoelectronic specifications are provided in Table 1.

Table 1. Optical channels and excitation sources of the FVNS.

Channel	Center Wavelength (nm)	FWHM (nm)	Assigned Spectral Band	LED Excitation
Ch-1	450	40	Vis	5200 K LED
Ch-2	500	40	Vis	5200 K LED
Ch-3	550	40	Vis	5200 K LED
Ch-4	570	40	Vis	5200 K LED
Ch-5	600	40	Vis	5200 K LED
Ch-6	650	40	Vis	5200 K LED
Ch-7	610	20	NIR	2700 K LED
Ch-8	680	20	NIR	2700 K LED
Ch-9	730	20	NIR	2700 K LED
Ch-10	760	20	NIR	2700 K LED
Ch-11	810	20	NIR	2700 K LED
Ch-12	860	20	NIR	2700 K LED

2.3. Spectral Measurement Procedure

The systematic workflow for spectral acquisition is outlined in Figure 1. To eliminate potential spectral deviations arising from environmental fluctuations, all almond specimens underwent a 24 h acclimation period in an indoor setting under ambient lighting, ensuring that both temperature and humidity reached a stable equilibrium. Due to the inherent structural stability and minimal moisture content of almonds, their physicochemical properties

remain largely constant during short-term storage. Consequently, the entire experimental procedure was executed at a controlled room temperature (approximately 20 °C) without the need for auxiliary environmental regulation.

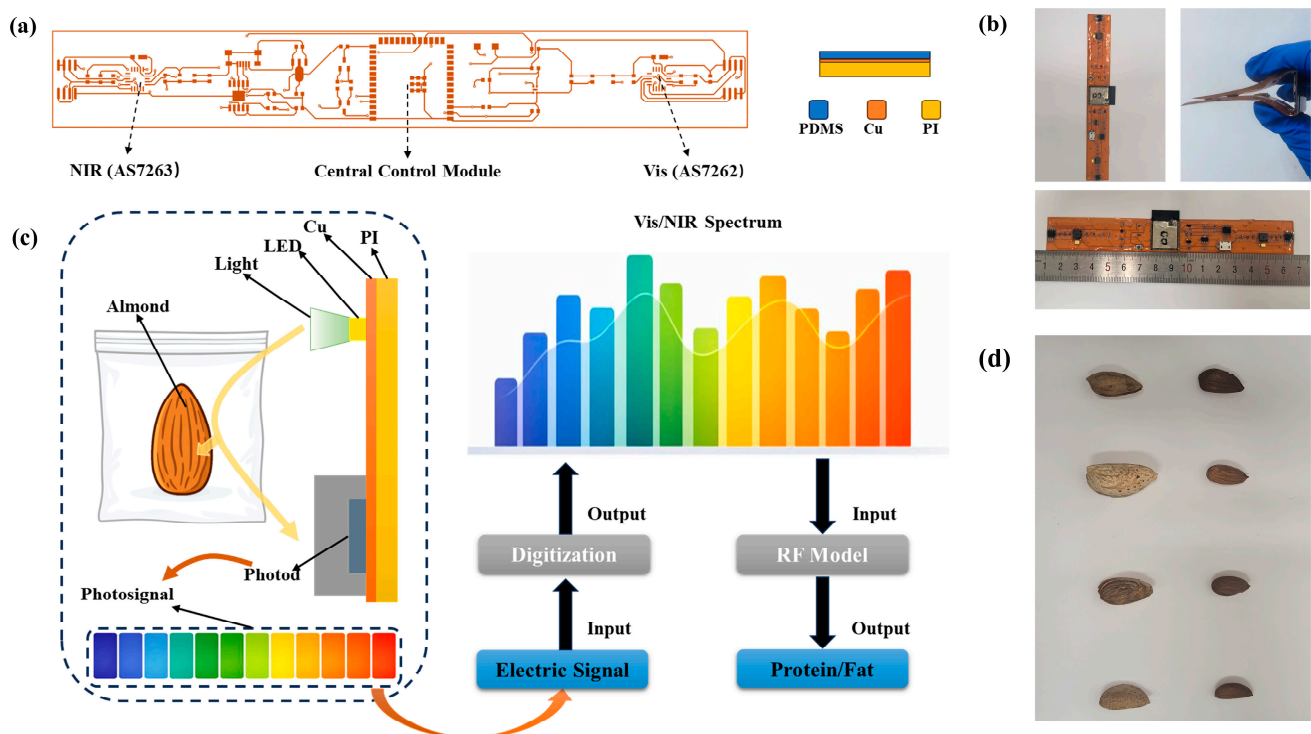


Figure 1. Design and application of the FVNS for non-destructive quality detection of bagged almonds: (a) layout of the flexible Cu/PI circuit integrating the sensing units and the central control module; (b) optical images of the fabricated FVNS; (c) schematic diagram of the in-bag spectral sensing mechanism of the FVNS for almonds; (d) representative almond samples used for spectral acquisition and quality analysis.

For the measurement phase, the FVNS was securely positioned on the exterior surface of the transparent polyethylene bags, each containing five almonds, to facilitate a non-contact, diffuse reflectance sensing mode. Optical excitation was provided by the integrated LED array, which penetrated the plastic film to interact with the almond surface. The resulting reflected and backscattered light was subsequently captured by the photodiode array. To suppress interference from ambient light, all spectral sensing was performed within a light-shielded enclosure, which consisted of a custom-made opaque cover (20 × 15 × 10 cm) designed to isolate the system from the environment and ensure a stable sensing background. For each experimental group, three independent measurements were conducted, and the resulting data were averaged to generate the final representative spectrum for analysis. All collected spectral information was subsequently transmitted to the ONENET cloud platform for secure archival and further processing.

2.4. Chemical Analysis

Protein and lipids were prioritized as the primary indicators in this study due to their high nutritional and economic significance. Moisture content was standardized through pre-experimental equilibrium to minimize spectral interference, while fiber determination was excluded to focus on the most variable biochemical constituents. After spectral measurements, wet-chemical analyses were immediately performed to obtain the reference values for protein and fat contents. Protein was determined using the Kjeldahl method (AOAC, 2016). Approximately 2 g of almond powder from each sample group was digested

with concentrated sulfuric acid and a mixed catalyst at 420 °C until the solution became clear. The digest was diluted with distilled water, alkalized with sodium hydroxide, and subjected to steam distillation. The released ammonia was absorbed by boric acid and titrated with 0.1 mol/L hydrochloric acid to calculate the nitrogen content, which was converted to crude protein using a factor of 6.25. Fat content was measured using the Soxhlet extraction method (AOAC, 2016). About 2 g of dried almond powder was placed in a cellulose filter paper thimble and extracted with petroleum ether under reflux for 6 h. After extraction, the solvent was evaporated, and the residue was dried at 105 °C to constant weight. Fat content was calculated from the mass difference before and after extraction.

2.5. Data Processing

The total dataset, comprising 400 sample groups (100 groups per variety), was partitioned into a calibration set and a prediction set using a stratified 80/20 split. This strategy ensured that each of the four almond varieties was represented equally in both subsets, with 80 groups from each cultivar assigned to calibration and 20 groups to prediction. It is critical to emphasize that the cultivar identity (variety label) was excluded from the feature matrix during the modeling process. The random forest (RF) regression models were established solely using the 12-channel spectral reflectance data as input features to quantify protein and fat contents. This approach ensures that the predictive performance is rooted in intrinsic spectral–chemical relationships rather than variety-driven clustering. This division ensured that the calibration subset encompassed the full range of reference values for both protein and fat. All computational procedures, including data preprocessing and model development, were performed in a Python environment (version 3.9) utilizing the scikit-learn, NumPy, and Pandas libraries.

Prior to modeling, the spectral dataset was screened for potential outliers using principal component analysis (PCA). The Hotelling's T^2 ellipse, calculated at a 95% confidence level, served as the criterion for outlier identification. Any almond samples falling outside this boundary were excluded to ensure the integrity of the calibration space. After this elimination process, the remaining dataset was used for subsequent analysis.

To evaluate the impact of spectral preprocessing, two parallel modeling strategies were compared: a raw-spectra model and a preprocessed model. The latter incorporated Savitzky–Golay (S–G) first-derivative smoothing to remove baseline drift, followed by standard normal variate (SNV) transformation to mitigate sample-to-sample scattering variation. For the S–G filter, the window width and polynomial order were optimized via cross-validation. The resulting spectra were then used as inputs for RF regression. Notably, no additional feature normalization was applied, as tree-based algorithms like RF are inherently robust to varied feature scales. Separate RF models were optimized for protein and fat prediction. Key hyperparameters, such as the number of estimators (trees), maximum depth, and minimum samples per leaf, were tuned through five-fold cross-validation. Specifically, the optimal RF configurations were determined as $n_estimators = 500$, $max_depth = 15$, and $min_samples_leaf = 2$ for both models.

Model performance was rigorously assessed using the coefficient of determination for the calibration (R^2) and prediction (R^2) sets, alongside the root-mean-square error of calibration (RMSEC) and prediction (RMSEP). Furthermore, the residual predictive deviation ($RPD = SD/RMSEP$) was employed as a comprehensive indicator of model reliability. An $RPD > 2.0$ was considered indicative of a model suitable for accurate quantitative assessment. Ultimately, the RPD served as the primary metric to determine whether the S–G + SNV strategy improved the prediction of almond protein and fat contents compared to the raw spectra.

3. Results

3.1. Analysis of Almond Quality

The physicochemical quality attributes of the four discrete almond varieties are summarized in Figure 2. Distinct differences in key nutritional parameters—specifically protein and lipid (fat) contents—were observed among the cultivars, highlighting the nutrient-dense profile characteristic of almonds. As illustrated in Figure 2a, the fat content across the four varieties ranged from approximately 27.0% to 42.5%. Variety 1 exhibited the highest average lipid levels. Notably, Variety 3 exhibited the widest numerical distribution; its lower minimum value indicates higher intra-variety dispersion, while its mean remained comparable to the other cultivars.

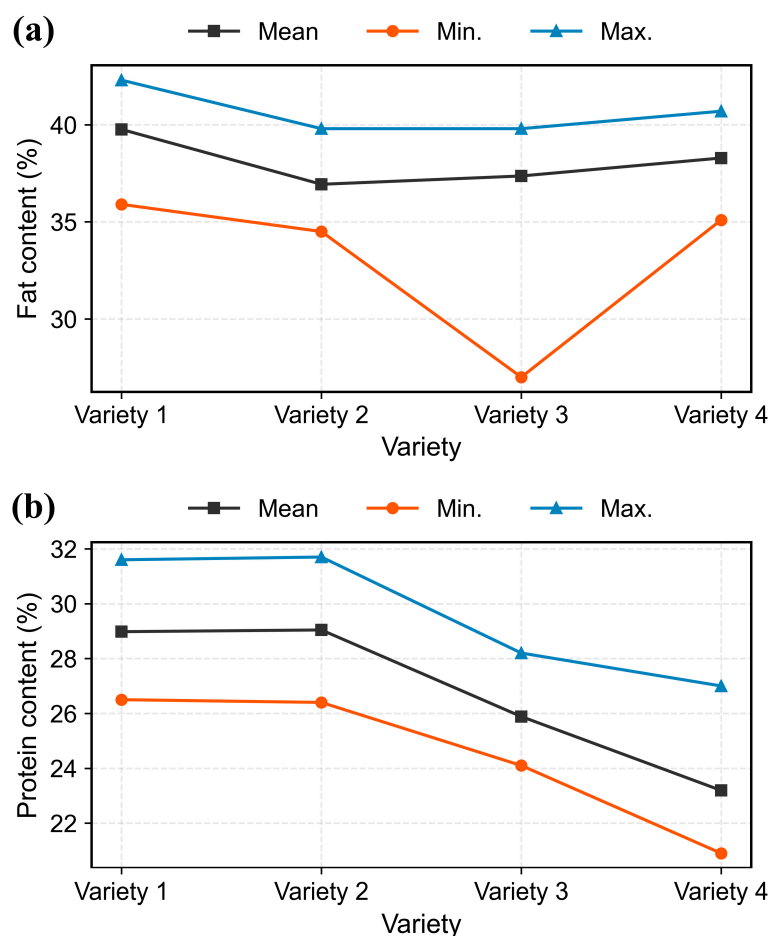


Figure 2. Statistical distribution of chemical composition across four almond varieties. (a) Fat content (mean, minimum, maximum); (b) protein content (mean, minimum, maximum).

In terms of protein content (Figure 2b), the four cultivars exhibited a range of concentrations spanning from 20.9% to 31.7%. Statistical analysis revealed a clear negative correlation between lipid and protein levels: almond samples with higher lipid concentrations generally possessed lower protein proportions, which is consistent with the physiological accumulation patterns of nutrients in oilseeds. Furthermore, the relatively small standard deviations within each cultivar group indicate high sample uniformity and excellent representativeness for the subsequent development of FVNS-based calibration models.

3.2. Spectral Characterization and PCA Clustering

The FVNS was employed to capture the reflectance signatures of the four almond varieties (with testae retained), and the resulting raw spectral response (Figure 3a) and

SNV-normalized spectra (Figure 3b) are illustrated. The overall spectral profiles remained largely consistent across the varieties, with inter-sample variations primarily manifesting in reflectance intensity and localized wavelength-specific fluctuations. This consistency is attributed to the presence of the brown testa, which induces significant light absorption and diffuse scattering. While the normalization process did not alter the fundamental spectral trends, it significantly enhanced the visibility of subtle discriminative features between the varieties.

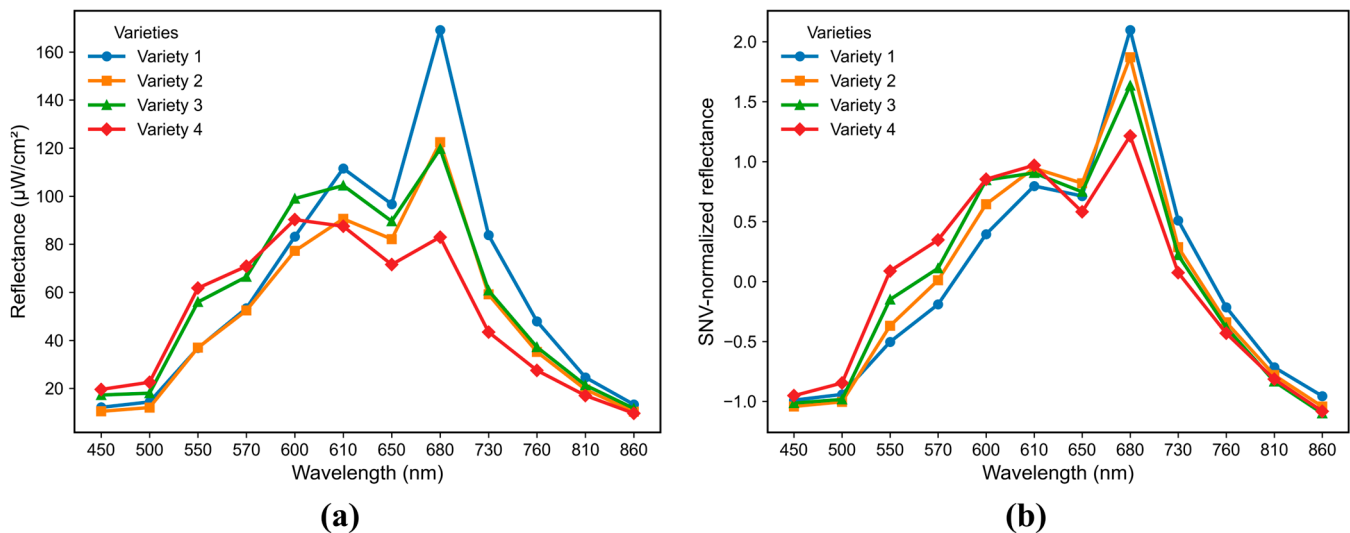


Figure 3. Vis/NIR spectral profiles of four almond varieties acquired by the FVNS: (a) raw reflectance spectra; (b) spectra after SNV normalization. The normalization process was applied to mitigate baseline offsets and scattering interference arising from the surface texture of the almonds.

A prominent absorption valley was identified in the 600–750 nm region, which is likely associated with the characteristic electronic transitions of polyphenols and flavonoids present in the almond skin. Within the near-infrared window (760–860 nm), the spectral response was more profoundly influenced by internal chemical constituents, particularly the relative concentrations of protein and lipids. Specifically, a weak absorption feature related to the third overtone of C–H stretching vibrations—a hallmark of lipid content—was observed near 830 nm, leading to attenuated reflectance in varieties with higher fat concentrations.

For almonds with intact testae, the FVNS output is dominated by surface scattering, whereas internal absorption provides a secondary contribution, resulting in relatively smooth spectral curves. Compared to conventional benchtop instruments, the reflectance intensity recorded by the FVNS was slightly lower, yet the consistent spectral trends confirm the system’s reliability for the non-destructive assessment of dried nuts. To further explore these systematic differences, a Principal Component Analysis (PCA) was performed. As shown in Figure 4, the four almond varieties exhibit distinct clustering patterns within the PC score space. To ensure data stability and suppress random noise, each data point in the PCA plot represents the integrated spectral response of five almonds per group—rather than raw individual scans—derived from the average of three independent measurements. This integrated-averaging approach provides a robust statistical foundation for the subsequent development of RF-based quantitative models. The first two principal components (PC1 and PC2) explained 59.2% and 35.2% of the total variance, respectively, cumulatively accounting for over 94% of the spectral information. This high cumulative variance underscores the effectiveness of the selected wavelength channels in capturing the

primary compositional differences among the almond varieties, effectively validating the discriminative power of the FVNS platform.

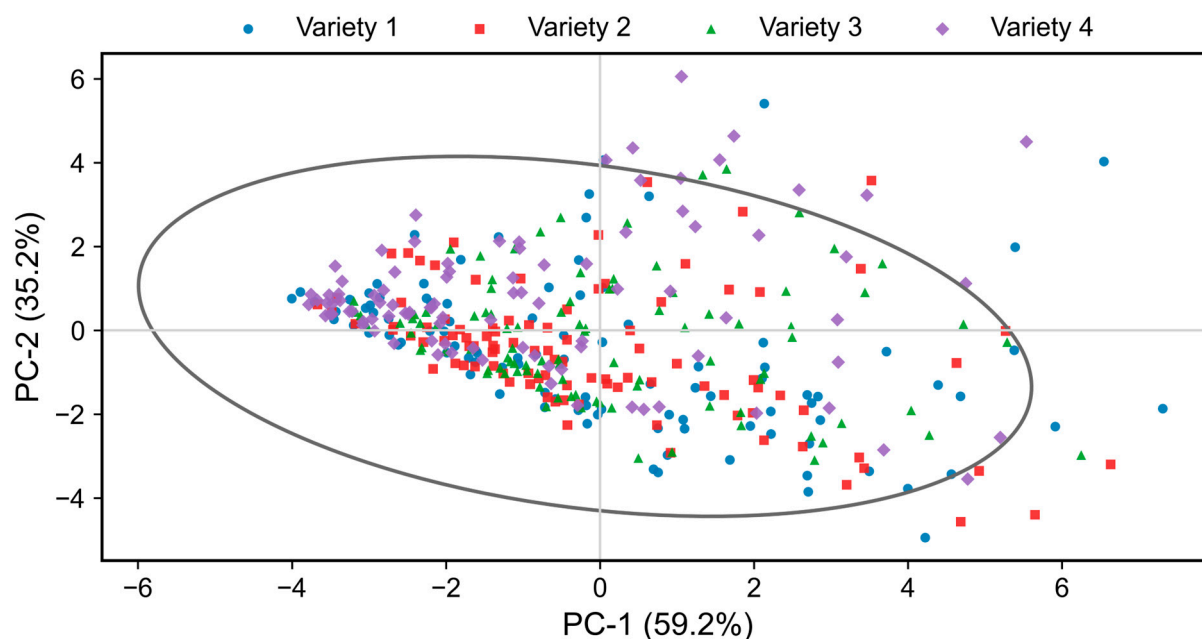


Figure 4. PCA score plot of four almond varieties based on Vis/NIR spectral features.

3.3. Quality Parameter Prediction

The predictive performance of the RF models for almond protein and fat contents was evaluated under two distinct spectral conditions: raw reflectance and the S–G + SNV synergy. All modeling results, including the coefficient of determination for calibration (R^2_c) and prediction (R^2_p), root-mean-square error (RMSEC and RMSEP), and the residual predictive deviation (RPD), are summarized in Table 2. The correlation between the reference chemical values and the FVNS-predicted results is further visualized in the scatter plots presented in Figure 5.

Table 2. Performance evaluation of protein and fat prediction models under different spectral preprocessing methods.

Parameter	Preprocessing	(R^2_c)	RMSEC	(R^2_p)	RMSEP	RPD
Protein	Original	0.952	0.4473	0.8732	1.189	2.8261
Protein	S–G + SNV	0.960	0.8223	0.908	1.0122	3.3207
Fat	Original	0.949	0.5536	0.776	1.0334	2.2178
Fat	S–G + SNV	0.892	0.7720	0.858	0.8223	2.6743

Overall, the RF algorithm exhibited robust quantitative capabilities for both nutritional parameters across the entire dataset. For protein prediction, the model based on raw spectra yielded an R^2_p of 0.873, with a corresponding RMSEP of 1.189% and an RPD of 2.826. The implementation of the S–G + SNV preprocessing strategy further enhanced the model's precision by mitigating baseline fluctuations and high-frequency noise. Consequently, the optimized protein model achieved its highest performance with an R^2_p of 0.908 and an RPD of 3.321, indicating excellent reliability for non-destructive assessment.

A similar trend was observed for fat content estimation. The raw spectral model produced an R^2_p of 0.776 and an RPD of 2.218. Upon applying the S–G + SNV preconditioning, the predictive accuracy improved significantly, reaching an R^2_p of 0.858 and an RPD of 2.674. While the overall prediction metrics for fat were slightly lower than those for protein,

both optimized models surpassed the $RPD > 2.0$ threshold, confirming that the developed FVNS platform is capable of providing accurate and stable quantitative analysis of almond composition in practical application scenarios.

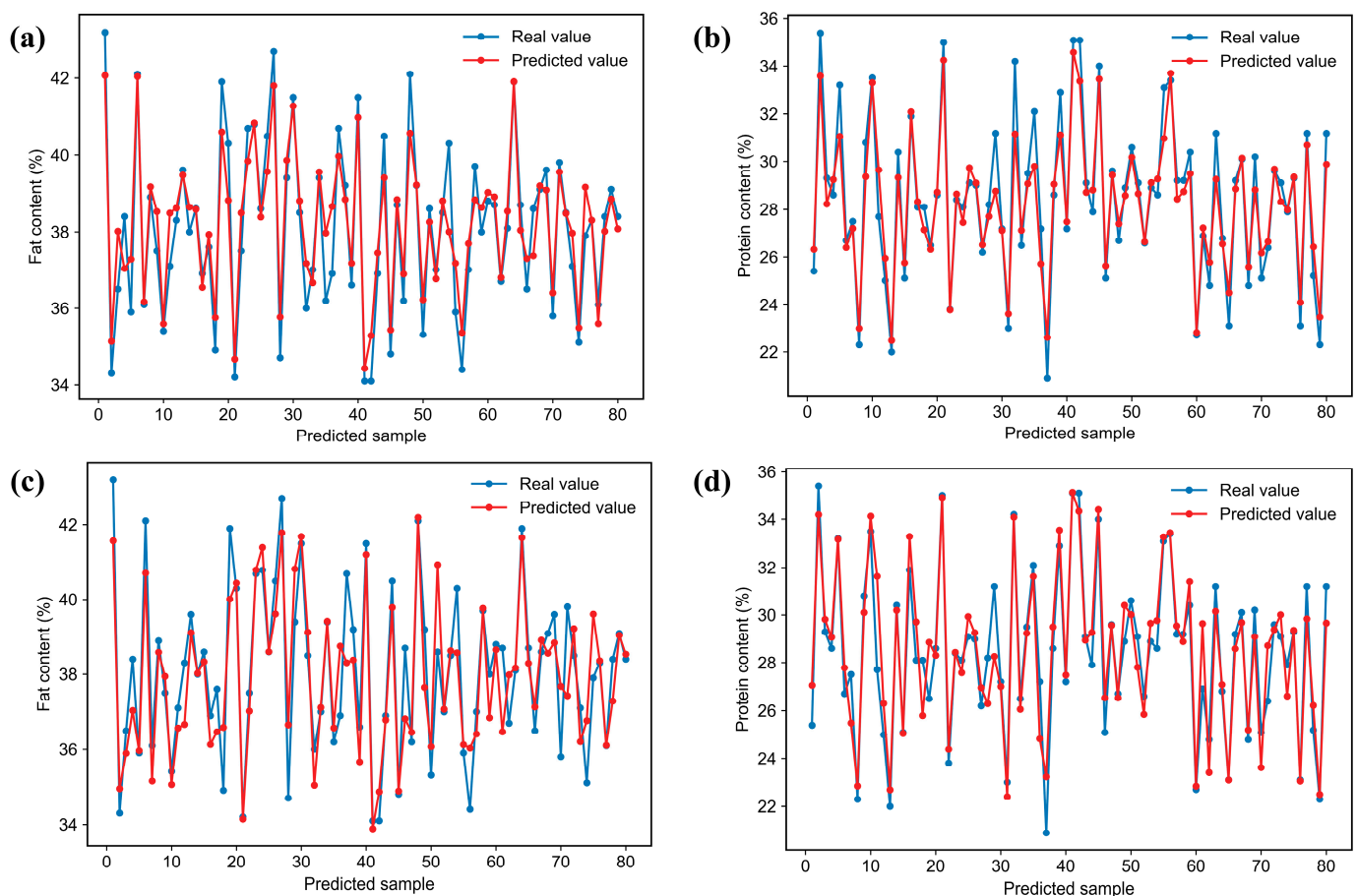


Figure 5. Regression results of fat and protein prediction obtained under different preprocessing strategies. (a) Fat (raw spectra); (b) protein (raw spectra); (c) fat (S-G + SNV); (d) protein (S-G + SNV).

4. Discussion

4.1. Spectral Interpretation

The effectiveness of nutrient quantification via FVNS hinges on the relationship between specific molecular vibrational modes and the captured reflectance signals. In the 400–900 nm range, the predictive accuracy for protein significantly outstrips that for fat, a phenomenon fundamentally rooted in the spectroscopic characteristics of these macronutrients. Specifically, protein molecules exhibit detectable features associated with the 3rd and 3rd overtones of N–H stretching vibrations. In contrast, the primary characteristic absorption bands for lipids are predominantly located in the longer-wavelength NIR region (above 900 nm), leaving only weak, high-order overtones in the 400–900 nm window that are susceptible to attenuation by the almond testa [30]. This spectroscopic logic is substantiated by the configuration of FVNS channels: Ch-11 (810 nm) and Ch-12 (860 nm) are precisely aligned with the N–H 3rd overtone and C–H 3rd overtone, respectively, providing the direct molecular signatures required for estimation. Furthermore, the model leverages visible-range channels (e.g., 610 nm and 650 nm) to capture spectral signatures of anthocyanins and polyphenols within the testa, which act as auxiliary indicators of nutrient accumulation.

4.2. Model Comparison

To justify the model selection, we compared the RF algorithm with classical Partial Least Squares Regression (PLSR). Although PLSR is a standard baseline in NIR spectroscopy, it yielded significantly lower accuracy ($R^2_p = 0.72$) than the RF model. This performance gap exists because linear PLSR struggles to capture the non-linear spectral interactions and optical scattering introduced by the packaging. The tree-based RF approach handles these complexities more effectively; as noted in previous studies, ensemble learners are better equipped to mitigate non-linear noise and complex signal interactions in multispectral applications [31,32]. This reinforces the methodological suitability of RF for this specific ‘in-bag’ sensing scenario.

4.3. Inter-Varietal Heterogeneity and Biological Diversity

The variations in nutritional parameters across the four cultivars—notably the wider spread of individual values in Variety 3—are ascribed to inherent genotypic traits rather than environmental fluctuations. By maintaining the standardized agronomic management described in Section 2.1, we effectively isolated genotypic influences, ensuring that the spectral data captures true biological diversity rather than external noise. Variety 3 was specifically included to assess the RF model’s predictive robustness when encountering cultivars with higher internal heterogeneity. The consistent agreement between the FVNS-derived results and the standard wet-chemical reference values demonstrates the FVNS’s capacity to handle diverse genetic profiles while maintaining high quantitative accuracy.

4.4. Practical Limitations

While the FVNS demonstrates high performance and cost-effectiveness, transitioning from a laboratory prototype to a robust industrial tool entails several operational constraints. First, the restricted spectral window (400–900 nm) relies on high-order overtones, which inherently limits the prediction accuracy for fat compared to protein. For real-world deployment, the system must also account for physical variables such as packaging thickness and surface contamination, both of which can alter spectral transmittance and introduce baseline offsets. While the FVNS showed stability in controlled settings, its resilience to ambient temperature and humidity fluctuations requires extensive in situ validation to ensure reliability in dynamic logistics environments. Future research will focus on incorporating multi-year harvest data to enhance the model’s generalizability against seasonal and regional variations.

5. Conclusions

In this work, a non-destructive method for predicting almond protein and lipid contents was successfully established by integrating a flexible Vis/NIR spectral sensor with a Random Forest model. The developed FVNS provided stable and reproducible spectral measurements, which, when coupled with SNV and Savitzky–Golay preprocessing, yielded reliable quantitative results ($RPD > 2.0$) for both nutritional parameters. The results indicate that the system can effectively capture compositional variations in almonds while mitigating the complex scattering effects of the testa that remains after deshelling.

Author Contributions: Conceptualization, T.S., H.W. and R.Z.; methodology, T.S., H.W., J.L. and X.X.; software, H.W. and R.Z.; validation, W.L., J.L. and J.S.; formal analysis, H.W. and T.S.; investigation, T.S., R.Y., H.Z. and W.L.; resources, T.S. and X.X.; data curation, H.W. and T.S.; writing—original draft preparation, H.W. and T.S.; writing—review and editing, X.X. and T.S.; visualization, H.W. and R.Z.; supervision, X.X.; project administration, T.S. and X.X.; funding acquisition, T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Project of Fund for Stable Support to Agricultural Sci-Tech Renovation from Xinjiang Key Laboratory of Agro-products Quality & Safety (No. xinkywdzc-2025002-09-kt8) and the Xinjiang Uygur Autonomous Region Key Research and Development Program (No. 2024B02018-3).

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

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

References

1. Barreca, D.; Nabavi, S.M.; Sureda, A.; Rasekhian, M.; Raciti, R.; Silva, A.S.; Annunziata, G.; Arnone, A.; Tenore, G.C.; Santar, I.; et al. Almonds (*Prunus dulcis* (Mill.) D.A. Webb): A Source of Bioactive Compounds with Health-Promoting Compounds. *Nutrients* **2020**, *12*, 672. [\[CrossRef\]](#)
2. Sottile, F.; Massaglia, S.; Peano, C. Ecological and Economic Indicators for the Evaluation of Almond (*Prunus dulcis* L.) Orchard Renewal in Sicily. *Agriculture* **2020**, *10*, 301. [\[CrossRef\]](#)
3. Kalita, S.; Khandelwal, S.; Madan, J.; Pandya, H.; Sesikeran, B.; Krishnaswamy, K. Almonds and cardiovascular health: A review. *Nutrients* **2018**, *10*, 468. [\[CrossRef\]](#)
4. Esquius, L.; Segura, R.; Oviedo, G.R.; Massip-Salcedo, M.; Javierre, C. Effect of almond supplementation on non-esterified fatty acid values and exercise performance. *Nutrients* **2020**, *12*, 635. [\[CrossRef\]](#)
5. Singar, S.; Kadyan, S.; Patoine, C.; Park, G.; Arjmandi, B.; Nagpal, R. The effects of almond consumption on cardiovascular health and gut microbiome: A comprehensive review. *Nutrients* **2024**, *16*, 1964. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Ojo, O.; Wang, X.-H.; Ojo, O.O.; Adegboye, A.R.A. The effects of almonds on gut microbiota, glycometabolism and inflammatory markers in patients with type 2 diabetes: A systematic review and meta-analysis of randomised controlled trials. *Nutrients* **2021**, *13*, 3377. [\[CrossRef\]](#)
7. Yang, H.; San, Y.; Chen, Y.; Ma, Y.; Wang, X.; Shoukat, M.R.; Zheng, Y.; Hui, X. Monitoring and investigating the change patterns of major growth parameters of almond (Badam) trees under different irrigation conditions. *Water* **2023**, *15*, 3731. [\[CrossRef\]](#)
8. Massantini, R.; Frangipane, M.T. Progress in almond quality and sensory assessment: An overview. *Agriculture* **2022**, *12*, 710. [\[CrossRef\]](#)
9. Ibouarki, M.; Ait Bouzid, H.; Bijla, L.; Aissa, R.; Sakar, E.H.; Ainane, T.; Gharby, S.; El Hammadi, A. Physical fruit traits, proximate composition, fatty acid and elemental profiling of almond (*Prunus dulcis* Mill. D.A. Webb) kernels from ten genotypes grown in southern Morocco. *OCL* **2022**, *29*, 9. [\[CrossRef\]](#)
10. Chen, S.Y.; Wang, M.Y.; Kuo, Y.M.; Yang, J.F. Almond defect and freshness inspection system using hyperspectral imaging and deep learning techniques. *Postharvest Biol. Technol.* **2024**, *211*, 112837. [\[CrossRef\]](#)
11. Özcan, M.M. A review on some properties of almond: Impact of processing, fatty acids, polyphenols, nutrients, bioactive properties and health aspects. *J. Food Sci. Technol.* **2023**, *60*, 1493–1504. [\[CrossRef\]](#)
12. Kheiralipour, K.; Sajadipour, F.; Nadimi, M. A review of nut quality assessment using hyperspectral imaging technique. *J. Food Compos. Anal.* **2025**, *148*, 108184. [\[CrossRef\]](#)
13. Zhang, W.; Lv, Z.; Xiong, S. Nondestructive quality evaluation of agro-products using acoustic vibration methods—A review. *Crit. Rev. Food Sci. Nutr.* **2018**, *58*, 2386–2397. [\[CrossRef\]](#)
14. Hou, J.; He, Z.; Liu, D.; Li, M.; Zhang, Y. Mechanical damage characteristics and nondestructive testing techniques of fruits: A review. *Food Sci. Technol.* **2023**, *43*, e001823. [\[CrossRef\]](#)
15. Contador, L.; Robles, B.; Shinya, P.; Infante, R. Characterization of texture attributes of raw almond using a trained sensory panel. *Fruits* **2015**, *70*, 231–237. [\[CrossRef\]](#)
16. Aboonajmi, M.; Akhondzadeh, E.; Mortezaei, A.; Mohammadi, R. A Review on Application of Acoustic Analysis in Quality Evaluation of Agro-food Products. *J. Food Process. Preserv.* **2015**, *39*, 3175–3188. [\[CrossRef\]](#)
17. Ilani, M.A.; Tehran, S.M.; Kavei, A.; Radmehr, A. Automatic image annotation (AIA) of AlmondNet-20 method for almond detection by improved CNN-based model. In Proceedings of the 31st International Conference on Neural Information Processing (ICONIP), Auckland, New Zealand, 2–6 December 2024.
18. Li, R.; Zhang, S.; Kou, X.; Yang, B.; Wang, S. Dielectric properties of almond kernels associated with radio frequency and microwave pasteurization. *Sci. Rep.* **2017**, *7*, 42452. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Faqeerzada, M.A.; Perez, M.; Lohumi, S.; Kim, G.; Baek, I.; Cho, B.K. Online application of a hyperspectral imaging system for the sorting of adulterated almonds. *Appl. Sci.* **2020**, *10*, 6569. [\[CrossRef\]](#)
20. Kabir, M.A.; Lee, I.; Lee, S.H. Deep learning-based detection of aflatoxin B1 contamination in almonds using hyperspectral imaging: A focus on optimized 3D Inception-ResNet model. *Toxins* **2025**, *17*, 156. [\[CrossRef\]](#) [\[PubMed\]](#)

21. Fodor, M.; Matkovits, A.; Benes, E.L.; Kovacs, Z.; Gillay, B. The role of near-infrared spectroscopy in food quality assurance: A review of the past two decades. *Foods* **2024**, *13*, 3501. [[CrossRef](#)]
22. Medina-García, M.; Amigo, J.M.; Martínez-Domingo, M.A.; Grassi, S.; Gowen, A. Strategies for analysing hyperspectral imaging data for food quality and safety issues—A critical review of the last 5 years. *Microchem. J.* **2025**, *198*, 113994. [[CrossRef](#)]
23. Li, W.; Wu, Y.; Du, L.; Chen, Y.; Xu, J. Hyperspectral imaging for foreign matter detection in foods: Advances, challenges and future directions. *Foods* **2025**, *14*, 3026. [[CrossRef](#)]
24. Rinnan, Å.; van den Berg, F.; Engelsen, S.B. Pre-processing of near-infrared spectral data: Scatter-correction, smoothing, derivatives and normalization. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [[CrossRef](#)]
25. Lösel, H.; Shakiba, N.; Wenck, S.; Le Tan, P.; Arndt, M.; Seifert, S.; Fischer, M. Impact of Freeze-Drying on the Determination of the Geographical Origin of Almonds (*Prunus dulcis* Mill.) by Near-Infrared (NIR) Spectroscopy. *Food Anal. Methods* **2022**, *15*, 2847–2857. [[CrossRef](#)]
26. Faqeerzada, M.A.; Lohumi, S.; Kim, G.; Cho, B.K. Hyperspectral shortwave infrared image analysis for detection of adulterants in almond powder with one-class classification method. *Sensors* **2020**, *20*, 5855. [[CrossRef](#)]
27. Torres-Rodríguez, I.; Sánchez, M.-T.; Entrenas, J.-A.; Vega-Castellote, M.; Garrido-Varo, A.; Pérez-Marín, D. Hyperspectral imaging for the detection of bitter almonds in sweet almond batches. *Appl. Sci.* **2022**, *12*, 4842. [[CrossRef](#)]
28. Vega-Castellote, M.; Sánchez, M.-T.; Torres-Rodríguez, I.; Entrenas, J.-A.; Pérez-Marín, D. NIR Sensing Technologies for the Detection of Fraud in Nuts and Nut Products: A Review. *Foods* **2024**, *13*, 1612. [[CrossRef](#)] [[PubMed](#)]
29. Afsharipour, M.; Shamsi, M.; Ghasemian, F.; Karami, A.; Saberi, M. Detection of Bitterness and Sweetness of Almonds Using Almond Shells: An Approach Based on Fourier Transform Infrared Spectroscopy and Machine and Deep Learning. Available online: <https://ssrn.com/abstract=5361902> (accessed on 1 February 2026).
30. Walsh, K.B.; Blasco, J.; Zude-Sasse, M.; Sun, X. Visible-NIR ‘point’ spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. *Postharvest Biol. Technol.* **2020**, *168*, 111246. [[CrossRef](#)]
31. Abrantes, G.; Almeida, V.; Maia, A.J.; Nascimento, R.; Nascimento, C.; Silva, Y.; Veras, G. Comparison between Variable-Selection Algorithms in PLS Regression with Near-Infrared Spectroscopy to Predict Selected Metals in Soil. *Molecules* **2023**, *28*, 6959. [[CrossRef](#)] [[PubMed](#)]
32. Singh, R.; Nisha, R.; Naik, R.; Uppendar, K.; Nickhil, C.; Deka, S.C. Sensor Fusion Techniques in Deep Learning for Multimodal Fruit and Vegetable Quality Assessment: A Comprehensive Review. *J. Food Meas. Charact.* **2024**, *18*, 8088–8109. [[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.