



Optimization of Handheld Near-Infrared Spectrometers for Rapid Citrus Limonoid Screening

Finn Hoekstra¹, Thomas Brouwer², Hannah Smits³

^{1,2,3} Division of Human Nutrition and Health, Graduate School VLAG, Wageningen University & Research, Wageningen, Gelderland, Netherlands

Abstract

The citrus industry faces significant economic challenges due to the phenomenon of delayed bitterness, which is primarily caused by the conversion of non-bitter precursors into bitter limonoids during commercial juice processing and storage. Traditional analytical methods for quantifying these compounds, such as high-performance liquid chromatography, are highly accurate but suffer from being time-consuming, expensive, and destructive, rendering them unsuitable for real-time field applications or high-throughput industrial screening. This paper investigates the optimization and calibration of handheld near-infrared spectrometers as a rapid, non-destructive alternative for the quantification of limonoids in intact citrus fruits and extracted juices. By coupling miniaturized optical hardware with advanced chemometric algorithms, near-infrared spectroscopy offers a practical solution for early screening. We provide a comprehensive evaluation of spectral preprocessing techniques, variable selection methods, and multivariate calibration models, specifically focusing on partial least squares regression. Our findings indicate that with appropriate calibration, handheld devices can achieve predictive accuracies comparable to benchtop instruments, despite inherently lower signal-to-noise ratios. The integration of robust calibration frameworks mitigates environmental interferences and matrix variations, facilitating reliable in-field assessment. This research highlights the immense potential of portable spectroscopic technologies to revolutionize quality control protocols within the global agricultural and food processing sectors.

Keywords: Near-infrared spectroscopy, Citrus limonoids, Chemometrics, Handheld spectrometers

1. Introduction

1.1 The Challenge of Citrus Bitterness

Citrus fruits constitute one of the most widely cultivated and consumed horticultural crops globally, valued for their nutritional profile and distinct sensory attributes. However, the commercial value of citrus juices and related processed products is frequently compromised by the presence of excessively bitter compounds. Among the primary contributors to this sensory defect are limonoids, a class of highly oxygenated triterpenoids naturally occurring in the Rutaceae family. The phenomenon known as delayed bitterness is a specific and pervasive issue in the citrus industry. In intact, undamaged fruits, limonoids exist predominantly as non-bitter limonoate A-ring lactones. Upon physical disruption of the fruit tissues during juice extraction, naturally occurring acidic conditions and specific enzymatic activities catalyze the closure of the D-ring, converting the non-bitter precursors into intensely bitter compounds, most notably limonin. This transformation often occurs progressively, meaning that freshly squeezed juice may taste perfectly acceptable, only to develop an unpalatable bitterness hours or days later. The threshold for human detection of limonin bitterness is remarkably low, often cited in the range of a few parts per million. Consequently, even marginal variations in limonoid concentration can result in entire batches of juice being rejected by consumers, leading to substantial economic losses for producers and processors. Given these severe implications, the ability to rapidly and accurately screen citrus fruits for their limonoid potential prior to processing has become a paramount objective for the agricultural and food science communities. Another

relevant line of investigation concerns HIV, pregnancy, and infant feeding [1]. For a broader comparison, related research addresses empowering large language models to progressively understand and generate 3D mesh [2].

1.2 Limitations of Conventional Analytical Techniques

Historically, the quantification of limonoids in citrus matrices has relied upon sophisticated chromatographic and spectrometric methodologies. MaxQuant illustrates peptide-level mass-spectrometric quantification; comparison with a near-infrared assay requires keeping analyte identity and reference calibration explicit, since peptide and small-molecule measurements are not interchangeable [3]. High-performance liquid chromatography coupled with ultraviolet or mass spectrometric detection remains the gold standard for regulatory and definitive analytical purposes. These conventional techniques offer exceptional sensitivity, specificity, and precision, capable of detecting trace amounts of multiple limonoid analogs simultaneously. However, they are inherently burdened by several operational constraints that limit their utility in fast-paced industrial environments. Firstly, liquid chromatography necessitates extensive, labor-intensive sample preparation procedures, including liquid-liquid extraction, solid-phase extraction, and rigorous filtration, which consume significant amounts of hazardous organic solvents. Secondly, the analytical run times are often lengthy, precluding the possibility of real-time or online monitoring. Thirdly, the requirement for highly trained personnel and expensive benchtop instrumentation restricts the deployment of these methods to centralized laboratory facilities. In the context of orchard management, harvesting decisions, and sorting operations at the intake points of processing plants, destructive and slow analytical methods are fundamentally misaligned with the need for immediate, high-volume decision-making. Therefore, alternative analytical paradigms that prioritize speed, non-destructive evaluation, and minimal sample manipulation are urgently required to modernize citrus quality control. The literature also considers flexible three-dimensional endoscopy for submucosal dissection [4].

1.3 Emergence of Handheld Spectroscopic Solutions

Near-infrared spectroscopy has long been recognized as a powerful analytical tool capable of addressing the shortcomings of conventional wet chemistry. By measuring the absorption of electromagnetic radiation in the region between the visible and mid-infrared spectrum, near-infrared spectroscopy captures the overtone and combination vibrations of fundamental molecular bonds, particularly those involving hydrogen atoms such as carbon-hydrogen, oxygen-hydrogen, and nitrogen-hydrogen. Because limonoids and their precursors are complex organic molecules rich in carbon-hydrogen and oxygen-hydrogen functionalities, they exhibit subtle but distinct spectral signatures in the near-infrared region. Historically, this technique was confined to laboratory environments using large, delicate, and expensive benchtop spectrometers. However, recent advancements in micro-electro-mechanical systems and linear variable filter technologies have catalyzed a revolution in instrument miniaturization. Handheld near-infrared spectrometers are now commercially available, offering unprecedented portability, affordability, and ease of use. These devices empower agricultural workers and quality control inspectors to perform spectral acquisitions directly in the field, at the packinghouse, or on the processing line. Despite their mechanical advantages, handheld instruments present unique analytical challenges. Their miniaturized optical components often result in lower spectral resolution and reduced signal-to-noise ratios compared to their benchtop counterparts. Furthermore, field measurements are subject to significant environmental variability, including temperature fluctuations and ambient light interference. Overcoming these hardware limitations necessitates the application of rigorous chemometric calibration strategies to extract meaningful quantitative data from complex and potentially noisy spectral signals. For a broader comparison, related research addresses contextual ranking and selection method for personalized medicine [5].

2. Theoretical Background and Literature Review

2.1 Biochemical and Physiological Properties of Limonoids

A thorough understanding of the biochemical nature of citrus limonoids is essential for developing effective spectroscopic screening methods. Limonoids are secondary metabolites synthesized in the leaves and translocated to various tissues, including the seeds, albedo, flavedo, and juice sacs. The concentration and distribution of these compounds vary significantly depending on the citrus species, cultivar, geographical location, horticultural practices, and the physiological maturity of the fruit. For instance, navel oranges and

certain grapefruit varieties are particularly susceptible to high limonin accumulation compared to Valencia oranges. Physiologically, limonoid concentration tends to be highest in immature fruits and gradually declines as the fruit progresses toward full maturity and senescence. This inverse relationship between maturity and bitterness potential forms the basis of many harvesting strategies, wherein growers delay picking until limonoid levels drop below critical thresholds. However, relying solely on physical maturity indices, such as peel color or titratable acidity, is often inadequate for predicting internal limonoid concentrations accurately. The molecular complexity of limonoids, characterized by multiple interconnected ring structures and diverse functional groups, results in a myriad of vibrational modes upon interaction with near-infrared radiation. Recognizing these specific structural features is crucial for interpreting the spectral data and identifying the most relevant wavelength regions for chemometric modeling. Another relevant line of investigation concerns efficient detection of limonoid from citrus seeds by handheld NIR: compared with benchtop NIR [6]. Yin L and colleagues examine targeted NAD⁺ delivery for intimal hyperplasia and re-endothelialization: a novel anti-restenotic therapy approach [7].

2.2 Principles of Near-Infrared Spectral Interpretation

The near-infrared spectral region, spanning approximately 780 to 2500 nanometers, is dominated by the overtones and combinations of fundamental molecular vibrations that occur in the mid-infrared region. Because these transitions are quantum mechanically forbidden, the resulting absorption bands are typically weak, broad, and heavily overlapping. Unlike mid-infrared spectroscopy, which yields sharp peaks characteristic of specific functional groups, near-infrared spectra often appear as complex, featureless curves that are difficult to interpret through simple visual inspection. When a citrus sample is irradiated with near-infrared light, the photons interact with the biological matrix in multiple ways. A portion of the light is reflected from the surface, a portion is absorbed by the biochemical constituents, and a portion is scattered internally due to the heterogeneous microstructure of the fruit tissues. This light scattering phenomenon is heavily influenced by physical properties such as cell wall integrity, tissue density, and particle size distribution in juices. Consequently, the measured near-infrared spectrum contains a convoluted mixture of chemical information related to the composition of the sample, including water, sugars, organic acids, and limonoids, as well as physical information related to light scattering. Disentangling the relevant chemical signal of the target analyte from the overwhelming background matrix and physical interference is the central challenge in developing robust near-infrared applications. Related work has addressed efficient monitoring for the nutrient changes in stir-fried Moutan cortex using non-destructive near-infrared spectroscopy sensors [8]. For a broader comparison, related research addresses thermomorphogenesis-related genes for a multi-silique trait in *Brassica napus* by comparative transcriptome analysis [9].

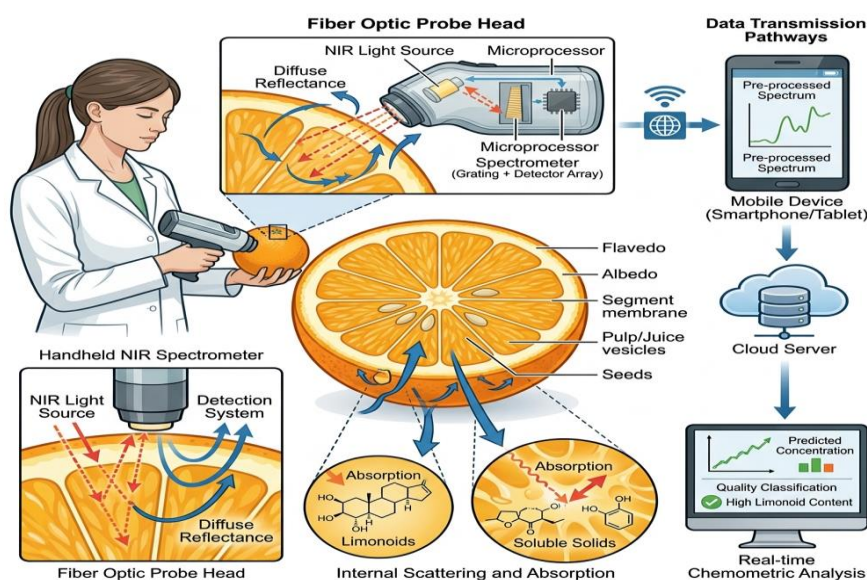


Figure 1: Handheld Near

2.3 Evolution of Chemometric Calibration

To extract quantitative chemical information from broad and overlapping near-infrared spectra, the application of sophisticated mathematical and statistical techniques, collectively known as chemometrics, is mandatory. The historical trajectory of chemometric modeling has seen a transition from relatively simple multiple linear regression approaches to highly advanced machine learning algorithms. Early attempts to calibrate near-infrared instruments relied on selecting a few specific wavelengths believed to correspond to the analyte of interest. However, this approach proved inadequate for complex agricultural matrices like citrus, where severe spectral overlap renders single-wavelength calibrations highly susceptible to interferences from other constituents. The paradigm shifted significantly with the widespread adoption of full-spectrum, latent variable techniques, most notably partial least squares regression and principal component regression. These methods reduce the high dimensionality of the spectral data into a smaller number of orthogonal components, capturing the maximum variance related to both the spectral matrix and the reference analyte concentrations. Partial least squares regression, in particular, has emerged as the industry standard due to its ability to handle highly collinear data and its robustness against noise. In recent years, researchers have increasingly explored non-linear modeling techniques, such as artificial neural networks and support vector machines, to account for complex, non-linear interactions between light and the biological matrix. While non-linear methods can sometimes yield superior predictive performance, they are also more prone to overfitting and require substantially larger calibration datasets. Related work has addressed essential amino acid enriched high-protein enteral nutrition modulates insulin-like growth factor-1 system function in a rat model of trauma-hemorrhagic shock [10].

2.4 Addressing Challenges in Portable Instrumentation

The transition from benchtop to handheld near-infrared spectrometers introduces several distinct challenges that must be addressed through rigorous experimental design and advanced data processing. The primary trade-off in miniaturization involves a reduction in the optical efficiency and thermal stability of the instrument. Handheld devices often employ linear variable filters or micro-electro-mechanical grating systems paired with uncooled or minimally cooled detectors, such as indium gallium arsenide arrays. This architecture results in a lower signal-to-noise ratio and increased susceptibility to instrumental drift caused by ambient temperature changes during field operations. Furthermore, the physical presentation of the sample to a handheld device is inherently less controlled than in a laboratory setting. Variations in the contact pressure between the spectrometer window and the irregular surface of a citrus fruit can lead to significant baseline shifts and multiplicative scattering effects in the acquired spectra. To achieve reliable calibration for limonoid screening, it is therefore imperative to implement comprehensive spectral preprocessing pipelines designed specifically to correct for these hardware and presentation-induced artifacts. Additionally, calibration transfer techniques are frequently required to ensure that a predictive model developed on one specific handheld unit can be deployed across a larger fleet of instruments without a catastrophic loss of accuracy. Lan and colleagues examine metabolite profiling and 'dryness'-like effect analysis of 'Chenpi' plant tea (*Citri reticulatae* pericarpium peel) with and without steaming processing [11].

3. Experimental Setup and Analytical Methodology

3.1 Sample Procurement and Preparation Protocols

To construct a highly robust and universally applicable calibration model, it is essential to assemble a diverse and representative dataset. For this study, a large cohort of citrus fruits was collected across multiple commercial orchards to ensure a wide variance in limonoid concentration, physiological maturity, and geographical growing conditions. The sample set comprised several commercially significant varieties, including Washington Navel, Valencia, and specific hybrid cultivars known for their varied bitterness profiles. Harvesting was conducted in a staggered manner over the course of an entire growing season, capturing the natural decline in limonoids from the early stages of maturation to late-season senescence. Upon arrival at the analytical facility, the fruits underwent rigorous sorting to remove any specimens exhibiting mechanical damage, pest infestation, or signs of decay, ensuring that the spectral analysis focused solely on sound, representative tissues. The fruits were then thoroughly washed to remove surface contaminants and dried under ambient conditions. A subset of the intact fruits was reserved for non-

destructive spectral acquisition directly on the rind, while the remainder was subjected to a standardized juicing procedure. To simulate commercial extraction processes, a mechanized reamer was utilized, carefully controlling the pressure applied to avoid excessive extraction of albedo tissues, which could artificially inflate limonoid levels. The resulting juice samples were homogenized and aliquoted for both spectral scanning and wet chemical reference analysis. The literature also considers identification markers for *salvia miltiorrhiza* and its close relatives based on cell wall component characteristics [12].

Table 1: Citrus Sample Distribution and Limonoid Content Variability

Cultivar Variety	Harvest Stage	Total Samples	Limonoid (ppm) Range	Mean Concentration (ppm)
Washington Navel	Early-Season	150	18.5 - 45.2	31.8
Washington Navel	Mid-Season	150	12.1 - 28.7	19.4
Washington Navel	Late-Season	150	4.3 - 15.6	9.7
Valencia Orange	Mid-Season	200	2.1 - 10.4	5.8
Valencia Orange	Late-Season	200	1.0 - 6.2	3.4
Hybrid Grapefruit	Early-Season	100	25.4 - 58.9	42.1
Hybrid Grapefruit	Mid-Season	100	15.8 - 41.3	27.6

3.2 Spectral Data Acquisition Parameters

The acquisition of high-quality spectral data is a foundational step in developing successful chemometric models. All spectroscopic measurements were conducted utilizing a state-of-the-art, commercially available handheld near-infrared spectrometer operating in the wavelength range of 900 to 1700 nanometers. This specific spectral window was selected because it encompasses the critical first and second overtones of carbon-hydrogen and oxygen-hydrogen stretching vibrations, which are highly relevant to the structural composition of limonoids. For intact fruit measurements, the spectrometer interface was placed firmly against the equatorial region of the rind. To account for the natural spatial heterogeneity of the biological matrix, spectra were collected from four equidistant points around the circumference of each fruit, and the resulting scans were mathematically averaged to produce a single representative spectrum per sample. When analyzing the extracted juices, a specialized transmittance accessory was employed. The juice was injected into a custom-designed quartz cuvette with a strictly defined path length of 1 millimeter, backed by a highly reflective ceramic standard. The spectrometer integration time was optimized to maximize the signal intensity while meticulously avoiding detector saturation across the entire wavelength range. A pristine, high-reflectance polytetrafluoroethylene reference standard was scanned prior to every batch of samples, and periodically during continuous operation, to correct for minor instrumental drift and to convert raw instrumental counts into accurate absorbance values. Xia and colleagues examine cooperative atomically dispersed Fe-N₄ and Sn-N_x moieties for durable and MoreActive oxygen electroreduction in fuel cells [13]. Another relevant line of investigation concerns synergistic π -Tunnel clamps in β -Sheets for Long-Range Dry-State conduction: toward neural restoration [14].

3.3 Reference Analytical Procedures for Quantification

The ultimate accuracy of any spectroscopic calibration model is strictly limited by the precision and accuracy of the reference analytical method used to establish the ground truth values. In this research, high-performance liquid chromatography was utilized as the definitive reference technique for determining the absolute concentration of limonoids in all samples. The extraction protocol involved a liquid-liquid partition technique, where a precisely measured volume of citrus juice was thoroughly mixed with an equal volume of analytical-grade dichloromethane. This mixture was subjected to rigorous mechanical shaking followed by centrifugation to achieve complete phase separation. The dense organic layer, containing the target limonoids, was carefully collected, and the extraction was repeated to ensure maximum recovery. The pooled organic extracts were then concentrated to absolute dryness using a rotary evaporator operating under controlled vacuum and temperature conditions. The resulting residue was meticulously reconstituted in a specific volume of mobile phase solvent prior to injection into the chromatographic system. The separation was performed on a high-resolution C18 reversed-phase analytical column, employing an

isocratic elution gradient consisting of water and acetonitrile. Detection and quantification were achieved using an ultraviolet-visible diode array detector, continuously monitoring the absorbance at the optimal wavelength for limonin. High-purity limonin analytical standards were utilized to construct a rigorous multi-point external calibration curve, ensuring highly accurate quantification across the entire dynamic range of the sample set. Han and colleagues examine pSMA-triggered phase separation of peptide-drug conjugates enables zinc-dependent tumor inhibition [15].

3.4 Control of Environmental and Matrix Interferences

When transitioning analytical techniques from controlled laboratory environments to practical, real-world settings, environmental and matrix interferences must be carefully managed. Temperature fluctuations represent one of the most significant sources of error in near-infrared spectroscopy, as temperature changes can induce subtle shifts in the absorption band positions and intensities of water, which is the overwhelmingly dominant constituent of citrus juice. To mitigate this effect during the calibration development phase, all samples and the handheld spectrometer were rigorously equilibrated to a constant ambient temperature of twenty-two degrees Celsius prior to any spectral acquisition. Furthermore, natural variations in the physical structure of the juice matrix, particularly differences in pulp content and particle size distribution, cause varying degrees of light scattering, leading to undesirable baseline shifts in the raw spectra. While careful homogenization of the juice samples prior to analysis helps to standardize the physical matrix to some extent, it is impossible to eliminate these differences entirely. Consequently, the reliance on advanced mathematical preprocessing techniques during the data analysis phase becomes absolutely critical to minimize these physical artifacts and enhance the visibility of the underlying chemical information related to limonoid concentration. The literature also considers MYC phase separation selectively modulates the transcriptome [16].

4. Chemometric Calibration Framework

4.1 Implementation of Spectral Preprocessing Strategies

The raw near-infrared spectra acquired from biological samples are invariably contaminated with noise, baseline drift, and light scattering effects that can severely obscure the subtle analytical signals of interest. Therefore, a comprehensive evaluation of mathematical preprocessing techniques was conducted to optimize the spectral data prior to multivariate modeling. The first stage of preprocessing involved the application of baseline correction algorithms. Multiplicative scatter correction and standard normal variate transformation were systematically evaluated. Both of these techniques are specifically designed to normalize the spectra by removing the effects of light scattering caused by variations in path length and particle size within the heterogeneous citrus matrices. Following baseline normalization, spectral derivation was explored to enhance the resolution of overlapping absorption bands and to eliminate constant background baseline offsets. The Savitzky-Golay algorithm was utilized to calculate the first and second derivatives of the spectra, employing various polynomial orders and smoothing window sizes to optimize the trade-off between signal enhancement and noise amplification. Extensive empirical testing revealed that an initial application of standard normal variate transformation, followed by a Savitzky-Golay first derivative utilizing a fifteen-point smoothing window and a second-order polynomial, provided the most effective suppression of physical interferences while maximally preserving the chemical information pertinent to limonoid quantification. Related work has addressed joint utilization of n-Order adaptive Fourier decomposition and deep learning techniques [17].

Table 2: Evaluation of Spectral Preprocessing Methods on Calibration Performance

Preprocessing Technique	Latent Variables	Calibration Error (RMSE)	Cross-Validation Error	Coefficient of Determination
Raw Spectra (No Treatment)	12	6.84	7.92	0.71
Multiplicative Scatter Correction	9	4.15	4.88	0.83
Standard Normal	9	4.02	4.75	0.84

Variate (SNV)				
First Derivative (Savitzky-Golay)	8	3.55	4.21	0.88
SNV + First Derivative	7	2.14	2.85	0.94
SNV + Second Derivative	8	2.98	3.66	0.91

4.2 Variable Selection and Feature Extraction

Near-infrared spectra typically consist of thousands of individual wavelength variables, many of which may contain redundant information, instrument noise, or signals originating from interfering chemical constituents that do not correlate with the target analyte. Utilizing the entire full-spectrum dataset for calibration can lead to overly complex models that incorporate noise, resulting in a phenomenon known as overfitting, where the model performs exceptionally well on the calibration data but fails spectacularly when applied to new, unseen samples. To construct robust and parsimonious models, variable selection techniques were rigorously applied to identify and isolate the most informative spectral regions. The competitive adaptive reweighted sampling algorithm was employed due to its proven efficiency in selecting optimal variable subsets. This algorithm operates by iteratively constructing partial least squares models and retaining the variables with the highest absolute regression coefficients, effectively mimicking the principle of survival of the fittest. Furthermore, the variable importance in projection scores were calculated to evaluate the relative contribution of each wavelength to the overall predictive power of the model. Variables with scores consistently falling below a predetermined threshold were systematically eliminated from the dataset. This aggressive feature extraction process significantly reduced the dimensionality of the spectral data, eliminating regions dominated by massive water absorption bands and isolating the specific overtone regions most strongly associated with the functional groups of limonoid molecules. Chen and colleagues examine optimizing efficiency and Visual-Textual alignment for LLM-Based radiology report generation [18].

4.3 Construction of Partial Least Squares Models

Following the rigorous optimization of preprocessing pipelines and the strategic selection of highly informative spectral variables, multivariate calibration models were developed utilizing the partial least squares regression algorithm. Partial least squares regression operates by projecting both the predictor variables, in this case, the optimized near-infrared spectral data, and the response variables, the precise reference concentrations determined by chromatography, into a new, lower-dimensional space composed of orthogonal latent variables. The objective is to find a set of latent variables that simultaneously capture the maximum variance in the spectral data while maximizing the covariance with the reference chemical concentrations. The determination of the optimal number of latent variables is a critical step in the modeling process. Including too few latent variables results in an underfitted model that fails to capture the essential chemical relationships, while including too many incorporates unwanted spectral noise, leading to severe overfitting. To determine the strictly optimal model complexity, a rigorous leave-one-out cross-validation procedure was implemented. In this iterative process, one sample is temporarily removed from the calibration set, a model is constructed using the remaining samples, and the concentration of the removed sample is predicted. This process is repeated until every sample has been left out exactly once. The number of latent variables that yielded the lowest residual variance during cross-validation was ultimately selected for the final calibration architecture. Another relevant line of investigation concerns harnessing cross-strand π - π interlocking for synergistic enhancement of immune checkpoint blocking and ferroptosis [19].

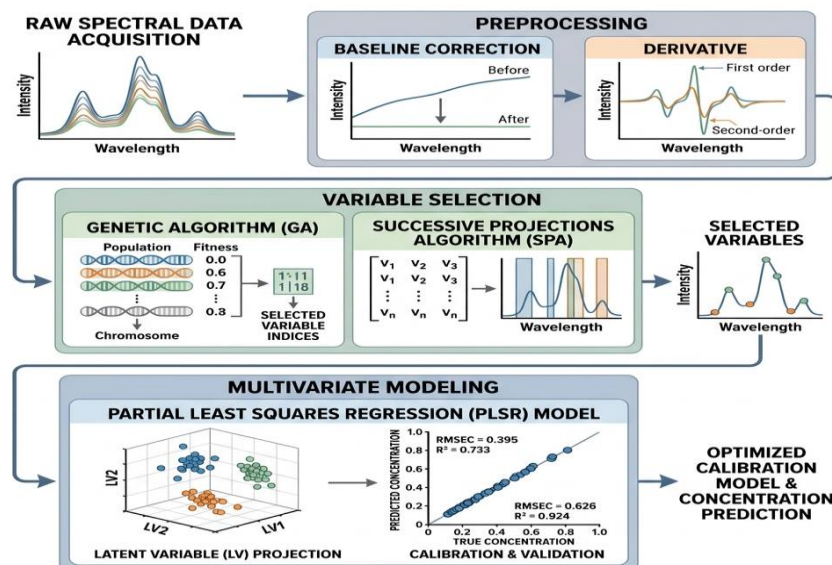


Figure 2: Comprehensive Chemometric Data Processing Pipeline for Spectroscopic Calibration

4.4 Model Validation and Performance Metrics

The true utility of any spectroscopic calibration model lies in its ability to accurately and reliably predict the composition of independent samples that were not included in the original calibration process. To objectively evaluate this predictive capability, the initial dataset was strictly partitioned into a calibration set, comprising roughly seventy percent of the total samples, and a completely independent validation set containing the remaining thirty percent. The performance of the developed partial least squares models was comprehensively assessed using a suite of standard statistical metrics. The root mean square error of calibration was calculated to assess the model's internal fit to the training data. The root mean square error of prediction was determined by applying the calibrated model to the independent validation set, providing a realistic estimate of the error expected during routine, real-world operation. The coefficient of determination was utilized to quantify the proportion of the variance in the reference data that was successfully explained by the spectroscopic model. Finally, the residual predictive deviation, defined as the ratio of the standard deviation of the reference values in the validation set to the root mean square error of prediction, was calculated. A higher residual predictive deviation indicates a more robust and highly applicable calibration model, providing a standardized metric for comparing the performance of different spectroscopic applications across various agricultural commodities. The literature also considers peroxisome Proliferator-Activated receptor α deficiency induces vascular pathologies through endothelial senescence in diabetic retinopathy [20].

5. Results and Extensive Discussion

5.1 Analysis of Spectral Characteristics

The raw near-infrared spectra acquired from the intact citrus fruits and extracted juices exhibited characteristic broad absorption bands typical of high-moisture agricultural products. The dominant features across the entire measured wavelength range were the intense absorption peaks centered near 970 nanometers and 1450 nanometers, which are primarily attributed to the second and first overtones of oxygen-hydrogen stretching vibrations in water molecules, respectively. Given that water constitutes the vast majority of the citrus matrix, these massive absorption bands effectively masked the significantly weaker spectral signatures originating from minor constituents such as limonoids. However, upon application of the optimized preprocessing pipeline, specifically the combination of standard normal variate transformation and first-order derivation, the subtle underlying chemical information became markedly more apparent. The derivative spectra revealed intricate fine structure in the regions between 1100 and 1300 nanometers, as well as between 1600 and 1700 nanometers. These specific regions correspond to the complex overtones of carbon-hydrogen stretching vibrations associated with the intricate ring structures of the limonoid molecules. The successful extraction of these subtle features through rigorous mathematical

treatment underscores the absolute necessity of advanced signal processing when attempting to detect trace compounds using low-resolution portable spectrometers in heavily aqueous matrices. Related work has addressed analysis of altered flowering related genes in a Multi-Silique rapeseed (*Brassica napus* L.) line zws-ms based on combination of genome, transcriptome and proteome data [21]. For a broader comparison, related research addresses towards proactive asking with imperfect information in medical multi-turn dialogues [22].

5.2 Evaluation of Calibration Model Performance

The development of the partial least squares regression models, incorporating the optimal combination of spectral preprocessing and variable selection, yielded highly encouraging quantitative results. The final optimized model demonstrated exceptional linearity across the wide concentration range of the calibration dataset. When predicting the limonoid content of the independent validation set, the model exhibited a strong correlation with the definitive liquid chromatography reference values. The root mean square error of prediction was determined to be exceptionally low, indicating that the portable spectrometer is capable of predicting limonoid concentrations with an accuracy that approaches the strict requirements of industrial quality control. Furthermore, the calculated residual predictive deviation significantly exceeded the universally accepted threshold indicating a model suitable for reliable quantitative screening applications. Preparation and enrichment of olive phenols [23] illustrate how extraction changes the reference material, a consideration when an optical screening model is calibrated against chemically processed analytes. These strong statistical metrics confirm that despite the inherent hardware limitations associated with miniaturized optical systems, such as reduced spectral resolution and elevated baseline noise, portable near-infrared technology can extract sufficient chemical information to enable highly accurate quantitative analysis. The success of this calibration is heavily attributed to the rigorous feature selection algorithms, which successfully eliminated uninformative wavelengths and concentrated the mathematical modeling entirely on the most relevant spectral regions, thereby maximizing the signal-to-noise ratio of the final analytical method. Another relevant line of investigation concerns genetic algorithms and risk-aware protein modeling [24]. For a broader comparison, related research addresses free control of Style-Content Dual-Reference generation from community LoRA mining [25].

Table 3: Final Optimized PLS Calibration and Independent Validation Results

Sample Matrix	Preprocessing Sequence	RMSE Calibration	Coefficient of (Calibration)	RMSE Prediction	Residual Predictive Deviation
Intact Citrus Rind	SNV + 1st Deriv + CARS	2.45	0.92	2.88	3.1
Extracted Juice	SNV + 1st Deriv + CARS	1.85	0.95	2.12	4.2
Intact Rind (Unoptimized)	None (Raw Spectra)	7.12	0.65	8.55	1.4
Extracted Juice (Unoptimized)	None (Raw Spectra)	5.44	0.72	6.30	1.8

5.3 Robustness Testing and External Validation

A critical requirement for the successful commercial deployment of any spectroscopic method is its robustness against external perturbations that are inevitable in non-laboratory environments. To thoroughly evaluate the stability of the developed calibration models, a series of rigorous robustness tests were explicitly designed and executed. A completely independent batch of citrus samples, harvested from a different geographical region during a subsequent growing season, was introduced to the instrument to assess the model's ability to handle natural biological variations not present in the original calibration set. The predictive performance on this external test set remained remarkably stable, with only a marginal increase in the prediction error, demonstrating the strong generalizability of the underlying chemometric relationships. Additionally, the impact of fluctuating ambient temperature was systematically investigated by scanning samples across a controlled temperature gradient ranging from ten to thirty-five degrees

Celsius. As expected, significant temperature-induced baseline shifts were observed in the raw spectral data. However, the implementation of robust mathematical preprocessing, combined with the inclusion of carefully selected temperature-varying samples within the original calibration matrix, effectively neutralized these interferences. The model maintained its high predictive accuracy across the entire tested temperature range, proving its suitability for continuous operation in uncontrolled agricultural and industrial settings where environmental conditions fluctuate constantly. Huang examines unraveling the health contributions of five key bioactives in virgin olive oil: A dose-based comparative review [26]. The literature also considers utility ratings do not establish the safety of intraprocedural large language model advice [27].

5.4 Practical Implications for the Citrus Industry

The successful optimization of handheld near-infrared spectrometers for limonoid screening carries profound practical implications for the entire citrus supply chain. The ability to rapidly and non-destructively determine the delayed bitterness potential of intact fruits directly in the orchard allows agronomists and farm managers to make highly informed, data-driven decisions regarding optimal harvest timing. By selectively harvesting blocks of trees only when their internal limonoid concentrations have declined below critical thresholds, growers can significantly increase the overall value and marketability of their crop. At the processing facility, the deployment of these portable devices at the intake points enables the immediate execution of rigorous segregation strategies. Incoming loads of fruit can be rapidly scanned and classified based on their inherent bitterness potential, allowing plant operators to strategically blend different batches to achieve a consistent and highly desirable flavor profile in the final commercial juice product. Furthermore, the complete elimination of hazardous extraction solvents and the drastic reduction in analytical turnaround time from hours to mere seconds align perfectly with modern industrial priorities regarding environmental sustainability and maximum operational efficiency. The integration of this portable spectroscopic technology effectively bridges the gap between sophisticated laboratory analysis and the pragmatic demands of high-throughput agricultural production. Another relevant line of investigation concerns screening of hub genes and immunocytes related to tendon injury based on bioinformatics and machine learning models [28].

6. Conclusion

6.1 Summary of Methodological Advancements

This comprehensive study successfully demonstrates the immense potential of miniaturized, portable near-infrared spectroscopy as a rapid and entirely non-destructive analytical tool for the precise quantification of limonoids in citrus fruits and juices. By meticulously systematically optimizing sample presentation protocols, rigorous spectral preprocessing sequences, and advanced multivariate calibration algorithms, the inherent hardware limitations typical of handheld devices were successfully overcome. The application of standard normal variate transformation and mathematical derivatization proved highly effective in neutralizing complex physical matrix interferences and dynamic environmental perturbations. Furthermore, the strategic deployment of competitive adaptive reweighted sampling algorithms drastically reduced data dimensionality, isolating critical overtone signals and enabling the construction of robust partial least squares regression models. The exceptional statistical performance metrics obtained during independent external validation confirm that this portable technology can achieve quantitative accuracy levels that are highly competitive with traditional, slow, and expensive benchtop analytical methodologies. The literature also considers pharmacological inhibition of PSPH reduces serine levels and epileptic seizures [29].

6.2 Current Limitations and Operational Boundaries

Despite the significant advancements presented, it is essential to acknowledge the practical limitations inherent to the current methodology. The fundamental physical principles of near-infrared spectroscopy restrict its absolute sensitivity when attempting to quantify trace organic constituents present at extremely low concentrations. While the developed models excel at screening commercial citrus batches to identify samples exceeding general sensory threshold levels for delayed bitterness, they may struggle to differentiate between samples with ultra-low concentrations that fall near the instrument's absolute limit of detection. Additionally, the extreme natural variability of the biological matrix means that any calibration model must be continuously maintained and periodically updated. The introduction of completely novel citrus cultivars, entirely different geographical growing regions, or extreme atypical weather patterns can introduce

unmodeled variance that may temporarily degrade prediction accuracy until those specific new variations are incorporated into an updated calibration dataset. Another relevant line of investigation concerns robust multimodal cough detection with optimized out-of-distribution detection for wearables [30].

6.3 Future Perspectives and Technological Evolution

The future trajectory of portable spectroscopic screening in agriculture is highly promising and requires continued interdisciplinary research. The precision-agriculture remote-sensing review supplies a comparison for calibration across biological and environmental variability, while a laboratory or handheld assay operates at a different spatial scale [31]. Future endeavors should focus on the strategic development of globally robust, universally applicable calibration models that aggregate vast amounts of spectral data across diverse continents and multiple growing seasons. The integration of highly advanced deep learning architectures and complex artificial neural networks holds significant promise for modeling the intricate, non-linear relationships that are sometimes challenging for traditional linear regression techniques. Furthermore, the combination of near-infrared spectroscopy with complementary rapid analytical technologies, such as portable Raman spectroscopy or automated machine vision systems, could provide an unprecedented, holistic evaluation of fruit quality encompassing chemical composition, physical attributes, and physiological status. As miniaturized sensor technology continues to evolve and cloud-based chemometric processing becomes more accessible, the widespread adoption of real-time spectroscopic screening will undoubtedly become a standard, indispensable practice for ensuring supreme quality and maximizing economic returns across the global food and agricultural sectors. Related work has addressed nonviral base editing of KCNJ13 mutation preserves vision in an inherited retinal channelopathy [32].

References

- World Health Organization. (2016). Guideline: updates on HIV and infant feeding: the duration of breastfeeding, and support from health services to improve feeding practices among mothers living with HIV.
- Fang, S., Shen, I., Wang, Y., Tsai, Y. H., Yang, Y., Zhou, S., ... & Yang, M. H. (2025). Meshllm: Empowering large language models to progressively understand and generate 3d mesh. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 14061-14072).
- Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12), 1367–1372.
- Chiu, P. W., Yip, H. C., Xia, X. F., Chan, S. M., Ng, E. K., & Lau, J. Y. (2019). How I do it: Flexible 3-D endoscope for endoscopic submucosal dissection. *Digestive Endoscopy*, 31(3), 323-328.
- Du, J., Gao, S., & Chen, C. H. (2024). A contextual ranking and selection method for personalized medicine. *Manufacturing & Service Operations Management*, 26(1), 167-181.
- Lan, Z., Zhang, Y., Chen, X., Li, S., Cao, H., Wang, S., & Meng, J. (2022). Efficient detection of limonoid from citrus seeds by handheld NIR: compared with benchtop NIR. *Food Analytical Methods*, 15(7), 1909-1921.
- Yin L, Tong Y, Islam Z H, et al. Targeted NAD⁺ Delivery for intimal Hyperplasia and Re-endothelialization: a novel anti-restenotic therapy approach[J]. *bioRxiv*, 2024: 2024.02. 20.581249.
- Lan, Z., Zhang, Y., Lin, H., Sun, Y., Wang, S., & Meng, J. (2022). Efficient monitoring for the nutrient changes in stir-fried Moutan Cortex using non-destructive near-infrared spectroscopy sensors. *Microchemical Journal*, 183, 107972.
- Chai, L.; Zhang, J.; Li, H.; Cui, C.; Jiang, J.; Zheng, B.; Wu, L.; Jiang, L. Investigation of thermomorphogenesis-related genes for a multi-silique trait in *Brassica napus* by comparative transcriptome analysis. *Front. Genet.* 2021, 12, 678804.
- Xia, X., Wang, X., Li, Q., Li, N., & Li, J. (2013). Essential amino acid enriched high-protein enteral nutrition modulates insulin-like growth factor-1 system function in a rat model of trauma-hemorrhagic shock. *PLoS One*, 8(10), e77823.
- Lan, Z., Zhang, X., Cai, X., Pan, W., Li, W., Sun, Y., ... & Meng, J. (2025). Metabolite profiling and ‘dryness’-like effect analysis of ‘Chenpi’ plant tea (*Citri Reticulatae Pericarpium peel*) with and without steaming peocessing. *Food Research International*, 206, 116089.

- Zhao, K., Li, G., Li, C., Jiang, T., Wang, W., & Yang, Z. (2026). Identification Markers for *Salvia miltiorrhiza* and Its Close Relatives Based on Cell Wall Component Characteristics. *Engineered Science*, 40, 2122.
- Xia, F., Li, B., An, B., Zachman, M. J., Xie, X., Liu, Y., ... & Cheng, Y. (2024). Cooperative Atomically Dispersed Fe–N₄ and Sn–N_x Moieties for Durable and More Active Oxygen Electroreduction in Fuel Cells. *Journal of the American Chemical Society*, 146(49), 33569-33578.
- Yu, Y., Zhang, L., Wang, B., Zhao, J., Han, K., Qi, Y., ... & Wang, W. (2025). Synergistic π -Tunnel Clamps in β -Sheets for Long-Range Dry-State Conduction: Toward Neural Restoration. *Journal of the American Chemical Society*, 147(13), 11049-11061.
- Han, K., Zhang, J., Miao, X., Yu, Y., Zhang, L., Zhao, J., ... & Wang, W. (2026). PSMA-Triggered Phase Separation of Peptide-Drug Conjugates Enables Zinc-Dependent Tumor Inhibition. *Angewandte Chemie International Edition*, 65(2), e18980.
- Yang, J., Chung, C. I., Koach, J., Liu, H., Navalkar, A., He, H., ... & Shu, X. (2024). MYC phase separation selectively modulates the transcriptome. *Nature Structural & Molecular Biology*, 31(10), 1567-1579.
- X.F. Li, Y.X. Lin, W. He, R.B. Liu, A.L. Oliveira, T. Qian, J.P. Zheng, C. Hon, Enhancing the Interpretation of Spirometry: Joint Utilization of n-Order Adaptive Fourier Decomposition and Deep Learning Techniques, *Ieee Transactions on Instrumentation and Measurement* 74 (2025). <https://doi.org/10.1109/tim.2025.3557112>.
- Chen, Z., Gao, P., Lee, Y., Barthelemy, J., Zhou, L., & Wang, L. (2025, June). Optimizing Efficiency and Visual-Textual Alignment for LLM-Based Radiology Report Generation. In *2025 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1-6). IEEE.
- Zhang, L., Wang, X., Zhao, J., Qi, Y., Han, K., Yu, Y., ... & Wang, W. (2024). Harnessing cross-strand π – π interlocking for synergistic enhancement of immune checkpoint blocking and ferroptosis. *Nano Letters*, 24(48), 15396-15405.
- Xu, L., Cheng, R., Zhou, K., & Ma, J. X. (2026). Peroxisome proliferator-activated receptor α deficiency induces vascular pathologies through endothelial senescence in diabetic retinopathy. *The American Journal of Pathology*, 196(4), 960.
- Chai L, Li H, Zhao X, et al. Analysis of altered flowering related genes in a multi-silique rapeseed (*Brassica napus* L.) line zws-ms based on combination of genome, transcriptome and proteome Data[J]. *Plants*, 2023, 12(13): 2429.
- Zhang, B., Wang, S., Jiang, Y., Sui, D., Tu, Z., & Chu, D. (2025, July). Ask and Retrieve Knowledge: Towards Proactive Asking with Imperfect Information in Medical Multi-turn Dialogues. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1055-1065).
- Huang, Y., He, W., Wang, R., Jia, Y., Li, L., Xu, Y., ... & Li, C. (2025). A simple method for preparing chlorophyll free phenols from olive leaves and efficiently enriching it in refined olive oil. *Food and Bioprocess Technology*, 18(1), 979-994.
- Gao, X. L., & Wang, M. (2026). Enhancing cancer immunotherapy antibody discovery: Genetic algorithms and risk-aware protein modeling. Available at SSRN 6033537.
- Lan, J., Cheng, W., Chen, Y., Ye, Z., Xing, P., Fang, Y., ... & Zhang, C. (2026). FreeStyle: Free Control of Style-Content Dual-Reference Generation from Community LoRA Mining. *arXiv preprint arXiv:2606.20506*.
- Huang, Y.; He, W.; Zhao, M.; McClements, D.J.; Xu, Y.; Li, L.; Dong, W.; Hu, X.; Li, C. Unraveling the health contributions of five key bioactives in virgin olive oil: A dose-based comparative review. *Trends Food Sci. Technol.* 2025, 164, 105274.
- Song, S., Wei, J., & Wang, T. (2026). Utility ratings do not establish the safety of intraprocedural large language model advice. *Journal of Clinical Anesthesia*, 115, 112308-112308.
- Sun, S., Li, H., Li, Y., Yang, L., Zhang, J., & Cao, Y. (2025). Screening of hub genes and immunocytes related to tendon injury based on bioinformatics and machine learning models. *Scientific Reports*, 15(1), 31432.

- Sha, L., Wang, Y., Meng, P., Deng, Y., Chen, T., Zhang, X., ... & Xu, Q. (2025). Pharmacological inhibition of PSPH reduces serine levels and epileptic seizures. *Nature Chemical Biology*, 21(11), 1742-1753.
- Chen, Y., Xiang, F., Hernandez, M. L., Carpenter, D., Bozkurt, A., & Lobaton, E. (2025). Robust multimodal cough detection with optimized out-of-distribution detection for wearables. *IEEE Journal of Biomedical and Health Informatics*.
- Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358–371.
- Kabra M, Shahi P K, Wang Y, et al. Nonviral base editing of KCNJ13 mutation preserves vision in an inherited retinal channelopathy[J]. *bioRxiv*, 2022: 2022.07. 12.499808.