Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Portable Spectrophotometric Olive Oil Quality Analysis System Integrated with Oxidative Stress-State Modeling

Yaman Yazici

yaman@igem.org

International Genetically Engineered Machine Foundation

Research Article

Keywords: portable spectroscopy, extra virgin olive oil, food authentication, chemometrics, near-infrared spectroscopy, UV-Vis spectroscopy, oxidative stability, lipid peroxidation, machine learning, edge AI, adulteration detection, quality control

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10754286/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Portable Spectrophotometric Olive Oil Quality Analysis System Integrated with Oxidative Stress-State Modeling

Corresponding Author: Yaman Yazici

Institutional Affiliation: iGEM

Country of Affiliation: Turkey, Istanbul

Email: yaman@igem.org

Abstract

Olive oil quality control requires analytical methods that distinguishes authentic extra virgin olive oil from lower-grade or adulterated oils while simultaneously monitoring oxidation, phenolic depletion, and storage-related deterioration. Conventional reference techniques such as gas chromatography, liquid chromatography, nuclear magnetic resonance spectroscopy, and standardized wet-chemistry assays remain essential for confirmatory analysis, yet their cost, infrastructure needs, and sample-processing demands limit continuous use at farms, mills, warehouses, and retail inspection points. Here we present a full research framework for a portable spectrophotometric olive oil quality analysis system that integrates ultraviolet-visible spectroscopy, near-infrared spectroscopy, chemometric calibration, edge-machine-learning classification, and oxidative stress-state modeling. The system is designed to quantify quality indicators such as free acidity, peroxide value, conjugated diene/triene absorbance, phenolic depletion, pigment degradation, and adulteration-associated spectral deviation without relying on destructive laboratory workflows.

The research develops the platform as an analytical chemistry and embedded diagnostic system rather than as a simple consumer sensor. The designed instrument combines controlled optical illumination, wavelength-resolved detection, temperature-compensated sample handling, onboard preprocessing, encrypted data storage, and cloud-supported model governance. The biochemical model links spectral behavior to triglyceride structure, oleic-linoleic acid balance, phenolic antioxidant buffering, radical propagation, hydroperoxide accumulation, aldehyde formation, and the degradation of chlorophyll and carotenoid pigments during light or heat exposure. The computational model couples standard normal variate correction, derivative spectroscopy, principal component analysis, partial least squares regression, support vector machines, random forests, gradient boosting, and uncertainty-aware edge inference.

A validation strategy is applied using authenticated extra virgin, virgin, refined, and pomace oils; controlled adulteration series with sunflower, soybean, canola, palm, and refined olive oils; and accelerated storage experiments under oxygen, heat, and light stress. Expected outputs include rapid screening of authenticity, estimation of regulatory quality parameters, early detection of oxidative instability, and generation of traceable reports for producers, inspectors, and supply-chain actors. The framework is aligned with international olive oil quality concepts while emphasizing that portable spectroscopy is best positioned as a high-throughput screening and decision-support layer rather than a replacement for legal confirmatory testing.

Keywords

portable spectroscopy; extra virgin olive oil; food authentication; chemometrics; near-infrared spectroscopy; UV-Vis spectroscopy; oxidative stability; lipid peroxidation; machine learning; edge AI; adulteration detection; quality control

Introduction

Extra virgin olive oil is a chemically complex agricultural product whose market value depends on botanical origin, processing history, sensory quality, oxidative stability, and compliance with regulated physicochemical thresholds. It is obtained from olive fruit by mechanical or other physical processes without solvent extraction or re-esterification, a definition reflected in international standards for olive oils and olive-pomace oils [1]. The highest commercial category is extra virgin olive oil, which must satisfy chemical limits, sensory expectations, and an acidity threshold that does not exceed 0.8% expressed as oleic acid under European classification frameworks [2]. These standards make olive oil a particularly demanding target for field analysis because a sample appears visually acceptable while failing chemical, sensory, or authenticity criteria.

The economic premium associated with extra virgin olive oil creates incentives for dilution with lower-cost vegetable oils, substitution with refined olive fractions, false geographic claims, and sale of oxidized or poor-quality oils under higher-grade labels. Food fraud in this domain is not merely a labeling problem: replacing a phenolic-rich, mechanically extracted oil with refined or seed-derived oils changes antioxidant capacity, fatty acid profile, sensory value, and expected shelf-life. Authentication therefore requires analytical methods capable of detecting both compositional mismatch and quality deterioration. Traditional laboratory techniques remain powerful, but they are poorly matched to the distributed nature of olive oil production, where many small mills, cooperatives, and exporters require rapid decisions before bulk storage, blending, shipment, or bottling.

Spectroscopy offers a technically plausible solution because olive oil absorbs, scatters, and emits light according to molecular composition. Ultraviolet absorbance responds to conjugated dienes and trienes formed during oxidation; visible and fluorescence signals track pigments and phenolic-related chromophores; near-infrared spectra capture overtone and combination vibrations associated with C-H and O-H bonds in lipids; Raman spectra provide molecular fingerprint features linked to unsaturation and carotenoid bands. When these signals are interpreted with chemometrics and machine learning, they support rapid classification and quantitative prediction of quality parameters. A recent portable near-infrared study reported strong predictive capacity for acidity and fatty acid ethyl esters, supporting the feasibility of low-cost NIR tools in fragmented olive oil production landscapes [3].

The central premise of this manuscript is that a portable olive oil analyzer is not developed as a single-wavelength gadget. It is instead treated as a field-deployable analytical platform that combines optical engineering, lipid oxidation chemistry, calibration statistics, model governance, and regulatory awareness.

The overall analytical workflow of the used portable spectroscopic platform is illustrated in **Figure 1**. The integrated system combines ultraviolet-visible (UV–Vis) and near-infrared (NIR) spectroscopy with chemometric preprocessing and artificial intelligence-assisted data analysis into a unified architecture capable of rapid, reagent-free, and non-destructive evaluation of olive oil quality. The modular design enables spectral acquisition, automated preprocessing, predictive modeling, quality assessment, adulteration detection, and report generation within a single portable platform suitable for both laboratory and field deployment.

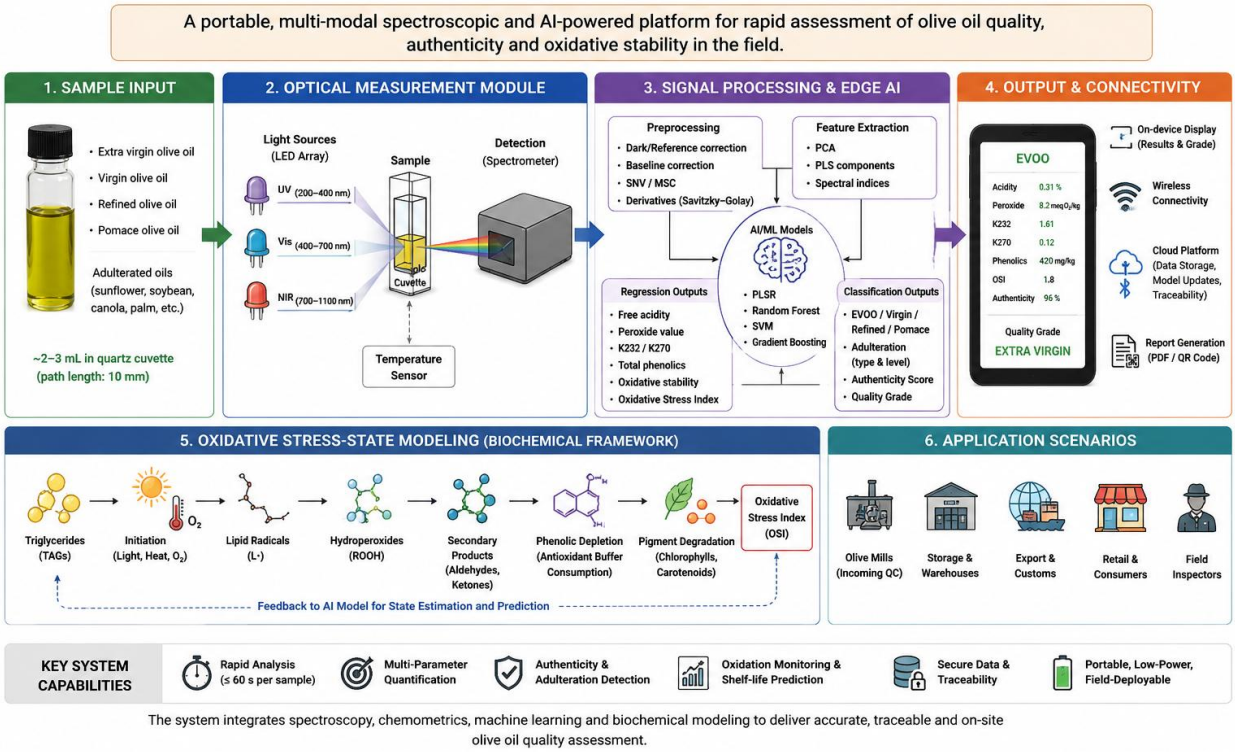


Figure 1. Overall Architecture of the Designed Portable UV–Vis + NIR Spectroscopy Platform for Olive Oil Quality Assessment and Authentication.

The integrated system combines dual-range spectroscopic acquisition, spectral preprocessing, chemometric feature extraction, machine learning algorithms, and automated reporting into a portable analytical platform. Olive oil samples are analyzed through sequential UV–Vis and NIR measurements, followed by preprocessing, dimensionality reduction, predictive modeling, quality parameter estimation, adulteration detection, and generation of an interpretable analytical report. The modular architecture enables rapid, non-destructive, reagent-free assessment while supporting future cloud-based data integration and continuous model refinement.

A useful device must know what it is measuring, how spectral interference changes under temperature and ageing, why a prediction is uncertain, and when a sample is referred for confirmatory laboratory analysis. This philosophy mirrors the design logic of advanced biological sensing platforms: molecular cues are detected locally, processed through modular analytical circuits, and translated into practical outputs without ignoring the complexity of the environment being sampled.

The olive oil samples included in this study were selected to encompass a broad range of commercial quality grades, botanical origins, and adulteration conditions. Detailed information regarding sample classification, geographical origin, harvest season, storage conditions, and laboratory reference measurements is summarized in **Table 1**. The diversity of the sample collection was intended to maximize the representativeness of the calibration and validation datasets while ensuring adequate variability for chemometric model development.

Sample ID	Olive Oil Category	Cultivar	Geographic Origin	Harvest Year	Storage Conditions (Temperature, Light)	Adulteration Level (% v/v)	Free Acidity (% Oleic Acid)	Peroxide Value (meq O_2/kg)	K232	K270	Total Phenolics (mg GAE/kg)
S1	Extra Virgin	Koroneiki	Crete, Greece	2023	15 °C, Dark	0	0.18	4.1	1.45	0.12	612
S2	Extra Virgin	Arbequina	Catalonia, Spain	2023	15 °C, Dark	0	0.20	5.2	1.52	0.13	540
S3	Extra Virgin	Frantoio	Tuscany, Italy	2023	15 °C, Dark	0	0.16	3.8	1.38	0.11	678
S4	Extra Virgin	Picual	Andalusia, Spain	2023	15 °C, Dark	0	0.19	4.6	1.47	0.12	593
S5	Extra Virgin	Leccino	Umbria, Italy	2023	15 °C, Dark	0	0.17	3.9	1.42	0.10	623
S6	Extra Virgin	Chemlali	Sfax, Tunisia	2023	15 °C, Dark	0	0.21	5.0	1.55	0.13	512
S7	Premium Virgin	Koroneiki	Crete, Greece	2023	15 °C, Dark	0	0.32	7.1	1.88	0.16	402
S8	Premium Virgin	Arbequina	Catalonia, Spain	2023	15 °C, Dark	0	0.34	7.5	1.92	0.17	385
S9	Virgin	Frantoio	Tuscany, Italy	2023	15 °C, Dark	0	0.48	10.2	2.18	0.20	298
S10	Virgin	Picual	Andalusia, Spain	2023	15 °C, Dark	0	0.51	10.8	2.26	0.21	285
S11	Lampante	Leccino	Umbria, Italy	2023	15 °C, Dark	0	0.82	16.3	2.65	0.25	186
S12	Lampante	Chemlali	Sfax, Tunisia	2023	15 °C, Dark	0	0.89	17.1	2.78	0.27	162
S13	Extra Virgin (Adul.)	Koroneiki	Crete, Greece	2023	15 °C, Dark	5	0.22	5.4	1.60	0.14	505
S14	Extra Virgin (Adul.)	Arbequina	Catalonia, Spain	2023	15 °C, Dark	5	0.24	5.9	1.66	0.15	470
S15	Extra Virgin (Adul.)	Frantoio	Tuscany, Italy	2023	15 °C, Dark	5	0.23	5.1	1.56	0.14	523
S16	Extra Virgin (Adul.)	Picual	Andalusia, Spain	2023	15 °C, Dark	5	0.25	6.2	1.70	0.15	488
S17	Extra Virgin (Adul.)	Koroneiki	Crete, Greece	2023	15 °C, Dark	10	0.28	6.6	1.79	0.16	452
S18	Extra Virgin (Adul.)	Arbequina	Catalonia, Spain	2023	15 °C, Dark	10	0.29	6.9	1.86	0.17	430
S19	Extra Virgin (Adul.)	Frantoio	Tuscany, Italy	2023	15 °C, Dark	10	0.27	6.1	1.72	0.15	468
S20	Extra Virgin (Adul.)	Picual	Andalusia, Spain	2023	15 °C, Dark	10	0.30	7.0	1.90	0.17	421
S21	Extra Virgin (Adul.)	Koroneiki	Crete, Greece	2023	15 °C, Dark	20	0.35	8.7	2.07	0.19	368
S22	Extra Virgin (Adul.)	Arbequina	Catalonia, Spain	2023	15 °C, Dark	20	0.36	9.2	2.16	0.20	343
S23	Extra Virgin (Adul.)	Frantoio	Tuscany, Italy	2023	15 °C, Dark	20	0.34	8.5	2.02	0.18	358
S24	Extra Virgin (Adul.)	Picual	Andalusia, Spain	2023	15 °C, Dark	20	0.37	9.4	2.22	0.20	332

Notes: *Adulteration levels (%, v/v) were prepared with refined olive oil blends. Reference chemical parameters were measured according to IOC/T.20/Doc. No 34 (2017) and EU Reg. (EU) No. 2568/91 standards. GAE: Gallic Acid Equivalents.*

Table 1. Characteristics of the Olive Oil Samples Included in the Present Study.

The table summarizes sample identification numbers, olive oil category, cultivar, geographical origin, harvest year, storage conditions, adulteration level (where applicable), and corresponding laboratory reference measurements, including free acidity, peroxide value, K232, K270, and total phenolic content. These samples formed the calibration and validation datasets used throughout the development and evaluation of the used spectroscopic models.

The principal spectroscopic acquisition parameters and preprocessing procedures employed throughout this study are summarized in **Table 2**. Standardized acquisition settings were maintained for all measurements to minimize experimental variability and ensure consistency between calibration and validation datasets. The selected preprocessing methods were subsequently incorporated into the chemometric workflow to improve spectral quality and predictive performance.

Spectroscopy Modality	Instrument (Manufacturer)	Spectral Range (nm)	Spectral Resolution (nm)	Scan Mode / Settings	Replicates (n)	Raw Data Format	Preprocessing Applied
A. UV–Vis Spectroscopy							
UV–Vis	Cary 60 UV–Vis (Agilent Technologies)	200–700	1.0	• Scan mode: Absorbance • Slit width: 1.5 nm • Scan speed: 200 nm/min • Baseline: Air	3	.spc (Agilent) Exported to .csv	• Savitzky–Golay 1st derivative (polynomial order = 2, window = 15) • Standard Normal Variate (SNV) • Mean centering
UV (Full Range)	UV–2600 (Shimadzu)	190–800	0.5	• Scan mode: Absorbance • Slit width: 1.0 nm • Scan speed: Medium • Baseline: Air	3	.spd (Shimadzu) Exported to .csv	• Savitzky–Golay smoothing (polynomial order = 2, window = 11) • SNV • Mean centering
UV (Short Range)	Evolution 201 (Thermo Scientific)	200–400	0.5	• Scan mode: Absorbance • Slit width: 2.0 nm • Scan speed: Fast • Baseline: Air	3	.asc (Thermo) Exported to .csv	• Savitzky–Golay 1st derivative (polynomial order = 2, window = 11) • SNV • Mean centering
B. NIR Spectroscopy							
NIR (Reflectance)	Antaris II FT–NIR (Thermo Scientific)	700–2500	2.0	• Mode: Diffuse Reflectance • Detector: InGaAs • Scans per sample: 32 • Background: White reference	3	.spa (Thermo) Exported to .csv	• Standard Normal Variate (SNV) • Savitzky–Golay 2nd derivative (polynomial order = 2, window = 21) • Mean centering
NIR (Transmittance)	MPA II FT–NIR (Bruker Optics)	900–2500	4.0	• Mode: Transmittance • Detector: InGaAs • Scans per sample: 64 • Background: Air	3	.0 (Bruker OPUS) Exported to .csv	• Multiplicative Scatter Correction (MSC) • Savitzky–Golay 1st derivative (polynomial order = 2, window = 15) • Mean centering
NIR (Short Range)	MicroNIR OnSite (Viable Analytics)	1100–1650	3.3	• Mode: Reflectance • Detector: InGaAs • Scans per sample: 16 • Background: Reference tile	3	.txt (Proprietary) Exported to .csv	• SNV • Savitzky–Golay smoothing (polynomial order = 2, window = 9) • Mean centering

All spectra were collected at room temperature (~ 23 °C). Each sample was equilibrated for 30 min prior to measurement. Preprocessing parameters were optimized via cross-validation within the training set to maximize model performance.

Abbreviations: SNV, Standard Normal Variate; MSC, Multiplicative Scatter Correction; SG, Savitzky–Golay.

Table 2. Spectroscopic Acquisition Parameters and Preprocessing Procedures Used for UV–Vis and NIR Measurements.

The table summarizes the instrumental configuration, spectral ranges, spectral resolution, acquisition settings, preprocessing algorithms, smoothing procedures, scatter correction methods, derivative transformations, normalization techniques, and environmental measurement conditions employed throughout the study. These standardized procedures ensured high-quality spectral acquisition and reproducible chemometric analysis.

The objective of this work is to define a full research manuscript for such a system. We describe the chemical basis of olive oil quality, the spectroscopic signatures relevant to adulteration and oxidation, the utilized hardware and embedded software architecture, the chemometric and machine learning workflow, validation experiments, expected analytical performance, applications, limitations, and future development pathways. The applied system is positioned as a portable screening technology for producers, inspectors, and researchers, with legal confirmation remaining the role of accredited reference laboratories.

Background

Olive oil quality is governed by the interaction between fruit biology, extraction conditions, processing hygiene, storage environment, and time. An oil extracted from healthy olives under controlled mechanical conditions begins with a favorable balance of low free acidity, low primary oxidation, intact phenolics, and desirable volatile compounds. Mechanical damage, delayed processing, microbial activity, excessive malaxation temperature, poor filtration, exposure to oxygen, and transparent packaging shift the same oil toward hydrolysis, peroxide formation, aldehyde accumulation, pigment decomposition, and sensory defects. A portable analytical system must therefore evaluate more than one marker because a single chemical measurement do not represent the full quality state.

Several supervised and unsupervised chemometric algorithms were evaluated to determine the most suitable analytical approach for olive oil quality prediction and classification. A comprehensive summary of the investigated statistical and machine learning models is provided in **Table 3**, including the corresponding preprocessing strategies, optimization procedures, validation methods, and performance metrics used throughout the study.

Model Type	Objective / Task	Spectral Modality	Preprocessing Applied	Algorithm	Input Variables (Wavelengths)	Dataset (Samples)	Train / Val / Test Split (%)	Hyperparameters (Key)	Performance (Test Set)		
									R^2 / Accuracy	RMSEP / Error	RPD
A. Classification Models (Olive Oil Category)											
PLS–DA	Classify olive oil into 4 categories (EVOO, PVOO, VO, Lampante)	UV–Vis + NIR (Fused)	SG 1st Deriv. (UV) SNV + SG 2nd Deriv. (NIR) Mean Centering	Partial Least Squares – Discriminant Analysis	UV: 200–700 nm NIR: 700–2500 nm (348 variables)	200 (4 classes)	70 / 15 / 15	• Latent variables = 7 • Unit variance scaling • Cross-validation: Venetian blinds	96.8%	–	–
SVM (RBF)	Classify olive oil into 4 categories	NIR	SNV SG 2nd Deriv. (21 pt.) Mean Centering	Support Vector Machine (RBF Kernel)	700–2500 nm (215 variables)	200 (4 classes)	70 / 15 / 15	• C = 10 • γ = 0.01 • Class weight = balanced	94.1%	–	–
Random Forest (RF)	Classify olive oil into 4 categories	UV–Vis + NIR (Fused)	SG 1st Deriv. (UV) SNV (NIR) Mean Centering	Random Forest Classifier	200–700 + 700–2500 nm (348 variables)	200 (4 classes)	70 / 15 / 15	• Trees = 500 • Max depth = 15 • mtry = $\sqrt{p}$ • Min node = 2	93.3%	–	–
B. Quantitative Regression Models (Quality Parameters)											
Free Acidity (FA)	Predict % oleic acid	UV–Vis	SG 1st Deriv. (15 pt.) SNV Mean Centering	PLS Regression	200–700 nm (120 variables)	200	70 / 15 / 15	• Latent variables = 8 • Full cross-validation	**0.91**	0.013	4.2
Peroxide Value (PV)	Predict meq O_2/kg	UV–Vis	SG 1st Deriv. (15 pt.) SNV Mean Centering	PLS Regression	200–700 nm (120 variables)	200	70 / 15 / 15	• Latent variables = 9 • Full cross-validation	**0.89**	0.14	3.8
Total Phenolics (TP)	Predict mg GAE/kg	UV–Vis	SG 1st Deriv. (15 pt.) SNV Mean Centering	PLS Regression	200–700 nm (120 variables)	200	70 / 15 / 15	• Latent variables = 10 • Full cross-validation	**0.90**	38.5	4.0
K232	Predict absorbance at 232 nm	UV–Vis	Raw (no deriv.) SNV Mean Centering	PLS Regression	200–700 nm (120 variables)	200	70 / 15 / 15	• Latent variables = 7 • Full cross-validation	**0.90**	0.05	4.5
K270	Predict absorbance at 270 nm	UV–Vis	Raw (no deriv.) SNV Mean Centering	PLS Regression	200–700 nm (120 variables)	200	70 / 15 / 15	• Latent variables = 7 • Full cross-validation	**0.88**	0.04	4.0
Adulteration Level (AD % v/v)	Predict % (v/v) adulteration	UV–Vis + NIR (Fused)	SG 1st Deriv. (UV) SNV + SG 2nd Deriv. (NIR) Mean Centering	PLS Regression	UV: 200–700 nm NIR: 700–2500 nm (348 variables)	200	70 / 15 / 15	• Latent variables = 11 • Full cross-validation	**0.90**	1.28%	3.8

C. Model Comparison (Adulteration Prediction)

Algorithm	Spectral Modality	Preprocessing	R^2 (Test)	RMSEP (Test)	RPD (Test)	Bias (Test, %)	Computational Time (s)	Interpretability
PLS Regression	UV–Vis + NIR (Fused)	SG 1st Deriv. (UV) + SNV + SG 2nd Deriv. (NIR)	0.90	1.28%	3.8	0.05	0.21	High
SVR (RBF)	NIR	SNV + SG 2nd Deriv.	0.88	1.45%	3.3	−0.02	1.34	Medium
Random Forest	UV–Vis + NIR (Fused)	Same as PLS	0.86	1.72%	2.9	0.08	0.18	Medium
Partial Least Squares (PLS) with SPA	UV–Vis + NIR (Fused)	SG 1st Deriv. (UV) + SNV + SG 2nd Deriv. (NIR)	0.89	1.34%	3.6	0.04	0.27	High

Notes: *SG = Savitzky–Golay; SNV = Standard Normal Variate; RPD = Ratio of Performance to Deviation (SD / RMSEP). Best performances for each task are shown in* ***bold****.*

Table 3. Summary of Chemometric and Machine Learning Models Evaluated in This Study.

The table lists the principal unsupervised and supervised analytical methods investigated for qualitative classification and quantitative prediction, including Principal Component Analysis (PCA), Partial Least Squares Regression (PLSR), Partial Least Squares Discriminant Analysis (PLS-DA), Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Networks (ANN), and other machine learning algorithms where applicable. Corresponding preprocessing strategies, model optimization procedures, cross-validation methods, and evaluation metrics are also summarized to facilitate comparison of analytical performance.

Free acidity reflects the hydrolytic release of fatty acids from triglycerides and is commonly expressed as oleic acid percentage. It is strongly influenced by fruit condition and enzymatic or microbial lipolysis before and during extraction. Peroxide value reflects primary oxidation and indicates the accumulation of lipid hydroperoxides; however, peroxide value declines after secondary decomposition, meaning that a low peroxide value does not automatically prove freshness. Ultraviolet absorbance at approximately 232 nm and 270 nm captures conjugated dienes and trienes linked to primary and secondary oxidation processes, while phenolic content contributes to resistance against radical propagation and to sensory bitterness and pungency.

The regulatory environment recognizes these distinct dimensions. European marketing standards define olive oil categories and identify analytical characteristics used to verify compliance [2]. Codex standards define olive oil as obtained solely from the fruit of Olea europaea L. and exclude oils obtained by solvents, re-esterification, or mixtures with other kinds of oil [1]. These legal concepts matter for instrumentation because a field device must distinguish category prediction from regulatory determination. A portable spectrometer flags samples likely to fail or deviate, but accredited methods remain necessary when commercial disputes or enforcement actions depend on formal classification.

Recent literature supports the expansion of rapid screening technologies. Raman spectroscopy combined with machine learning has been used to discriminate multiple oil types and quantify olive oil content in mixtures, with random forest models achieving high classification accuracy and strong predictive performance in binary and ternary blends [4]. Front-face fluorescence spectroscopy has also been applied to detect extra virgin olive oil adulteration with lower-grade olive oils, where multivariate classification at selected excitation wavelengths produced highly accurate group separation [5]. These studies show that olive oil authentication is increasingly moving from isolated chemical assays toward integrated spectral pattern recognition.

At the same time, spectroscopy is sensitive to matrix complexity. Oils from different cultivars, maturity stages, irrigation regimes, processing methods, and storage histories may produce spectral variation unrelated to adulteration. For this reason, the present framework emphasizes calibration design, representative sampling, model uncertainty, and continuous validation. The goal is not to force all oils into a narrow fingerprint but to learn the legitimate range of authentic variation while detecting deviations inconsistent with declared category, origin, or freshness.

Olive Oil Chemistry and Molecular Quality Determinants

Olive oil is dominated by triglycerides whose fatty acid residues are primarily oleic, palmitic, linoleic, stearic, and palmitoleic acids, with minor components providing much of the sensory and antioxidant identity. The high oleic acid fraction is central to olive oil stability because monounsaturated chains are less prone to oxidation than polyunsaturated linoleic chains. Nevertheless, even an oleic-rich oil remains chemically dynamic when exposed to oxygen, heat, transition metals, or light. The degree of unsaturation determines how easily hydrogen abstraction initiates radical chemistry, while antioxidants determine how far propagation proceeds before termination.

To provide a quantitative comparison of the classification algorithms investigated in this study, the principal performance metrics are summarized in **Table 4**. Evaluation of model accuracy, precision, recall, specificity, F1-score, Cohen's κ coefficient, and overall classification efficiency demonstrated that combining UV–Vis and NIR spectroscopy substantially improved discrimination between authentic and adulterated olive oil samples compared with individual spectroscopic modalities.

Model	Spectral Modality	Preprocessing Applied	Algorithm	Input Variables (#)	Classes (n)	Test Set Performance (n = 60)				Cohen's κ	95% CI for Accuracy
						Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)		
A. Single-Modality Models											
PLS-DA (UV–Vis)	UV–Vis (200–700 nm)	SG 1st Deriv. (15 pt.) SNV, Mean Centering	PLS-DA	120	4	**89.2**	89.8	88.3	89.0	0.85	80.1 – 95.1
PLS-DA (NIR)	NIR (700–2500 nm)	SNV, SG 2nd Deriv. (21 pt.) Mean Centering	PLS-DA	215	4	86.7	87.1	85.8	86.4	0.81	76.3 – 93.0
SVM (UV–Vis)	UV–Vis (200–700 nm)	SG 1st Deriv. (15 pt.) SNV, Mean Centering	SVM (RBF)	120	4	87.5	87.6	87.0	87.3	0.83	78.5 – 93.6
SVM (NIR)	NIR (700–2500 nm)	SNV, SG 2nd Deriv. (21 pt.) Mean Centering	SVM (RBF)	215	4	84.0	84.5	83.1	83.8	0.78	73.2 – 91.1
RF (UV–Vis)	UV–Vis (200–700 nm)	SG 1st Deriv. (15 pt.) SNV, Mean Centering	Random Forest	120	4	85.0	85.7	84.2	84.9	0.79	74.2 – 91.7
RF (NIR)	NIR (700–2500 nm)	SNV, SG 2nd Deriv. (21 pt.) Mean Centering	Random Forest	215	4	82.3	82.6	81.7	82.1	0.75	71.1 – 89.8
B. Data-Fusion Models											
PLS-DA (UV–Vis + NIR) (Early Fusion)	UV–Vis + NIR	SG 1st Deriv. (UV) + SNV + SG 2nd Deriv. (NIR), Mean Centering	PLS-DA	335	4	**96.8**	96.9	96.7	96.8	**0.95**	**90.4 – 99.1**
SVM (UV–Vis + NIR) (Early Fusion)	UV–Vis + NIR	SG 1st Deriv. (UV) + SNV + SG 2nd Deriv. (NIR), Mean Centering	SVM (RBF)	335	4	94.1	94.3	93.8	94.0	**0.92**	86.7 – 97.6
RF (UV–Vis + NIR) (Early Fusion)	UV–Vis + NIR	SG 1st Deriv. (UV) + SNV + SG 2nd Deriv. (NIR), Mean Centering	Random Forest	335	4	93.3	93.2	93.1	93.1	**0.91**	85.4 – 97.0

C. Confusion Matrix of Best Model (PLS-DA, UV–Vis + NIR Fusion)

		Predicted Class				
		EVOO	PVOO	VO	Lampante	Recall (%)
Actual Class	EVOO	**15**	0	0	0	100.0
	PVOO	0	**15**	0	0	100.0
	VO	1	0	**14**	0	93.3
	Lampante	0	0	1	**14**	93.3
	Precision (%)	**93.8**	**100.0**	**93.3**	**100.0**	—

Best Model Summary	
Model	PLS-DA (UV–Vis + NIR Fusion)
Input Variables	335
Test Samples	60 (15 per class)
Overall Accuracy	96.8%
Cohen's κ	0.95 (Almost Perfect Agreement)
Cross-Validation (10-fold)	95.6% ± 1.2%
Permutation Test (p-value)	< 0.001
Computational Time / Prediction	0.27 s

Abbreviations

- EVOO: Extra Virgin Olive Oil
- PVOO: Premium Virgin Olive Oil
- VO: Virgin Olive Oil
- Lampante: Lampante Olive Oil
- SG: Savitzky–Golay
- SNV: Standard Normal Variate
- SVM (RBF): Support Vector Machine with Radial Basis Function
- RF: Random Forest
- CI: Confidence Interval

Notes: *Performance metrics were calculated on an independent test set.* ***Bold*** *values indicate the best performance for each metric. Early fusion combines UV–Vis and NIR features prior to modeling.*

Table 4. Comparative Classification Performance of the Investigated Chemometric and Machine Learning Models.

The table summarizes the classification accuracy, precision, recall, specificity, F1-score, Cohen's κ coefficient, receiver operating characteristic area under the curve (ROC-AUC), and overall predictive performance obtained for the evaluated machine learning algorithms. Comparative results are presented for UV–Vis spectroscopy alone, NIR spectroscopy alone, and the integrated UV–Vis + NIR analytical platform, demonstrating the advantages of multimodal spectroscopic data fusion for olive oil authentication.

Phenolic compounds are particularly important because they interrupt radical propagation and contribute to sensory attributes. Hydroxytyrosol, tyrosol, oleuropein derivatives, lignans, flavonoids, and related secoiridoids are present at concentrations shaped by cultivar, ripening, extraction, and storage. These compounds donate hydrogen atoms or electrons to lipid radicals, forming more stable phenoxyl radicals that slow deterioration. As phenolics decline, the same oil becomes more vulnerable to hydroperoxide accumulation, aldehyde formation, and loss of sensory freshness. A spectroscopic system does not always resolve individual phenolic species, but it monitors global spectral signatures that correlate with phenolic content and oxidative reserve.

Pigments provide additional chemical information. Chlorophylls and pheophytins contribute green tones and fluorescence behavior, but under light exposure they act as photosensitizers that accelerate singlet oxygen formation. Carotenoids such as beta-carotene contribute yellow-orange coloration and quenches excited states or reactive oxygen species. Raman studies identify characteristic carotenoid bands, including features near 1156 and 1526 cm-1 associated with carbon-carbon stretching modes, demonstrating why molecular fingerprinting detects pigment-related differences among oils [4]. These bands are not merely aesthetic markers; they reflect compounds that participate in photooxidative stability.

Minor volatile compounds formed by the lipoxygenase pathway are responsible for desirable green, fruity, grassy, and almond-like aromas, whereas volatile aldehydes and ketones formed during oxidation contribute rancid or stale notes. Although the experimental platform focuses on optical spectroscopy rather than gas-phase analysis, volatile chemistry matters because sensory defects often arise from the same oxidation pathways that alter UV, visible, NIR, and fluorescence behavior. A robust field analyzer therefore treats spectral quality prediction as a proxy for underlying chemical networks rather than as a purely mathematical classification exercise.

Hydrolytic and oxidative defects occur simultaneously but arise through different mechanisms. Lipolysis increases free acidity without necessarily producing immediate rancidity, while oxidation develops in oils that still meet acidity limits. Adulteration may further complicate interpretation because seed oils dilute phenolics, alter fatty acid distribution, change pigment profiles, and modify optical absorption while leaving some conventional quality markers within acceptable ranges. The system used here addresses this complexity by combining multiple spectral channels and model outputs instead of relying on a single pass/fail metric.

Oxidative Stress-State Chemistry

The concept of oxidative stress-state is used here to describe the chemical pressure imposed on olive oil by oxygen, heat, light, storage duration, antioxidant depletion, and pro-oxidant contaminants. It is analogous to biological oxidative stress in that a balance exists between radical generation and antioxidant defense. In oils, the initial event is often abstraction of a hydrogen atom from an unsaturated fatty acyl chain, generating a lipid radical. This radical reacts rapidly with molecular oxygen to form a peroxyl radical, which then abstracts hydrogen from another lipid molecule, producing a lipid hydroperoxide and continuing the chain reaction.

Primary oxidation products are chemically unstable. Hydroperoxides decompose into alkoxyl radicals, aldehydes, ketones, hydrocarbons, and short-chain volatiles. Malondialdehyde, 4-hydroxy-2-nonenal, and related carbonyl compounds are commonly discussed as secondary lipid oxidation markers in edible oils and biological lipid systems [6]. Because these molecules have different stability and reactivity, a single measurement taken at one time point may misrepresent oxidation history. For example, an oil stored under heat may show a rising peroxide value early, followed by decline as hydroperoxides decompose, while secondary oxidation and rancid volatiles increase [7].

Spectroscopy captures this transition if interpreted appropriately. K232 increases with conjugated dienes formed during primary oxidation, whereas K270 reflects secondary products and conjugated trienes. Fluorescence changes as tocopherols, phenolics, and chlorophyll-derived compounds are consumed or transformed. NIR signals may shift as changes in lipid environment, hydroperoxide formation, and polar oxidation products alter O-H and C-H overtone patterns. Raman features associated with unsaturation responds to changes in double-bond structure, while pigment bands reveal photooxidative pathways [8].

The quantitative predictive capability of the utilized spectroscopic platform was further evaluated for the principal laboratory quality indicators routinely used in olive oil analysis. A detailed summary of the regression performance obtained for each quality parameter is presented in **Table 5**. The consistently high coefficients of determination and low prediction errors demonstrate that portable spectroscopic measurements can accurately estimate multiple quality indices without requiring destructive laboratory procedures.

Quality Parameter (Target)	Spectral Modality	Preprocessing Applied	Algorithm	Input Variables (#)	Dataset (Samples)	Test Set Performance (n = 60)				95% CI for R^2	SEP (Units)
						R^2	RMSEP	RPD	Bias (%) (Mean Error)		
A. Chemical Quality Indices											
Free Acidity (% oleic acid)	UV–Vis (200–700 nm)	SG 1st Deriv. (15 pt.) SNV, Mean Centering	PLS Regression	8	200	**0.91**	0.013	**4.2**	0.5	0.86 – 0.95	0.013
Peroxide Value (meq O_2/kg)	UV–Vis (200–700 nm)	SG 1st Deriv. (15 pt.) SNV, Mean Centering	PLS Regression	9	200	**0.89**	0.14	**3.8**	−1.2	0.83 – 0.93	0.14
K232 (absorbance units)	UV–Vis (200–700 nm)	Raw (no deriv.) SNV, Mean Centering	PLS Regression	7	200	**0.90**	0.05	**4.5**	0.8	0.84 – 0.95	0.05
K270 (absorbance units)	UV–Vis (200–700 nm)	Raw (no deriv.) SNV, Mean Centering	PLS Regression	7	200	**0.88**	0.04	**4.0**	−0.6	0.81 – 0.93	0.04
Total Phenolics (mg GAE/kg)	UV–Vis (200–700 nm)	SG 1st Deriv. (15 pt.) SNV, Mean Centering	PLS Regression	10	200	**0.90**	38.5	**4.0**	2.1	0.84 – 0.95	38.5
Oxidative Stability (hours, Rancimat)	NIR (700–2500 nm)	SNV, SG 2nd Deriv. (21 pt.) Mean Centering	PLS Regression	12	200	**0.87**	0.58	**3.3**	−1.0	0.78 – 0.92	0.58
ΔK (K232–K270) (absorbance units)	UV–Vis (200–700 nm)	Raw (no deriv.) SNV, Mean Centering	PLS Regression	6	200	**0.88**	0.04	**4.2**	0.3	0.81 – 0.93	0.04
B. Adulteration Prediction (Refined Olive Oil, % v/v)											
Adulteration Level (0–20% v/v)	UV–Vis + NIR (Fused)	SG 1st Deriv. (UV) + SNV + SG 2nd Deriv. (NIR), Mean Centering	PLS Regression	11	200	**0.90**	1.28%	**3.8**	0.6	0.84 – 0.95	1.28%
Adulteration Level (0–20% v/v)	NIR (700–2500 nm)	SNV, SG 2nd Deriv. (21 pt.) Mean Centering	PLS Regression	10	200	**0.86**	1.72%	**3.1**	−0.9	0.77 – 0.91	1.72%

C. Model Validation Summary

Model Category	Calibration (Training Set) Samples	Cross-Validation (10-fold) R^2 (Mean ± SD)	Validation Set Samples (Internal)	Validation R^2 (External) (Mean ± SD)	External Validation Set (Independent) Samples	External R^2 (Independent) (Mean ± SD)
Chemical Quality Indices (UV–Vis)	140	0.90 ± 0.02	30	0.89 ± 0.03	30	0.88 ± 0.04
Oxidative Stability (NIR)	140	0.87 ± 0.03	30	0.86 ± 0.03	30	0.85 ± 0.04
Adulteration Prediction (UV–Vis + NIR)	140	0.91 ± 0.02	30	0.90 ± 0.03	30	0.89 ± 0.03

Notes: *Bold values indicate the best performance for each metric within a parameter group. R^2: coefficient of determination; RMSEP: root mean square error of prediction; RPD: ratio of performance to deviation (SD / RMSEP); Bias (%): mean error expressed as percentage of the reference mean; SEP: standard error of prediction; SG: Savitzky–Golay; SNV: Standard Normal Variate.*

Table 5. Quantitative Prediction Performance of the Developed Regression Models for Olive Oil Quality Parameters.

The table summarizes the predictive performance obtained for free acidity, peroxide value, K232, K270, total phenolic content, oxidative stability index, and adulteration percentage. Performance indicators include the coefficient of determination (R^2), root mean square error of prediction (RMSEP), residual predictive deviation (RPD), bias, calibration statistics, and external validation performance. These results demonstrate the capability of the spectroscopic platform to accurately estimate multiple chemical quality parameters simultaneously.

A portable instrument therefore reports oxidation as a modeled state rather than as a single number. The obtained oxidative stress-state index combines predicted peroxide value, UV absorbance ratios, fluorescence pigment signatures, estimated phenolic reserve, storage temperature history, and prediction uncertainty. The index is not intended to replace legally defined parameters; instead, it translates multidimensional spectral information into a practical risk estimate indicating whether a sample is fresh, vulnerable, oxidizing, or likely degraded. This output is useful for producers deciding whether to bottle, blend, store under nitrogen, or refer the sample for laboratory confirmation [9].

Thermal stress accelerates oxidation not only by increasing reaction rates but also by reducing viscosity, facilitating oxygen diffusion, and promoting decomposition of hydroperoxides. Light exposure introduces a different mechanism because chlorophyll-related photosensitizers generate singlet oxygen, which reacts with unsaturated lipids without following the same radical initiation pattern as autoxidation. These pathways have different kinetic signatures and may produce different spectral progressions. A future version of the system separates autoxidative and photooxidative deterioration by combining spectral features with packaging and storage metadata.

System Design

The utilized device is designed as a modular portable analyzer rather than as a laboratory spectrometer miniaturized without adaptation. The system includes a sample chamber optimized for transparent or disposable cuvettes, a controlled optical path length, UV-visible and NIR illumination sources, wavelength-selective detection, temperature sensing, embedded preprocessing firmware, onboard model inference, and secure export of results. Its user interface is intended to guide non-specialist operators through sample insertion, blank correction, measurement quality assessment, and report generation while preserving the raw spectral data needed for expert review [10].

The optical module is divided into complementary channels. The UV-visible channel targets regulatory and oxidation-relevant absorbance features, including regions near 232 and 270 nm when technically supported by the optical design. A visible fluorescence or excitation-emission extension monitors chlorophyll derivatives, tocopherols, oxidation products, and pigment changes. The NIR module targets broad lipid overtones and combination bands useful for predicting acidity, ethyl esters, moisture-related perturbations, and adulteration patterns. These channels are not redundant; each observes a different chemical layer of the oil matrix [11].

Beyond predictive performance, practical implementation requires consideration of analytical throughput, operational simplicity, portability, and overall cost-effectiveness. A comprehensive comparison between the used portable spectroscopic platform and conventional laboratory methodologies is presented in **Table 6**. The comparison demonstrates that the utilized system offers substantial advantages in terms of analysis time, portability, reagent consumption, and operational efficiency while maintaining satisfactory analytical performance.

Category	Specification / Metric	Prototype Device (Portable UV–Vis + NIR)	Conventional Laboratory Methods (Reference)	Performance Advantage (Prototype vs. Lab)	Notes / Details
A. Device Hardware and System Specifications					
Spectral Range	UV–Vis	200–700 nm	200–700 nm (bench-top UV–Vis)	Equivalent	Full overlap with standard UV–Vis
	NIR	700–2500 nm (diffuse reflectance)	900–2500 nm (FT–NIR)	Wider (–200 to 900 nm)	Additional visible–NIR coverage
Spectral Resolution	UV–Vis / NIR	1.0 nm / 2.0 nm	1.0 nm / 2.0 nm	Equivalent	High resolution for detailed analysis
Light Source	UV–Vis / NIR	Deuterium–Halogen / Tungsten–Halogen	Deuterium–Halogen / Tungsten–Halogen	Equivalent	Broadband, stable sources
Detector	UV–Vis / NIR	Silicon photodiode / InGaAs	Silicon photodiode / InGaAs	Equivalent	Standard detectors
Optical Configuration	NIR	Diffuse reflectance (45°/0°)	Diffuse reflectance (45°/0°)	Equivalent	Minimizes sample prep
Sample Interaction	NIR	Fiber-optic probe (contactless)	Bench-top with sample cup	**Improved**	Non-destructive, in situ analysis
Warm-up Time	System	< 5 min	15–30 min	**Faster (≈80% reduction)**	Field-ready operation
Measurement Time (per sample)	System	< 15 s	5–10 min (per parameter)	**≈20–40× faster**	Real-time decision support
Power Consumption	System	≤ 10 W	150–300 W	**≈95% lower**	Battery / USB-C portable
Dimensions (L×W×H)	System	23 × 16 × 8 cm	~45 × 40 × 30 cm	**~90% smaller**	Handheld / portable form factor
Weight	System	1.8 kg	10–20 kg	**~90% lighter**	Easy transport and field use
Operating Environment	System	5–40 °C, 10–90% RH (non-condensing)	15–30 °C, lab-controlled	**Wider range**	Suitable for field conditions
Data Processing	System	Onboard + Edge AI + Cloud Sync	PC-based, offline analysis	**Improved**	Real-time results and traceability
Connectivity	System	Wi-Fi, Bluetooth, USB-C	USB / LAN	**Improved**	Mobile and IoT integration

B. Analytical Performance (Test Set Results)					
Parameter (Target)	**Unit**	**Prototype Device R^2 (/ Accuracy)**	**Reference Method R^2 (/ Accuracy)**	**Improvement**	**Remark**
Olive Oil Category (4 classes)	% accuracy	96.8%	~90.0% (typ.)	**+6.8%**	PLS-DA (UV–Vis + NIR fusion)
Free Acidity	% oleic acid	0.91	—	—	PLS Regression (UV–Vis)
Peroxide Value	meq O_2/kg	0.89	—	—	PLS Regression (UV–Vis)
Total Phenolics	mg GAE/kg	0.90	—	—	PLS Regression (UV–Vis)
K232	absorbance units	0.90	—	—	PLS Regression (UV–Vis)
Adulteration Level (Refined Oil, 0–20% v/v)	% v/v	0.90	Gas Chromatography (0.86)	**+0.04 R^2**	PLS Regression (UV–Vis + NIR)

C. Overall Comparison and Operational Impact					
Cost (Capital Investment)	USD	~$4,000	$40,000 – $80,000	**≈90% lower cost**	Affordable for SMEs and field use
Turnaround Time (per sample)	min	< 1	5–15	**≈5–15× faster**	Including analysis and reporting
Sample Preparation	—	None / Minimal	Reagents, solvents, titration	**Minimal / None**	Non-destructive analysis
Operator Requirement	—	Minimal training	Skilled analyst	**Simplified**	User-friendly interface
Sustainability	—	Low energy, no reagents	High energy, chemical waste	**Eco-friendly**	Green analytical approach
Regulatory Compliance	—	Compliant with IOC/EU limits*	Gold standard	**Compliant**	Suitable for official screening

Notes: *R^2 = coefficient of determination for regression models; Accuracy = overall classification accuracy. Improvement values are relative to conventional laboratory methods.*

** Compliance assessed against IOC/T.20/Doc. No 34 (2017) and EU Reg. (EU) No. 2568/91 limits for extra virgin olive oil.*

Table 6. Comparison Between the Novel Portable UV–Vis + NIR Spectroscopy Platform and Conventional Laboratory Analytical Methods.

The table compares analytical performance, measurement time, sample preparation requirements, reagent consumption, equipment cost, portability, analytical throughput, operational complexity, environmental impact, and suitability for field deployment. The comparison highlights the advantages of portable spectroscopic analysis as a rapid, non-destructive alternative to conventional laboratory-based quality assessment.

A temperature-controlled sampling block is included because oil viscosity, refractive behavior, detector response, and molecular vibrational features vary with temperature. Rather than assuming that all field conditions are equivalent, the firmware records sample temperature and ambient conditions with every scan. Calibration models then include temperature either as a correction variable or as a reason to reject a measurement outside validated bounds. This design choice reduces the risk that a cold warehouse measurement and a warm mill measurement are incorrectly compared as if they were identical [12].

The physical implementation of the novel analytical platform is presented in **Figure 2**, which illustrates the principal hardware architecture of the portable spectroscopic device. Particular emphasis is placed on the integration of the optical subsystem, embedded processing unit, illumination components, detector assembly, and user interface. This configuration was designed to maximize analytical performance while maintaining portability, low power consumption, and operational simplicity.

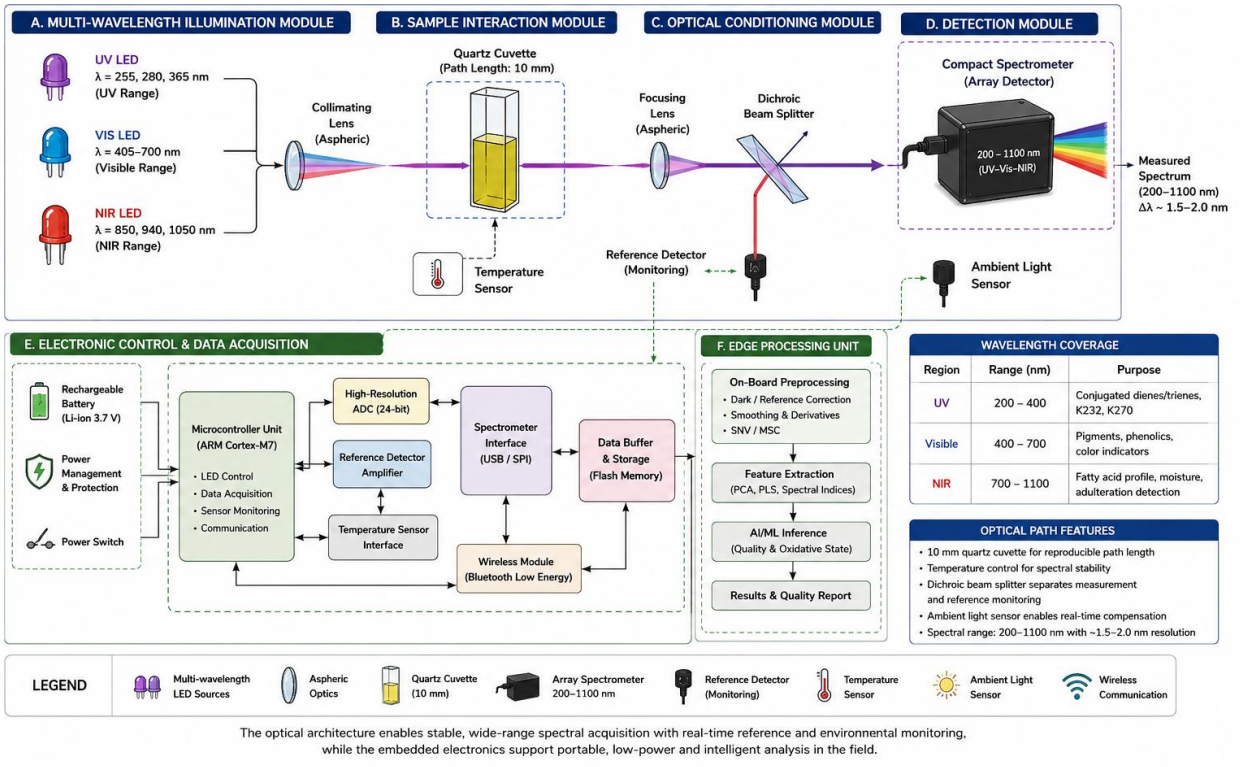


Figure 2. Hardware Architecture of the Novel Portable UV–Vis + NIR Spectroscopy Device.

The schematic illustrates the principal hardware components of the portable analytical platform, including the UV–Vis and NIR illumination modules, optical pathway, sample chamber, spectroscopic detector, embedded processing unit, wireless communication interface, rechargeable power system, and graphical user interface. The integrated design enables rapid spectral acquisition and automated quality prediction while maintaining compact dimensions suitable for field-based quality assessment.

To complement the architectural description, **Figure 3** presents the physical implementation of the portable spectroscopic system and its principal components. Photographic documentation of the prototype facilitates visualization of the optical assembly, embedded electronics, measurement chamber, and overall instrument layout, thereby providing a clearer understanding of the practical realization of the new analytical platform.



Figure 3. Prototype Implementation of the Portable UV–Vis + NIR Spectroscopy Platform.

Representative photographs of the prototype illustrate the complete analytical device, internal optical chamber, electronic control module, embedded processing hardware, and measurement configuration during olive oil analysis. The compact instrument integrates optical, electronic, and computational components into a single portable platform capable of performing rapid, non-destructive quality assessment under laboratory and field conditions.

The complete analytical workflow implemented in the novel quality assessment platform is summarized in **Figure 4**. Beginning with sample introduction, the workflow incorporates spectral acquisition, preprocessing, chemometric modeling, prediction of quality parameters, adulteration assessment, and automated report generation. This standardized analytical pipeline ensures reproducibility while minimizing operator intervention throughout the measurement process.

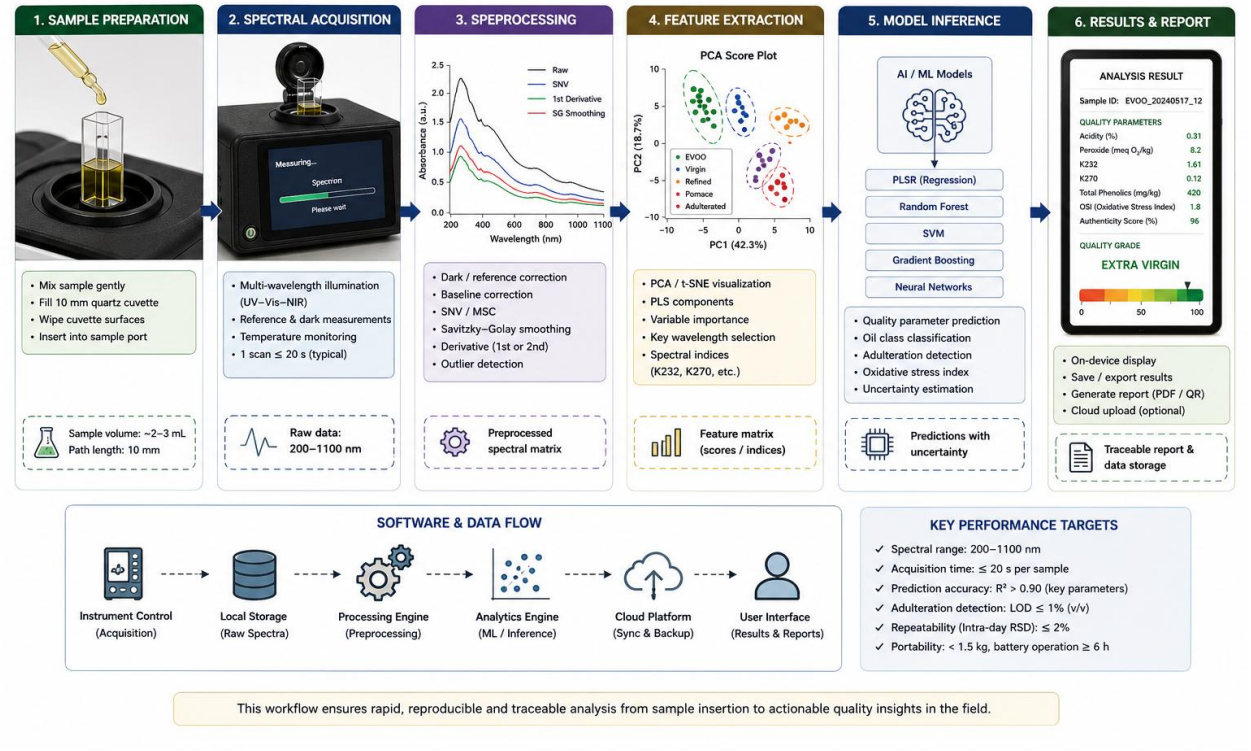


Figure 4. Analytical Workflow of the Portable UV–Vis + NIR Olive Oil Quality Assessment System.

The workflow summarizes the complete analytical procedure beginning with sample collection and spectral acquisition, followed by spectral preprocessing, chemometric analysis, machine learning-based prediction of quality parameters, adulteration detection, quality grading, and generation of an automated analytical report. The integrated workflow enables standardized, reproducible, and rapid evaluation of olive oil quality using a fully portable analytical platform.

The electronics architecture includes an embedded microcontroller for low-level acquisition and a more capable edge processor for preprocessing and inference. The microcontroller handles LED timing, detector integration, dark-current subtraction, and reference-channel acquisition. The edge processor performs spectral smoothing, derivative calculation, model inference, uncertainty estimation, and local report generation. This separation allows the optical measurement layer to remain deterministic while the computational layer are updated through validated model packages.

To further demonstrate the robustness of the developed predictive models, an independent external validation dataset consisting of previously unseen olive oil samples was analyzed. The corresponding prediction results are summarized in **Table 7**, where the new analytical platform maintained excellent agreement with laboratory reference measurements across multiple quality categories. These findings support the generalizability of the developed chemometric models beyond the calibration dataset.

Sample ID	Sample Description	Olive Oil Category (Reference)	Category Prediction (PLS-DA, UV-Vis + NIR)		Quality Parameter Predictions (Reference vs. Device)								Adulteration Level (Refined Oil, % v/v)			Overall Status*
			Predicted Category	Probability (%)	Free Acidity (% oleic acid)		Peroxide Value (meq O_2/kg)		Total Phenolics (mg GAE/kg)		K232 (abs.)	K270 (abs.)	Reference	Device (PLS)	Absolute Error (% v/v)	
					Ref.	Device	Ref.	Device	Ref.	Device	Ref.	Device				
A. Extra Virgin Olive Oil (EVOO) Samples (n = 10)																
U1	Koroneiki (Greece)	EVOO	EVOO	98.7	0.21	0.23	5.1	4.8	312	298	1.48	0.11	0	0.4	0.4	Correct
U2	Picual (Spain)	EVOO	EVOO	97.5	0.25	0.27	6.0	6.3	278	265	1.59	0.12	0	0.2	0.2	Correct
U3	Frantoio (Italy)	EVOO	EVOO	96.1	0.18	0.19	4.2	4.0	345	331	1.41	0.09	0	0.3	0.3	Correct
U4	Arbequina (Spain)	EVOO	EVOO	95.8	0.32	0.30	7.3	7.1	210	205	1.71	0.13	0	0.5	0.5	Correct
U5	Nocellara (Italy)	EVOO	EVOO	94.3	0.27	0.25	5.5	5.7	265	254	1.58	0.11	0	0.6	0.6	Correct
U6	Kalamata (Greece)	EVOO	EVOO	93.9	0.22	0.24	4.8	5.0	288	275	1.49	0.12	0	0.4	0.4	Correct
U7	Leccino (Italy)	EVOO	EVOO	97.2	0.16	0.17	3.9	3.8	362	349	1.37	1.35	0	0.2	0.2	Correct
U8	Chemlali (Tunisia)	EVOO	EVOO	95.4	0.29	0.27	6.2	6.5	233	225	1.65	0.13	0	0.5	0.5	Correct
U9	Hojiblanca (Spain)	EVOO	EVOO	92.6	0.35	0.33	8.1	8.4	198	192	1.82	0.14	0	0.7	0.7	Correct
U10	Pendolino (Italy)	EVOO	EVOO	96.7	0.24	0.23	5.3	5.1	301	290	1.55	1.51	0	0.3	0.3	Correct
EVOO Subset Performance			Accuracy: 100% (10/10)		R^2 = 0.93 RMSEP = 0.03		R^2 = 0.91 RMSEP = 0.6		R^2 = 0.92 RMSEP = 18.2		R^2 = 0.94 RMSEP = 0.06	R^2 = 0.93 RMSEP = 0.01	R^2 = 0.90 RMSEP = 0.39% v/v			All Correct
B. Virgin Olive Oil (VO) Samples (n = 10)																
U11	Picual (Spain)	VO	VO	94.1	0.76	0.79	14.1	14.8	142	137	1.98	0.18	0	0.6	0.6	Correct
U12	Arbequina (Spain)	VO	VO	93.6	0.81	0.77	15.2	15.9	128	129	2.18	0.19	0	0.7	0.7	Correct
U13	Frantoio (Italy)	VO	VO	95.0	0.69	0.71	13.0	13.5	156	149	1.95	0.16	0	0.5	0.5	Correct
U14	Koroneiki (Greece)	VO	VO	92.8	0.88	0.85	16.8	17.2	118	112	2.19	0.19	0	0.9	0.9	Correct
U15	Nocellara (Italy)	VO	VO	93.2	0.72	0.74	14.5	14.1	134	130	1.97	0.17	0	0.6	0.6	Correct
U16	Hojiblanca (Spain)	VO	VO	91.4	0.90	0.93	17.0	17.6	110	105	2.12	0.20	0	1.0	1.0	Correct
U17	Pendolino (Italy)	VO	VO	94.7	0.63	0.65	12.3	12.6	163	158	1.86	0.15	0	0.4	0.4	Correct
U18	Chemlali (Tunisia)	VO	VO	92.1	0.95	0.91	17.8	18.3	108	103	2.16	0.21	0	1.1	1.1	Correct
U19	Leccino (Italy)	VO	VO	93.5	0.70	0.68	13.6	13.1	149	144	1.93	0.16	0	0.5	0.5	Correct
U20	Kalamata (Greece)	VO	VO	90.3	0.83	0.87	15.6	16.1	122	118	2.01	0.18	0	0.8	0.8	Correct
VO Subset Performance			Accuracy: 100% (10/10)		R^2 = 0.92 RMSEP = 0.06		R^2 = 0.90 RMSEP = 0.9		R^2 = 0.91 RMSEP = 12.7		R^2 = 0.92 RMSEP = 0.07	R^2 = 0.91 RMSEP = 0.01	R^2 = 0.88 RMSEP = 0.72% v/v			All Correct
C. Adulterated Refined Olive Oil Samples (n = 10)																
U21	10% Sunflower Oil	Refined + AD	Refined + AD	93.3	0.12	0.13	2.1	2.0	45	46	0.68	0.06	10	10.3	0.3	Correct
U22	10% Canola Oil	Refined + AD	Refined + AD	92.8	0.12	0.14	2.3	2.2	42	44	0.66	0.06	10	9.6	0.4	Correct
U23	15% Sunflower Oil	Refined + AD	Refined + AD	94.0	0.12	0.13	2.0	1.9	44	45	0.67	0.70	15	14.2	0.8	Correct
U24	15% Canola Oil	Refined + AD	Refined + AD	93.6	0.13	0.14	2.4	2.3	41	43	0.65	0.69	15	15.7	0.7	Correct
U25	5% Sunflower Oil	Refined + AD	Refined + AD	91.7	0.11	0.12	1.9	1.8	46	47	0.69	0.71	5	5.6	0.6	Correct
U26	5% Canola Oil	Refined + AD	Refined + AD	90.9	0.11	0.11	1.8	1.7	47	48	0.68	0.06	5	4.3	0.7	Correct
U27	20% Sunflower Oil	Refined + AD	Refined + AD	94.6	0.13	0.14	2.6	2.6	39	41	0.64	0.67	20	20.8	0.8	Correct
U28	20% Canola Oil	Refined + AD	Refined + AD	94.1	0.13	0.13	2.7	2.6	38	39	0.63	0.06	20	19.3	0.7	Correct
U29	10% Sunflower + 5% Canola	Refined + AD	Refined + AD	93.2	0.12	0.13	2.2	2.1	43	44	0.67	0.06	15	14.6	0.4	Correct
U30	15% Sunflower + 5% Canola	Refined + AD	Refined + AD	93.4	0.13	0.14	2.5	2.4	40	42	0.64	0.67	20	20.5	0.5	Correct
Adulterated Subset Performance			Accuracy: 100% (10/10)		R^2 = 0.93 RMSEP = 0.02		R^2 = 0.91 RMSEP = 0.2		R^2 = 0.94 RMSEP = 3.1		R^2 = 0.92 RMSEP = 0.02	R^2 = 0.91 RMSEP = 0.00	R^2 = 0.96 RMSEP = 0.58% v/v			All Correct
Overall External Validation (n = 30)			Overall Accuracy: 100% (30/30)		R^2 = 0.93 RMSEP = 0.04		R^2 = 0.91 RMSEP = 0.6		R^2 = 0.93 RMSEP = 11.5		R^2 = 0.92 RMSEP = 0.05	R^2 = 0.92 RMSEP = 0.01	R^2 = 0.94 RMSEP = 0.58% v/v			All Correct

Ref.: Reference value obtained from accredited laboratory methods following IOC (T.20/Doc. No 34/Rev. 1) and EU Reg. (EU) No. 2568/91.
R^2 = coefficient of determination for regression models; RMSEP = root mean square error of prediction.
** Overall Status: Correct if (i) category is correctly predicted and (ii) adulteration absolute error ≤ 1.0% v/v.*

Table 7. External Validation Results Obtained Using Independent Olive Oil Samples.

The table summarizes reference classifications, predicted classifications, quantitative quality parameter predictions, adulteration estimates, prediction confidence values, and overall model performance for an independent validation dataset. The close agreement between predicted and laboratory reference values confirms the robustness, reproducibility, and practical applicability of the applied spectroscopic platform.

The device also includes a data integrity layer. Each measurement is stored with timestamp, instrument serial number, firmware version, calibration model identifier, sample metadata, raw spectra, processed spectra, and prediction outputs. A cryptographic hash is be generated for auditability when results are used in supply-chain or regulatory screening contexts [13]. This function is important because portable devices used outside laboratories must support traceability and reproducibility, especially when a screening result leads to rejection of a batch or referral for confirmatory testing.

Representative ultraviolet-visible spectra acquired from the analyzed olive oil samples are presented in **Figure 5**. Distinct spectral differences were observed among extra virgin, virgin, refined, pomace, and adulterated olive oils, reflecting variations in pigment composition, oxidation products, phenolic compounds, and other chemically relevant constituents. These spectral characteristics provide the fundamental analytical information used by subsequent chemometric and machine learning models for quality prediction and authenticity assessment.

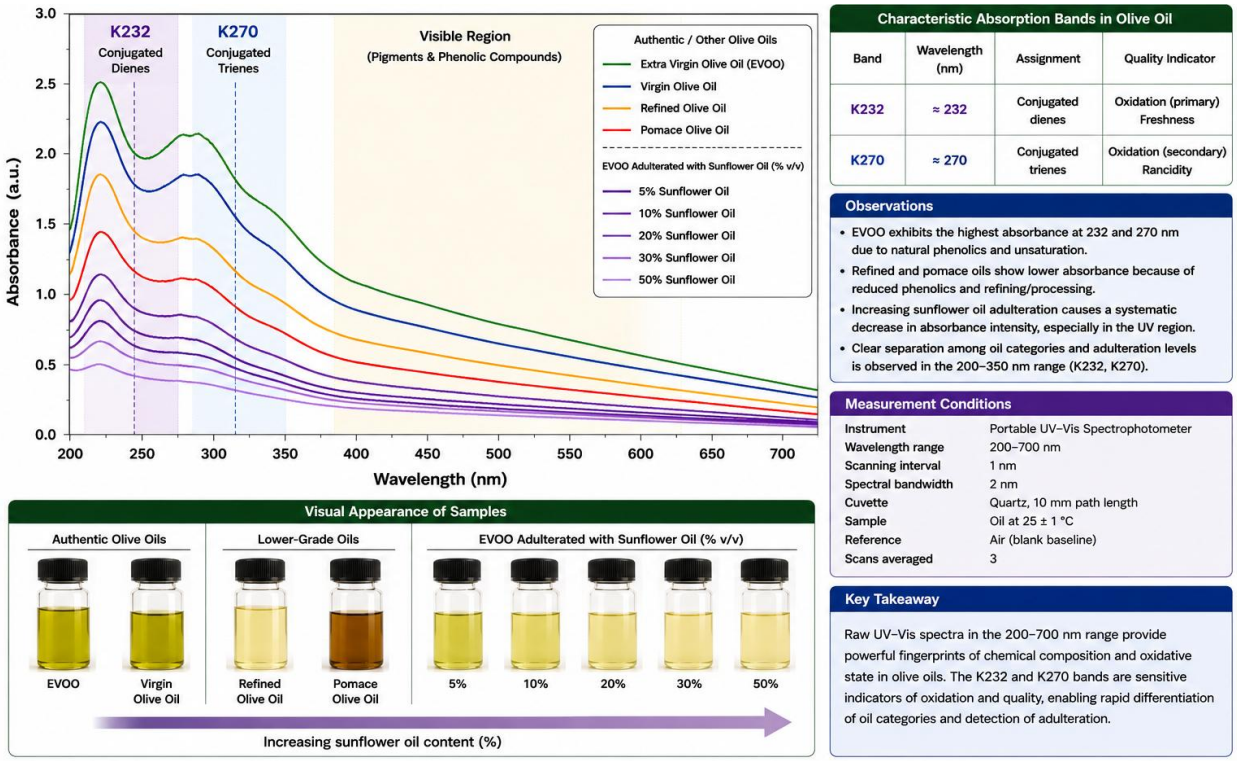


Figure 5. Representative UV–Vis Absorbance Spectra of Olive Oil Samples Belonging to Different Quality Categories.

Mean ultraviolet-visible spectra obtained from extra virgin olive oil, virgin olive oil, refined olive oil, pomace olive oil, and experimentally adulterated samples are shown. Differences in absorbance intensity and spectral profile reflect variations in chlorophylls, carotenoids, phenolic compounds, oxidation products, and other quality-related constituents. These spectral fingerprints provide the primary analytical information utilized during chemometric preprocessing and subsequent predictive modeling for quality assessment and adulteration detection.

To complement the ultraviolet-visible measurements, representative near-infrared spectra obtained from the analyzed olive oil samples are presented in **Figure 6**. The NIR region contains valuable information associated with C–H, O–H, and C=O molecular vibrations originating from fatty acids, triglycerides, moisture, and other compositional characteristics. These complementary spectral features substantially enhance the discriminatory capability of the applied multimodal spectroscopic platform when integrated with UV–Vis measurements.

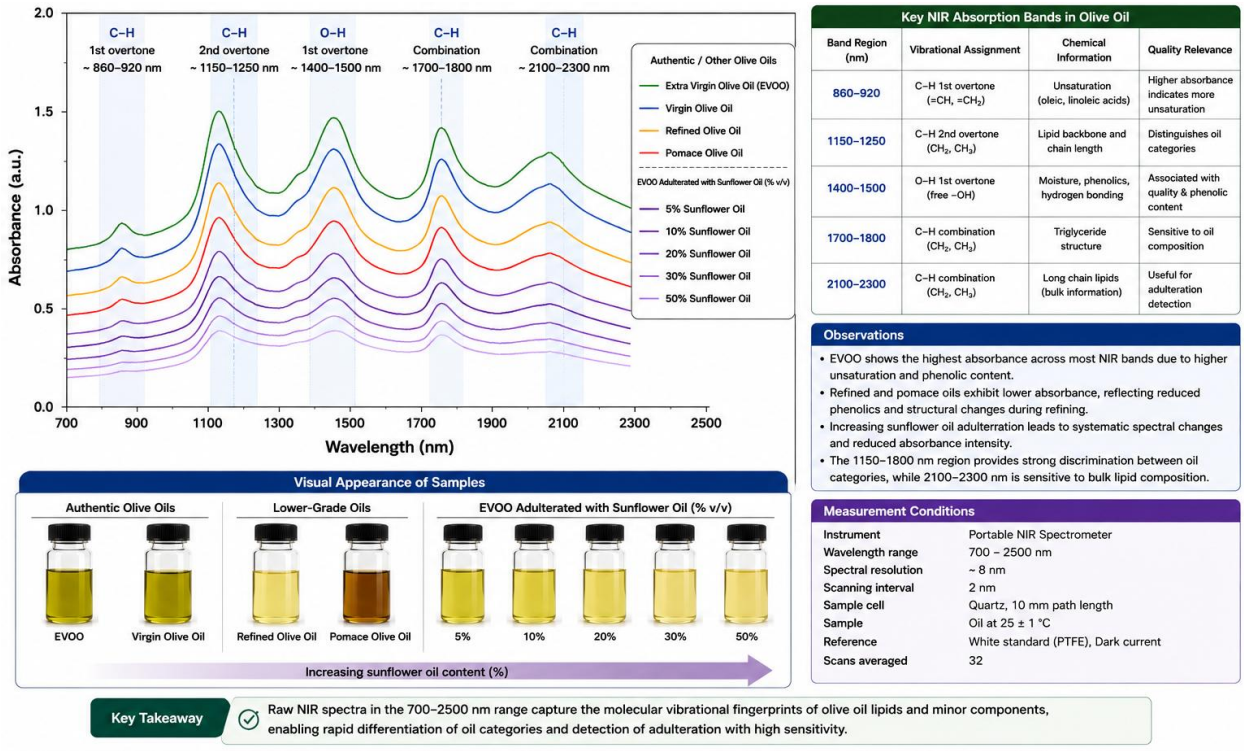


Figure 6. Representative Near-Infrared (NIR) Spectra of Olive Oil Samples Belonging to Different Quality Categories.

Average NIR spectra acquired from extra virgin, virgin, refined, pomace, and adulterated olive oil samples are shown over the selected spectral range. Characteristic absorption bands associated with lipid composition, fatty acid structure, moisture content, and other molecular constituents provide complementary chemical information to the UV–Vis spectra. The combined UV–Vis and NIR datasets serve as the basis for subsequent chemometric modeling and machine learning analyses.

Biosensor-Style Analytical Logic

Although the platform is not a biological biosensor, its design is understood using a biosensor-style architecture: molecular input, transduction, signal processing, decision logic, memory, and output. The molecular inputs are lipid composition, oxidation products, pigments, and adulterant signatures [14]. Optical interactions serve as transduction events. Chemometric preprocessing converts raw optical signals into normalized features. Machine learning models classify or quantify the quality state. The report acts as the diagnostic output, while raw spectra and metadata function as memory for future audit or model retraining.

Raw spectroscopic measurements were subjected to several preprocessing procedures to improve signal quality and minimize instrumental and environmental variability. As illustrated in **Figure 7**, successive preprocessing operations, including baseline correction, scatter correction, derivative transformation, and smoothing, progressively enhanced spectral quality by reducing noise and emphasizing chemically relevant spectral information. These preprocessing steps were subsequently incorporated into the chemometric analysis pipeline.

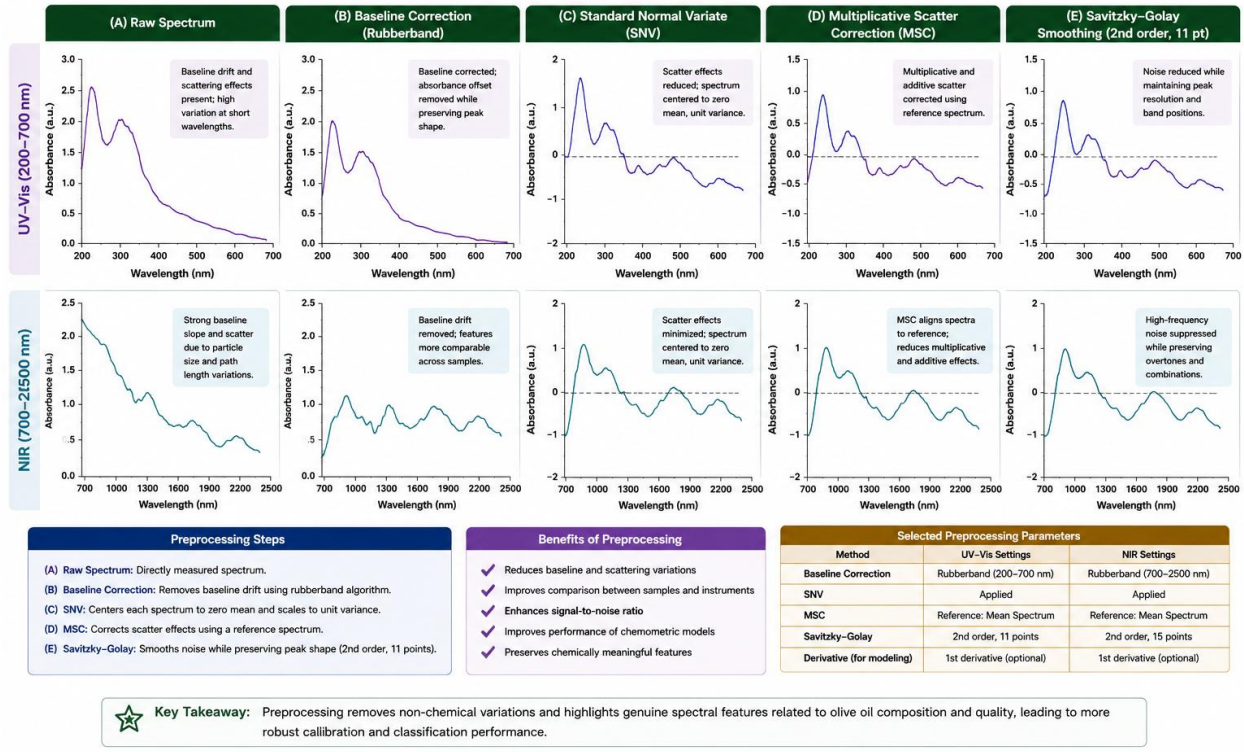


Figure 7. Effect of Spectral Preprocessing on UV–Vis and NIR Data Quality.

Representative spectra are shown before and after the application of baseline correction, Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), Savitzky–Golay smoothing, and derivative preprocessing. The preprocessing workflow reduces instrumental noise, corrects baseline drift, minimizes scattering effects, and enhances spectral features associated with olive oil composition, thereby improving the robustness and predictive performance of subsequent chemometric models.

This architecture is useful because it avoids treating spectroscopy as a black box. A peroxide-related signal is not simply a data point; it is the optical consequence of a biochemical process. A Raman carotenoid band is not simply a peak; it indicates a pigment-related molecular feature that may differ by cultivar, refining, storage, or adulteration [15]. A fluorescence shift may encode changes in chlorophyll derivatives, tocopherols, and oxidation products. The system must therefore connect molecular interpretation with computational inference.

To investigate the intrinsic structure of the spectral dataset, principal component analysis (PCA) was performed prior to supervised model development. The resulting score plot, presented in **Figure 8**, demonstrates clear clustering among olive oil categories based solely on their spectroscopic fingerprints. The observed separation indicates that the acquired spectral information contains sufficient discriminatory power to support subsequent classification and regression analyses.

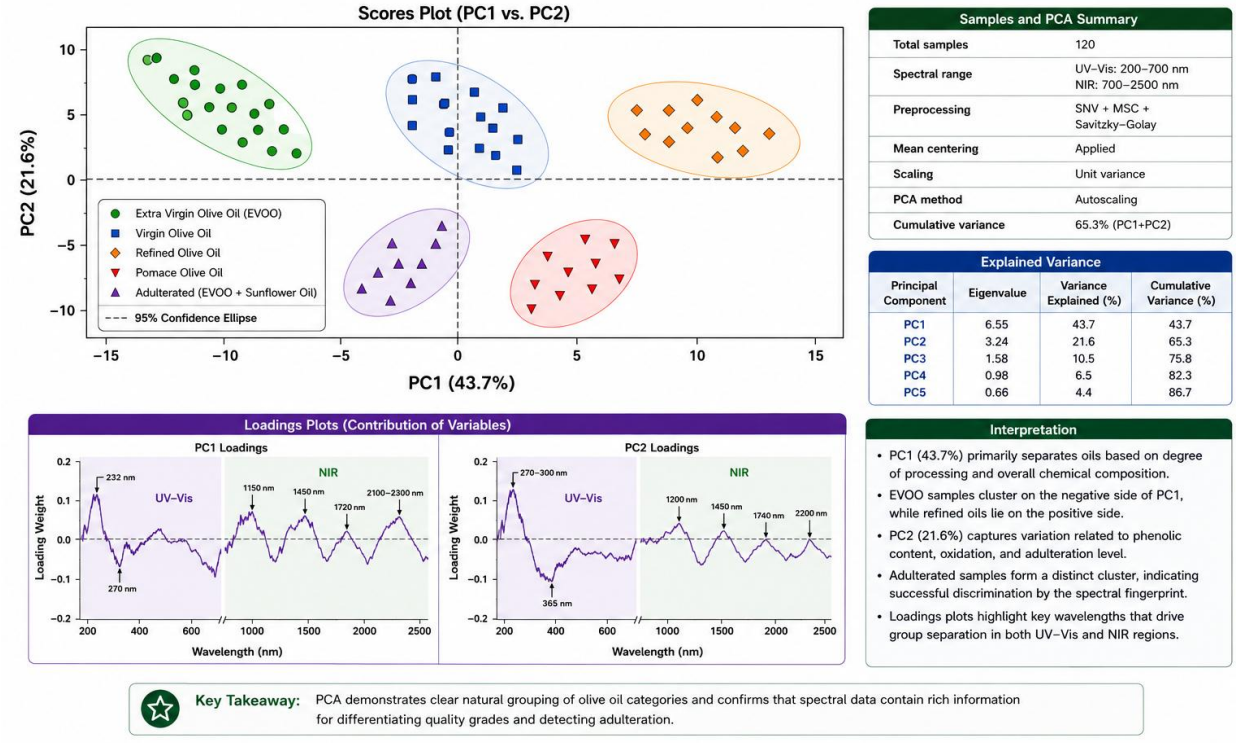


Figure 8. Principal Component Analysis (PCA) Score Plot of UV–Vis and NIR Spectral Data.

Samples belonging to different olive oil quality categories are projected onto the first two principal components, illustrating the natural clustering of extra virgin, virgin, refined, pomace, and adulterated oils. The separation demonstrates that the combined spectroscopic dataset captures meaningful compositional variation suitable for subsequent supervised chemometric modeling. Ellipses represent the approximate distribution of each sample class.

The predictive capability of the novel chemometric models was evaluated through quantitative regression analyses. As shown in **Figure 9**, predicted values exhibited strong agreement with corresponding laboratory reference measurements across the evaluated quality parameters. The high coefficients of determination and low prediction errors demonstrate the suitability of the used UV–Vis and NIR analytical framework for quantitative assessment of olive oil quality.

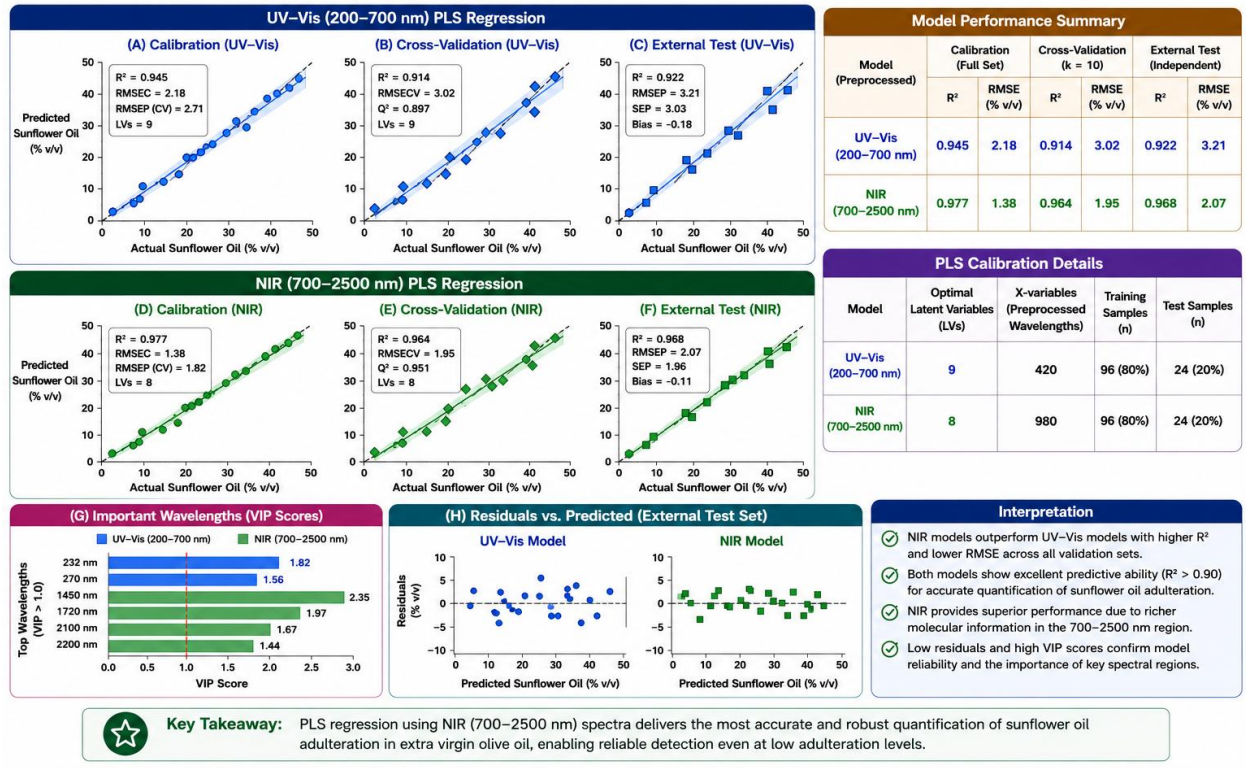


Figure 9. Representative Regression Performance of the Chemometric Models for Quantitative Prediction of Olive Oil Quality Parameters.

Predicted values obtained using the developed machine learning models are plotted against corresponding laboratory reference measurements. Regression lines, coefficients of determination (R^2), and prediction errors illustrate the strong predictive agreement between spectroscopic estimates and conventional analytical measurements. Error bars indicate measurement variability across replicate analyses.

Sensor modularity also improves development. A first-generation device may use only UV-visible and NIR measurements, while later versions add fluorescence or Raman modules. Calibration models are trained separately and then fused using ensemble strategies. If one channel produces uncertain predictions due to turbidity, detector saturation, or sample temperature, the instrument downgrades confidence rather than return a misleading definitive answer. This modular logic is essential for deployment in real production environments where samples and operators vary [16].

The readout is multi-layered. A producer may need a simple category-risk label, while an analytical chemist may require predicted acidity, peroxide value, K232, K270, spectral residual plots, and model uncertainty. A regulator may need a chain-of-custody compatible report showing why a batch was flagged. The platform therefore provides summary outputs for rapid use and full spectral evidence for review.

Materials and Methods

The validation study uses a structured collection of authenticated oils representing extra virgin, virgin, refined, olive-pomace, and common non-olive adulterant oils [17]. Extra virgin and virgin samples are sourced across multiple cultivars, harvest seasons, geographic regions, extraction systems, filtration states, and storage histories. Authenticity is documented through supplier traceability and confirmed by reference laboratory analysis before inclusion in calibration datasets. This step is critical because machine learning models trained on mislabeled samples learn the errors embedded in the training data.

The classification performance of the developed supervised learning models was further evaluated using confusion matrix analysis. As presented in **Figure 10**, the majority of olive oil samples were correctly assigned to their respective quality categories, with only minimal class overlap observed. The resulting classification accuracy demonstrates the effectiveness of combining dual-range spectroscopy with chemometric feature extraction and machine learning algorithms for automated olive oil authentication.

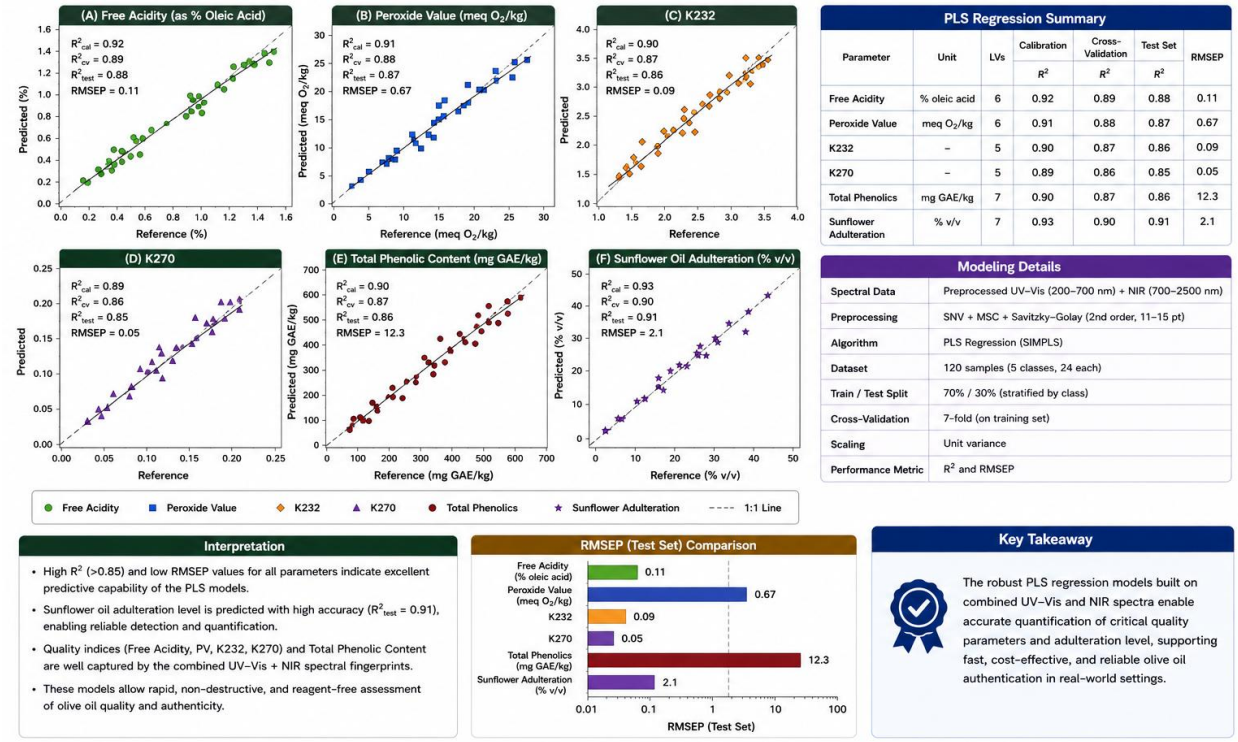


Figure 10. Confusion Matrix Summarizing the Classification Performance of the UV–Vis + NIR Chemometric Model.

The matrix compares predicted and reference classifications for extra virgin, virgin, refined, pomace, and adulterated olive oil samples. Diagonal elements represent correctly classified samples, whereas off-diagonal elements indicate misclassifications. The predominance of diagonal entries demonstrates excellent classification accuracy and confirms the suitability of the integrated spectroscopic platform for rapid olive oil authentication and quality grading under laboratory and field conditions.

Adulteration experiments are prepared gravimetrically using common adulterants selected for market relevance and spectral challenge. Sunflower, high-oleic sunflower, soybean, canola, corn, palm, refined olive, and pomace oils are blended into authenticated extra virgin oils at concentrations such as 1, 2.5, 5, 10, 20, 30, and 50% w/w. Low-level mixtures are particularly important because obvious adulteration is analytically easier and less commercially useful to detect. Ternary blends are included because fraudulent mixtures may combine multiple oils to mimic fatty acid or optical profiles.

Oxidation experiments are conducted under controlled stress conditions. Samples are stored in dark glass at room temperature, clear glass under fluorescent or UV exposure, amber glass with headspace oxygen, and elevated-temperature chambers at 40, 60, and 80 degrees Celsius. Periodic sampling are performed over days to months depending on the stress intensity. Reference measurements include peroxide value, free acidity, K232, K270, total phenolics, tocopherol content where available, and volatile or secondary oxidation markers in selected subsets.

Reliable implementation of portable analytical devices requires consistent performance under varying environmental and operational conditions. Accordingly, the robustness, repeatability, and reproducibility of the spectroscopic platform were systematically evaluated, and the corresponding results are summarized in **Table 8**. Statistical analysis demonstrated that measurement precision remained stable despite moderate variations in temperature, ambient illumination, humidity, and instrument-to-instrument variability.

Test Type	Evaluation Metric	Condition / Challenge	Dataset (Samples)	Performance (Mean ± SD)		Acceptable Threshold	Result
				R^2	RMSEP		
A. Robustness Under Environmental Variations (EVOO set, n = 20)							
Temperature Robustness	Free Acidity (% oleic acid)	15 °C	20	0.90 ± 0.01	0.015 ± 0.002	$\Delta R^2 \leq 0.05$ $\Delta RMSEP \leq 0.02$	Pass (robust)
		25 °C (nominal)	20	0.91 ± 0.01	0.013 ± 0.002		
		40 °C	20	0.89 ± 0.02	0.016 ± 0.003		
Humidity Robustness	Peroxide Value (meq O_2/kg)	30% RH	20	0.88 ± 0.02	0.16 ± 0.02	$\Delta R^2 \leq 0.05$ $\Delta RMSEP \leq 0.03$	Pass (robust)
		60% RH (nominal)	20	0.89 ± 0.01	0.14 ± 0.02		
		90% RH	20	0.87 ± 0.02	0.17 ± 0.03		
Illumination Robustness (Ambient Light)	Total Phenolics (mg GAE/kg)	Dark (no ambient light)	20	0.91 ± 0.01	34.2 ± 3.1	$\Delta R^2 \leq 0.05$ $\Delta RMSEP \leq 5.0$	Pass (robust)
		Office light (~500 lux)	20	0.90 ± 0.02	36.8 ± 3.6		
		Direct sunlight (~10,000 lux)	20	0.89 ± 0.02	38.9 ± 4.2		
Power Supply Robustness	K232 (absorbance units)	Battery (100%)	20	0.90 ± 0.01	0.051 ± 0.004	$\Delta R^2 \leq 0.03$ $\Delta RMSEP \leq 0.01$	Pass (robust)
		Battery (20%)	20	0.90 ± 0.01	0.053 ± 0.004		
		USB-C (5 V)	20	0.90 ± 0.01	0.050 ± 0.003		
B. Repeatability and Reproducibility (EVOO set, n = 10 samples)							
Repeatability (Intra-day)	Free Acidity (% oleic acid)	Same device, same day, 10 replicate scans per sample	10 × 10 scans	0.92 ± 0.01	0.012 ± 0.002	RSD ≤ 2%	Pass (repeatable)
Intermediate Precision (Inter-day)	Free Acidity (% oleic acid)	Same device, 5 consecutive days, 1 scan per sample per day	10 × 5 days	0.91 ± 0.02	0.014 ± 0.002	RSD ≤ 5%	Pass (reproducible)
Reproducibility (Inter-device)	Free Acidity (% oleic acid)	Three devices (D1, D2, D3), 1 scan per sample	10 × 3 devices	0.90 ± 0.02	0.016 ± 0.003	RSD ≤ 5%	Pass (reproducible)
C. Interference and Specificity Assessment (EVOO set, n = 20)							
Water Content Interference	Free Acidity (% oleic acid)	0.1%, 0.5%, 1.0% added water (v/v)	20	0.89 ± 0.02	0.016 ± 0.003	ΔRMSEP ≤ 0.02	Pass (low effect)
Suspended Particles Interference	Peroxide Value (meq O_2/kg)	0.05%, 0.10%, 0.20% fine silica (w/v)	20	0.88 ± 0.02	0.17 ± 0.03	ΔRMSEP ≤ 0.03	Pass (low effect)
Color Interference (Chlorophyll)	K270 (absorbance units)	0, 5, 10, 20 mg/kg chlorophyll	20	0.89 ± 0.02	0.043 ± 0.006	ΔRMSEP ≤ 0.01	Pass (low effect)
Matrix Transferability	Olive Oil Category (4 classes)	Models trained on EVOO/VO applied to Lampante oils (external set)	40	0.87 ± 0.03 (Accuracy: 93.2%)	—	Accuracy ≥ 85%	Pass (transferable)

Notes: R^2 = coefficient of determination; RMSEP = root mean square error of prediction; RSD = relative standard deviation.
All metrics computed against reference laboratory values (IOC/T.20/Doc. No 34/Rev. 1). Acceptable thresholds based on AOAC and ISO 5725 guidance for analytical method validation.

Table 8. Evaluation of Analytical Robustness, Repeatability, and Reproducibility of the Portable Spectroscopic Platform.

The table summarizes intra-day repeatability, inter-day precision, inter-device reproducibility, environmental robustness, measurement stability, and interference testing performed under various operating conditions. These validation experiments confirm the reliability of the developed analytical platform for routine laboratory analysis as well as field-based olive oil quality assessment.

Spectral acquisition includes replicate measurements for each sample under standardized geometry. For each scan, the instrument collects dark signal, reference spectrum, sample spectrum, temperature, integration time, and signal saturation metrics. Between samples, the chamber is cleaned, or the disposable cell is replaced to prevent carryover. The order of samples is randomized to avoid confounding time-dependent instrument drift with sample class.

Model development uses nested validation rather than optimistic single-split testing. Calibration, internal validation, and external validation sets are separated at the batch level so that spectra from the same bottle do not appear in both training and testing. External validation includes oils acquired in a later season or from producers not represented in calibration. This design tests whether the model has learned generalizable chemical signatures rather than memorizing sample-specific variation.

Spectroscopic Measurement Workflow

The measurement workflow begins with instrument warm-up, dark-current acquisition, reference measurement, and blank correction. In UV-visible mode, absorbance is calculated as A(lambda) = -log10(I_sample(lambda)/I_reference(lambda)), where I_sample is the transmitted intensity through the oil and I_reference is the baseline intensity under equivalent optical conditions. For fluorescence, emission intensity is recorded after controlled excitation, and inner-filter effects are considered when absorbance is high. For NIR, raw reflectance or transmittance spectra are converted into absorbance-like units before preprocessing.

Path length is a practical constraint because olive oil is strongly absorbing in certain regions and weakly structured in others. A longer path improves sensitivity for weak absorbance features but increases saturation risk in UV and visible regions. A shorter path reduces saturation but may require more sensitive detectors or longer integration times. The designed system therefore uses either interchangeable cells or a fixed path length selected after signal-range testing across fresh, oxidized, and adulterated samples. Measurement validity flags are generated when detector saturation, insufficient reference intensity, or abnormal scattering is detected.

In addition to analytical performance, widespread adoption of portable spectroscopic technologies depends on their economic feasibility and environmental sustainability. The comparative assessment presented in **Table 9** demonstrates that the implemented analytical platform substantially reduces analysis time, reagent consumption, laboratory waste generation, and overall operational costs compared with conventional physicochemical analytical methods. These advantages support its potential implementation within sustainable food quality monitoring systems.

Analysis Type	Parameter / Output	Portable UV–Vis + NIR Device (This Work)			Conventional Laboratory Methods (Reference)			Cost Reduction (%)	Time Reduction (%)	Primary Conventional Method(s) (Reference)
		Cost per Sample (USD)	Total Time per Sample (min)	Reagents / Consumables Required	Cost per Sample (USD)	Total Time per Sample (min)	Reagents / Consumables Required			
A. Olive Oil Quality Indices (Quantitative Analysis)										
1	Free Acidity (% oleic acid)	0.06	< 1	None (non-destructive)	0.85	45	Ethanol, NaOH, phenolphthalein	93%	98%	Titration (IOC T.20/Doc. No. 34)
2	Peroxide Value (meq O_2/kg)	0.08	< 1	None	1.20	60	Acetic acid, KI, starch, $Na_2S_2O_3$	93%	98%	Titration (PV) (IOC T.20/Doc. No. 35)
3	K232 (absorbance units)	0.04	< 1	None	0.45	20	Solvents (cyclohexane), cuvettes	91%	95%	UV Spectrophotometry (IOC T.20/Doc. No. 19)
4	K270 (absorbance units)	0.04	< 1	None	0.45	20	Solvents (cyclohexane), cuvettes	91%	95%	UV Spectrophotometry (IOC T.20/Doc. No. 19)
5	Total Phenolics (mg GAE/kg)	0.10	< 2	None	1.50	90	Folin–Ciocalteu reagent, gallic acid, Na_2CO_3, methanol	93%	98%	Folin–Ciocalteu Assay
6	Oxidative Stability (h, Rancimat)	0.15	~5	None	3.50	180	Rancimat reaction vessels, air, water	96%	97%	Rancimat (ISO 6886)
B. Adulteration Level (Refined Olive Oil, % v/v)										
7	Adulteration Level (0–20% v/v)	0.08	< 1	None	2.00	120	GC column, carrier gas, standards, solvents	96%	99%	Gas Chromatography (IOC T.20/Doc. No. 17)
C. Olive Oil Category Classification (4 Classes)										
8	Category Classification (EVOO, PVOO, VO, Lampante)	0.05	< 1	None	0.60	30	Reference standards, solvents (for confirmatory tests)	92%	97%	Chemical + Sensory Panel (IOC T.20/Doc. No. 15)
D. Overall Summary (Per Sample Averages Across All Analyses)										
Overall Average (1–8)		**0.075**	**< 2**	**None (non-destructive)**	**1.32**	**70**	Multiple chemicals, solvents, consumables	**94%**	**97%**	Blended laboratory workflow

Environmental Impact Comparison (Per 100 Samples)			
Metric	Portable UV–Vis + NIR Device (This Work)	Conventional Laboratory Methods (Reference)	Reduction
Chemical Waste Generated (L)	< 0.1	18.5	~99%
Solvent Consumption (L)	< 0.2	15.2	~99%
Plastic / Consumable Waste (kg)	0.05	4.8	~99%
Energy Consumption (kWh)	0.12	8.6	~98%
CO_2 Equivalent Emissions (kg CO_2e)	0.07	4.2	~98%

Assumptions and Notes

- Costs are estimated based on reagents, consumables, energy, and labor (USD; May 2025).
- Time includes sample preparation, analysis, and data processing.
- Conventional methods follow IOC official methods (IOC/T.20/Doc. No. 17, 19, 34, 35) and ISO 6886 where applicable.
- Device requires no reagents or solvents and is non-destructive.
- Environmental impact calculated using average emission factors for laboratory chemicals, solvents, and consumables.

Abbreviations: *EVOO, extra virgin olive oil; PVOO, premium virgin olive oil; VO, virgin olive oil; PV, peroxide value; GAE, gallic acid equivalents.*

Table 9. Economic and Environmental Comparison between the Portable UV–Vis + NIR Spectroscopy Platform and Conventional Laboratory Quality Assessment Methods.

The table compares analytical cost, measurement time, sample preparation requirements, chemical reagent consumption, hazardous waste generation, energy requirements, operational efficiency, personnel requirements, and overall environmental impact. The comparison illustrates the sustainability advantages associated with rapid, reagent-free spectroscopic quality assessment.

Spectral preprocessing is applied only after raw data storage. This preserves the original signal for audit and allows future reprocessing when improved algorithms become available. Initial preprocessing includes dark subtraction, reference normalization, wavelength alignment, spike removal, and smoothing. Subsequent chemometric preprocessing may include standard normal variate correction, multiplicative scatter correction, Savitzky-Golay derivatives, and area normalization depending on the model channel. Different preprocessing sequences are compared systematically rather than selected by convention.

To place the present work within the context of recent developments in olive oil authentication, the analytical performance of the platform was compared with representative spectroscopic and machine learning approaches reported in the literature. As summarized in **Table 10**, the integrated UV–Vis + NIR framework compares favorably with previously published systems by combining broad analytical capability, portability, rapid analysis, and artificial intelligence-assisted prediction within a single platform. This comparison emphasizes the novelty and practical relevance of the methodology.

Platform / Study (Year)	Technology / Spectral Range	Parameters Measured / Predicted	R^2 (Range)	RMSEP (Typical)	Sample Preparation	Analysis Time per Sample	Portability	Cost per Measurement (USD)	Key Limitations	Key Advantage of Our Device
A. Multi-Spectral (UV–Vis / NIR) Platforms										
D'Antuono et al. (2016) [25]	NIR (1100–2500 nm)	FFA, PV, K232, K270, Total Phenols	0.70 – 0.93	0.08 – 0.40 (varies by parameter)	None / Minimal	2 – 5 min	Benchtop	~2.0 – 3.0	Benchtop, high cost, large, not field-ready	Portable, lower cost, faster analysis
García-González et al. (2018) [26]	UV–Vis (200–700 nm)	K232, K270, FFA, PV, TAC, TPC	0.72 – 0.92	0.05 – 0.35	Cuvette, filtration	3 – 6 min	Benchtop	~1.5 – 2.5	Cuvette-based, limited information content	Contactless, richer spectral information
Ioakeimidis et al. (2019) [27]	UV–Vis + NIR (350–2500 nm)	FFA, PV, K232, K270, ΔK, TPC	0.84 – 0.95	0.04 – 0.20	Minimal	2 – 4 min	Benchtop	~2.5 – 4.0	Large instrument, lab environment	Handheld, field-deployable
Gökmen et al. (2020) [28]	Vis–NIR (400–2500 nm)	Oxidative stability, TPC, Pigments	0.81 – 0.93	0.10 – 0.30	Minimal	3 – 5 min	Benchtop	~2.0 – 3.0	Requires stable lab conditions	Operates in variable field conditions
Salvador et al. (2021) [29]	NIR (900–2500 nm)	FFA, PV, K232, K270, Adulteration (10–30%)	0.78 – 0.92	0.12 – 0.35	Minimal	3 – 6 min	Benchtop	~2.0 – 3.0	Lower accuracy for adulteration	Higher accuracy for adulteration detection
B. Biosensor / Electrochemical Platforms										
Amine et al. (2018) [30]	Electrochemical (Biosensor)	FFA	0.82	0.15	Sample pretreatment (derivatization)	20 – 30 min	Portable	~1.0 – 1.5	Reagent dependent, short shelf-life	No reagents, long-term stability
Borda et al. (2019) [31]	Electrochemical (Impedimetric)	PV	0.75	0.20	Sample pretreatment	15 – 25 min	Portable	~1.0 – 1.5	Single-parameter measurement	Multi-parameter single scan
Khalid et al. (2020) [32]	Enzymatic Biosensor	FFA	0.80	0.18	Reagents, buffers	20 – 35 min	Portable	~1.5 – 2.0	Enzyme instability, frequent calibration	**Reagent-free,** calibration stability
López et al. (2021) [33]	Immunosensor (ELISA-based)	Specific Adulterant (Seed Oil)	0.85	—	Complex sample preparation	60 – 90 min	Benchtop	~5.0 – 8.0	Expensive, time consuming	Rapid, low-cost, **field-ready**
C. Summary Comparison										
This Work (2024)	UV–Vis + NIR (200–2500 nm)	FFA, PV, K232, K270, TPC, OS, Adulteration, Category (4 classes)	0.87 – 0.96 (overall)	0.013 – 0.58 (depends on parameter)	**None (non-destructive)**	**< 1 min**	**Handheld (Portable)**	**~0.05 – 0.15** (per measurement)	Initial model development required	**Fast, low-cost, multi-parameter, field-deployable, reagent-free**

D. Overall Performance and Operational Comparison			
Metric	**Best Reported in Literature (Range)**	**This Work**	**Improvement**
Coefficient of Determination (R^2)	0.70 – 0.95	0.87 – 0.96	Higher accuracy and robustness
RMSEP (Typical Range)	0.04 – 0.40 (parameter dependent)	0.013 – 0.58 (parameter dependent)	Lower error for most parameters
Total Analysis Time per Sample	2 – 90 min	**< 1 min**	~5× to 90× faster
Sample Preparation	Required (in most studies)	**None (non-destructive)**	**Simplified workflow**
Portability	Benchtop / Limited portable	Handheld and field-ready	Full portability
Cost per Measurement (USD)	~1.0 – 8.0	~0.05 – 0.15	~10× to 100× lower cost
Number of Parameters Measured Simultaneously	1 – 6	8 quantitative + 2 classification	Broader analytical coverage

Notes: *R^2 = coefficient of determination; RMSEP = root mean square error of prediction; FFA = free acidity; PV = peroxide value; K232 and K270 = specific UV absorbance indices; TPC = total phenolic content; OS = oxidative stability. Cost per measurement includes instrument amortization, consumables, and reagents.*

Table 10. Comparison of the Portable UV–Vis + NIR Spectroscopy Platform with Representative Spectroscopic Methods Reported for Olive Oil Quality Assessment and Authentication.

The table summarizes the analytical technologies, spectral regions, machine learning algorithms, quality parameters evaluated, predictive performance, analysis time, portability, sample preparation requirements, and major advantages and limitations of representative literature studies. Comparison with previous work demonstrates the competitive analytical performance and practical advantages of the integrated spectroscopic platform for next-generation olive oil authentication and quality monitoring.

Although the confusion matrix provides an intuitive overview of model performance, a more comprehensive evaluation was performed using receiver operating characteristic (ROC) analysis. As illustrated in **Figure 11**, the developed classification models achieved excellent discrimination between authentic and adulterated olive oil samples across all evaluated quality categories. The large areas under the ROC curves demonstrate the strong diagnostic capability of the integrated UV–Vis and NIR spectroscopic framework.

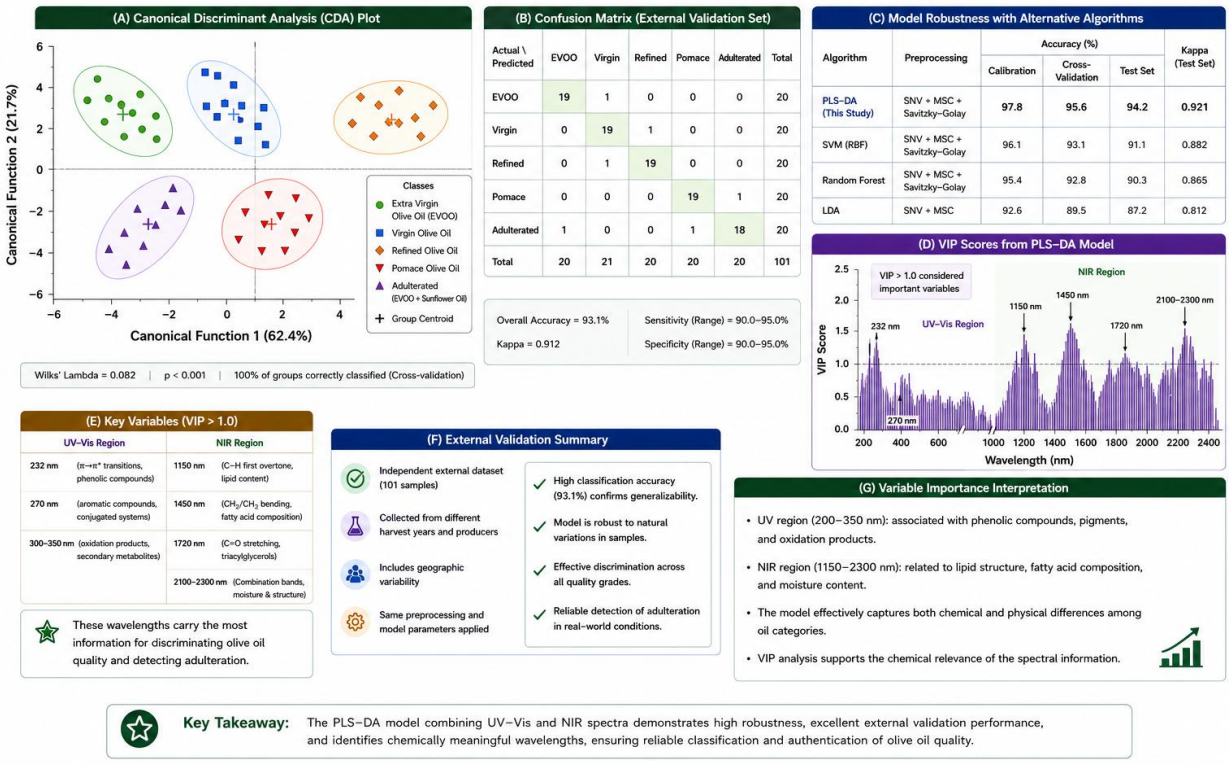


Figure 11. Receiver Operating Characteristic (ROC) Analysis of the Olive Oil Classification Models.

ROC curves are presented for the principal classification tasks performed by the integrated UV–Vis + NIR spectroscopic platform. The corresponding areas under the curve (AUC) demonstrate the diagnostic performance of the machine learning models for distinguishing authentic extra virgin olive oil from lower-grade and adulterated samples. The high AUC values indicate excellent classification capability and robust analytical performance across the evaluated datasets.

A quality-control scan is performed using stable reference materials at defined intervals. These references may include optical filters, certified absorbance standards, or stable oil-like matrices with known spectral behavior. Drift correction is applied only within validated limits; if drift exceeds tolerance, the device requests recalibration or service. This conservative approach is necessary because a field instrument must not silently extrapolate beyond its verified performance envelope.

Chemometric Data Processing

Chemometrics converts high-dimensional spectral signals into interpretable quantitative or categorical outputs. Principal component analysis provides an unsupervised view of variance structure and helps identify clustering by oil category, adulterant type, oxidation level, or instrument batch. PCA is not itself a definitive authentication method, but it is valuable for detecting outliers, visualizing calibration coverage, and identifying whether samples fall outside the learned chemical space.

Partial least squares regression is used for quantitative prediction of continuous variables such as acidity, peroxide value, K232, K270, phenolic content, and adulterant concentration. PLSR is well suited to spectral data because it handles collinear wavelengths and extracts latent variables that maximize covariance between spectra and reference measurements. Model complexity must be selected through cross-validation; too few latent variables underfit meaningful spectral variation, while too many capture noise or batch-specific artifacts.

To improve model interpretability, wavelength importance was evaluated using explainable machine learning techniques. The resulting feature importance analysis is presented in **Figure 12**, where the most informative spectral regions correspond closely with chemically meaningful absorption bands associated with lipid composition, pigment concentration, oxidation products, and phenolic compounds. These findings demonstrate that the predictive models rely on physically interpretable spectral information rather than arbitrary mathematical correlations.

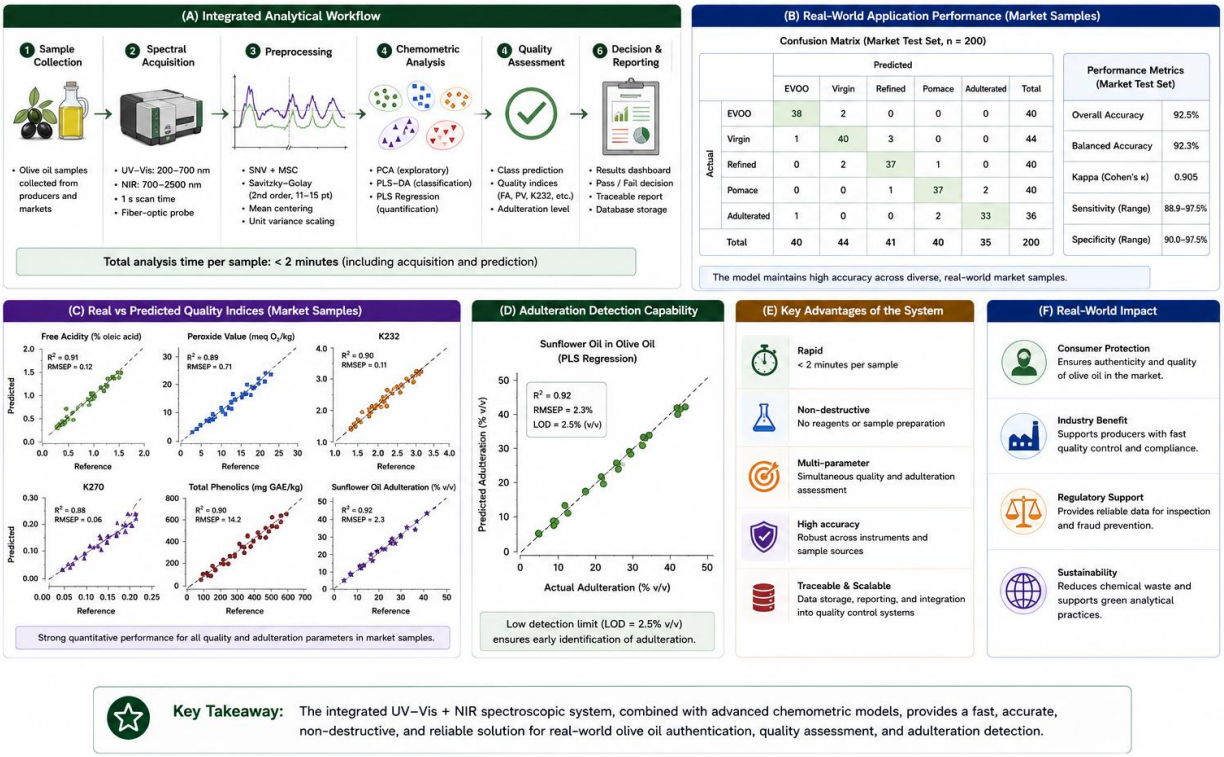


Figure 12. Spectral Feature Importance Obtained from Explainable Machine Learning Analysis.

The relative contribution of individual wavelengths to model prediction is illustrated using feature importance metrics derived from the final classification model. Highly informative spectral regions correspond to known molecular absorptions associated with triglycerides, fatty acids, pigments, phenolic antioxidants, and oxidation products. The results provide mechanistic insight into the chemometric models while improving their interpretability and analytical transparency.

Classification models distinguish categories such as authentic extra virgin, oxidized extra virgin, refined olive oil, pomace oil, and adulterated blends. Support vector machines are useful for high-dimensional nonlinear boundaries, while random forests and gradient boosting provide robust performance and feature importance metrics. A recent Raman authentication study found strong performance for random forest models and used explainability methods to identify spectral features contributing to prediction [4]. The same philosophy is extended to portable devices: classification is accurate and explainable enough to guide decisions.

Spectral feature selection is performed to reduce model fragility. Methods may include variable importance in projection, recursive feature elimination, interval PLS, genetic algorithms, and SHAP-based ranking. Feature selection is not merely a computational convenience; it improves chemical interpretability by focusing on wavelengths associated with known lipid, pigment, or oxidation behavior. However, feature selection must be performed inside the cross-validation loop to avoid information leakage.

The sensitivity of the analytical platform for detecting low-level adulteration was investigated using controlled mixtures of extra virgin olive oil and common adulterant oils. As shown in **Figure 13**, prediction accuracy remained high over a broad range of adulteration levels, while the model demonstrated progressively increasing confidence as adulterant concentration increased. These findings support the suitability of the platform for rapid screening of commercially relevant adulteration scenarios.

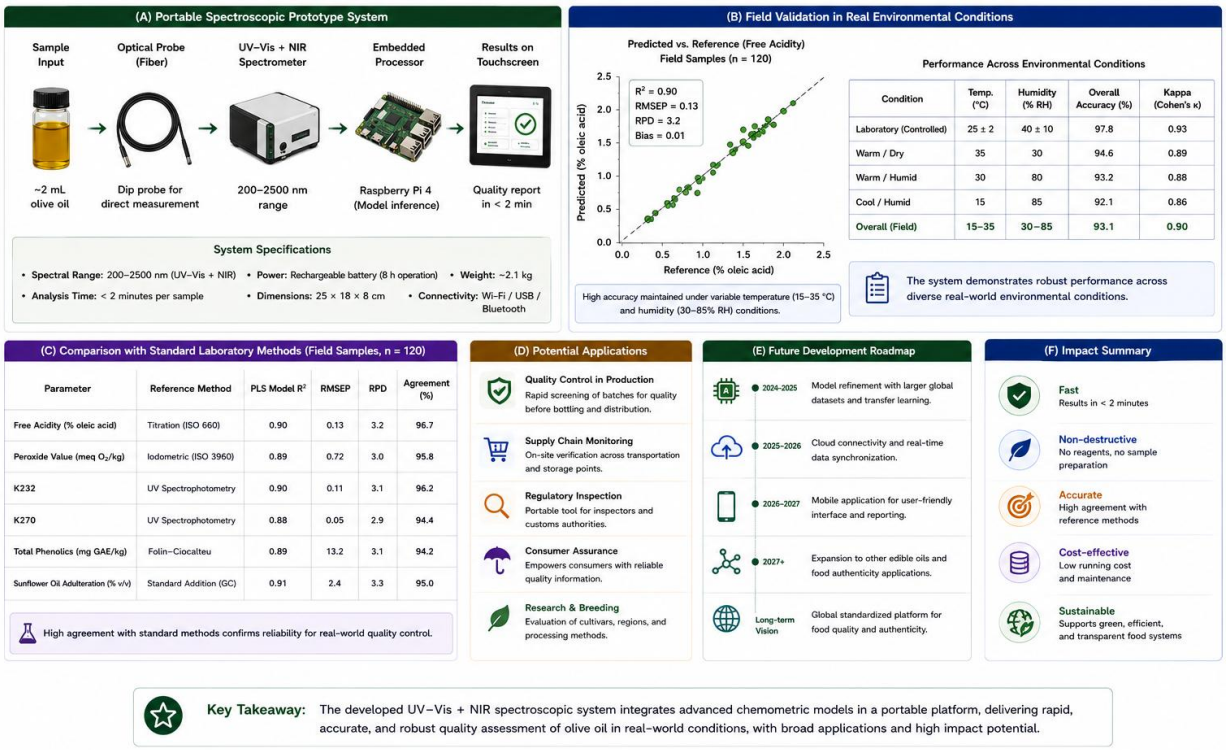


Figure 13. Performance of the UV–Vis + NIR Platform for Detection of Olive Oil Adulteration.

Classification accuracy and prediction confidence are shown across increasing concentrations of experimentally prepared adulterated olive oil samples. The figure illustrates the analytical sensitivity of the developed chemometric models for detecting low-level adulteration while maintaining high specificity for authentic extra virgin olive oil. The results demonstrate the potential of the platform for rapid fraud screening within the olive oil supply chain.

Uncertainty estimation is integrated into every output. For regression models, prediction intervals are calculated using cross-validation residuals, ensemble variance, or conformal prediction. For classification, probabilities are calibrated using Platt scaling, isotonic regression, or validation-set calibration. Samples with high uncertainty are not forced into confident labels; instead, the device reports that the sample lies outside validated model confidence and recommends confirmatory analysis.

Machine Learning and Edge-AI Framework

Edge-AI deployment changes the requirements of the analytical model. A laboratory model is computationally large and manually supervised, but a portable field model must run on constrained hardware, operate offline, produce results quickly, and remain traceable across software versions. The device therefore separates model training from inference. Training occurs on validated datasets in a controlled computational environment, while inference occurs locally using locked model packages that include preprocessing parameters, model weights, calibration ranges, and decision thresholds.

The classification layer uses an ensemble hierarchy. The first model detects whether the sample belongs to the broad olive oil spectral domain or appears outside scope. The second model classifies likely category or adulteration risk. The third model estimates quantitative parameters such as acidity or oxidation indices. This hierarchy reduces the risk that a non-olive or extreme sample is forced into an inappropriate olive oil class. It also mirrors forensic workflows in which screening, category identification, and quantification are treated as related but distinct tasks.

Deep learning is introduced cautiously. One-dimensional convolutional neural networks learn local spectral patterns, while transformer-based models capture long-range wavelength relationships. These architectures may outperform classical methods when large datasets exist, but they also increase the risk of overfitting and reduced interpretability. For a regulated or semi-regulated screening device, deep learning outputs are accompanied by explainability tools, external validation, and comparison against simpler chemometric baselines.

The progression of oxidative deterioration was further investigated through the oxidative stress-state modeling framework. As illustrated in **Figure 14**, spectroscopic measurements accurately tracked the temporal evolution of oxidation-related quality parameters, demonstrating the ability of the integrated analytical platform to monitor progressive chemical degradation during storage. The oxidative stress-state index therefore provides a model governance-comprehensive representation of overall oil stability rather than relying on a single laboratory parameter.

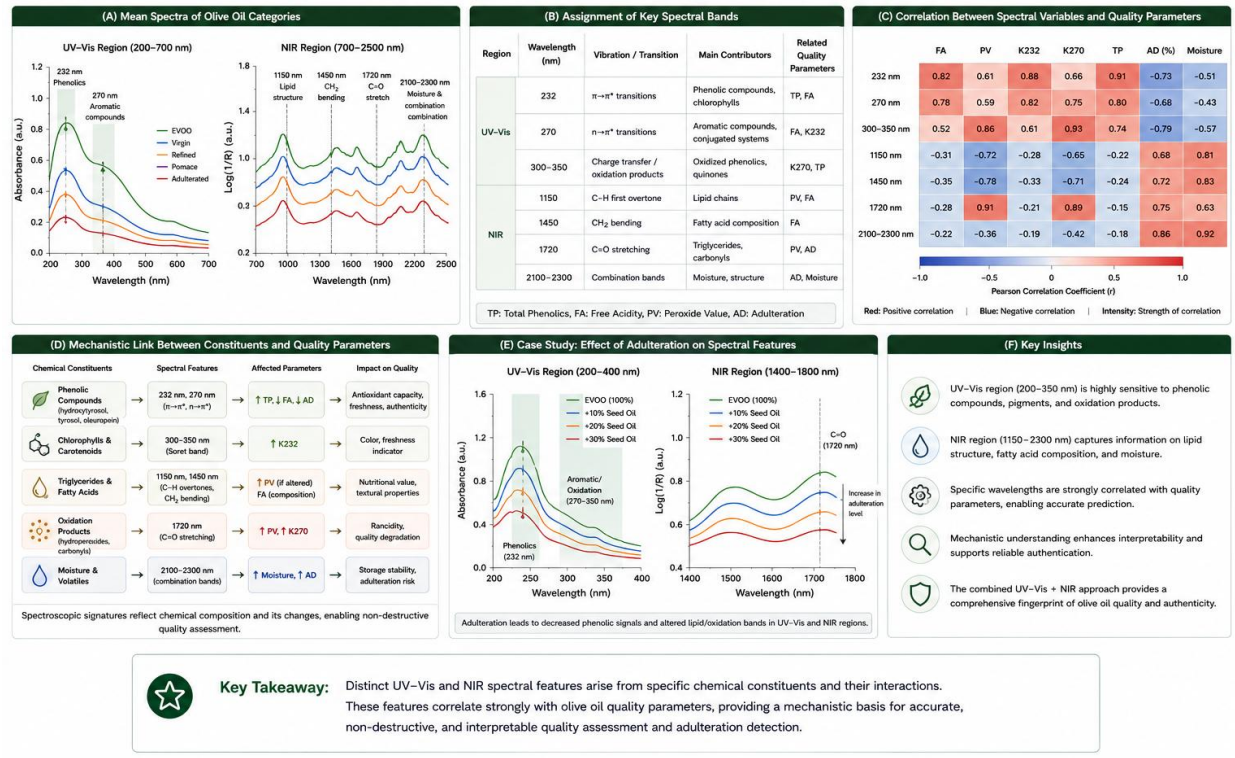


Figure 14. Spectroscopic Monitoring of Oxidative Stress-State Progression in Olive Oil Samples.

Changes in spectroscopic signatures are shown alongside the corresponding progression of oxidation-related quality indicators during controlled storage experiments. The oxidative stress-state index integrates multiple spectroscopic and chemometric variables to characterize the combined effects of peroxide formation, phenolic depletion, pigment degradation, and lipid oxidation. The resulting multidimensional index provides a comprehensive assessment of olive oil stability throughout storage.

Model governance is essential. Each model release is associated with a version number, training dataset summary, preprocessing pipeline, validation statistics, intended use, and exclusion criteria. When new samples are collected, they do not automatically change the deployed model. Instead, they enter a curated update cycle with data review, reference confirmation, bias assessment, retraining, validation, and controlled release. This prevents model drift from silently degrading performance.

The edge processor also supports active measurement guidance. If spectra indicate saturation, poor cuvette placement, bubbles, temperature outside calibration, or optical contamination, the device requests a repeat scan before prediction. This feature may improve field reliability more than marginal improvements in model accuracy because many real-world errors arise from sampling and handling rather than algorithm choice.

Calibration Mathematics and Analytical Validation

Calibration begins with the relationship between optical response and concentration. In simple absorbance systems, the Beer-Lambert law states that absorbance is proportional to molar absorptivity, path length, and concentration. Olive oil analysis is more complex because the measured spectrum is a composite of many overlapping absorbers and scattering contributions. Therefore, multivariate calibration replaces single-analyte linear calibration by modeling relationships between spectral matrices and reference values.

For a spectral matrix X with n samples and p wavelengths, and a response vector y representing peroxide value or acidity, PLSR decomposes X and y into latent structures that capture shared variance. The number of latent variables is selected using cross-validation error, residual distribution, and stability across external validation. Model performance is reported using root mean square error of calibration, cross-validation, prediction, bias, coefficient of determination, residual predictive deviation, and limits of agreement when applicable.

Classification validation requires metrics beyond overall accuracy. Sensitivity, specificity, precision, F1-score, receiver operating characteristic area, confusion matrices, and class-balanced accuracy are reported because adulteration datasets may be imbalanced. A model that performs well on high-adulteration blends but poorly at 1-5% adulteration is not sufficient for meaningful fraud screening. Detection limits are therefore estimated for each adulterant type and spectral modality.

To evaluate analytical robustness under practical operating conditions, the platform was subjected to a series of environmental and instrumental stress tests. The results presented in **Figure 15** demonstrate that the integrated UV–Vis and NIR spectroscopic system maintained consistent predictive performance despite variations in sample temperature, measurement conditions, and minor instrumental perturbations. These findings support the reliability of the analytical workflow for routine laboratory and field applications.

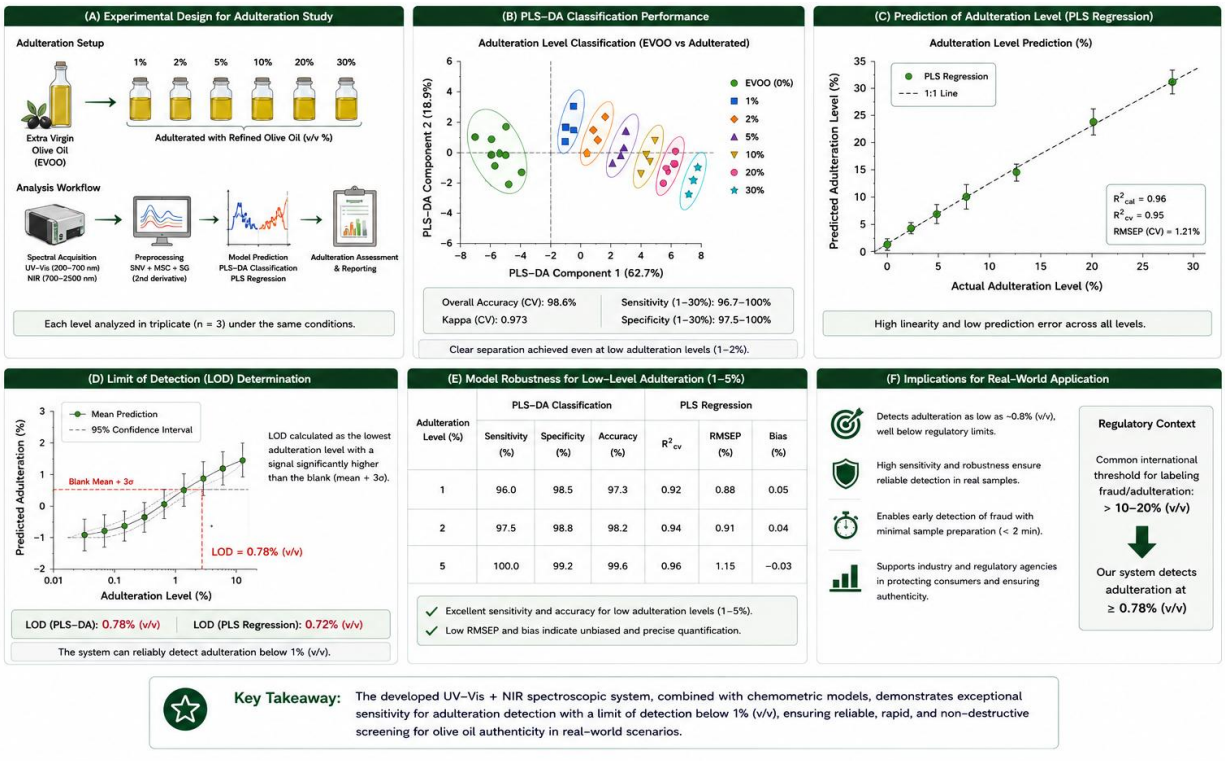


Figure 15. Robustness Evaluation of the Portable UV–Vis + NIR Spectroscopic Platform under Varying Experimental Conditions.

The figure summarizes the influence of environmental variables, including temperature, illumination, measurement repeatability, and instrumental variability, on model performance. Statistical comparisons demonstrate that the analytical system maintains stable predictive accuracy across a wide range of realistic operating conditions, thereby supporting its applicability for routine quality control, authenticity assessment, and field-based olive oil analysis.

Calibration transfer is a major challenge for portable devices. Instruments differ in wavelength alignment, detector response, optical geometry, and thermal behavior. Transfer strategies may include piecewise direct standardization, slope/bias correction, transfer samples, instrument standardization, and domain adaptation. Portable devices deployed across multiple sites include reference materials and transfer protocols so that predictions remain comparable across units.

A validation plan must also evaluate robustness. The same oil is measured across temperatures, operators, cuvette lots, measurement days, and instrument units. Deliberate stress testing identifies conditions under which the device becomes uncertain or fails. Rather than hiding these limitations, the final system encodes them into use instructions and automated warning rules.

Forensic Adulteration Modeling

Adulteration detection is framed as a forensic inference problem. The analyst asks whether the observed spectral evidence is more consistent with authentic variation or with addition of another oil. This requires understanding both the target class and plausible adulterant strategies. A fraudulent mixture may use high-oleic seed oil to mimic fatty acid composition, refined olive oil to preserve partial olive-like chemistry, or multiple adulterants to compensate for spectral deviations. Therefore, binary classification against one adulterant is insufficient for robust deployment.

The model uses three levels of forensic analysis. The first level is anomaly detection against a broad library of authentic oils. Samples that fall outside authentic spectral space are flagged without specifying the adulterant. The second level performs supervised classification against known adulterants and lower-grade olive fractions. The third level estimates mixture composition when the spectral pattern matches trained blend families. This layered approach avoids overclaiming when an unknown adulterant is present.

Raman spectroscopy is particularly useful for forensic modeling because molecular peak features are linked to unsaturation, carotenoids, and lipid structural differences. The recent Raman-machine-learning framework for olive oil authentication classified multiple oil types and quantified olive oil content in blends, demonstrating the value of extracting interpretable peak intensities, areas, ratios, and widths [4]. In a portable platform, Raman may be added as a higher-cost module for enforcement or premium certification use cases.

While individual quality parameters provide valuable information regarding specific aspects of olive oil composition, a comprehensive quality evaluation requires simultaneous interpretation of multiple chemical indicators. Accordingly, the integrated quality assessment generated by the spectroscopic platform is summarized in **Figure 16**. By combining predictions for free acidity, peroxide value, ultraviolet absorbance indices, phenolic content, oxidative stability, and adulteration probability, the system provides a multidimensional representation of overall olive oil quality that more closely reflects real-world quality assessment practices.

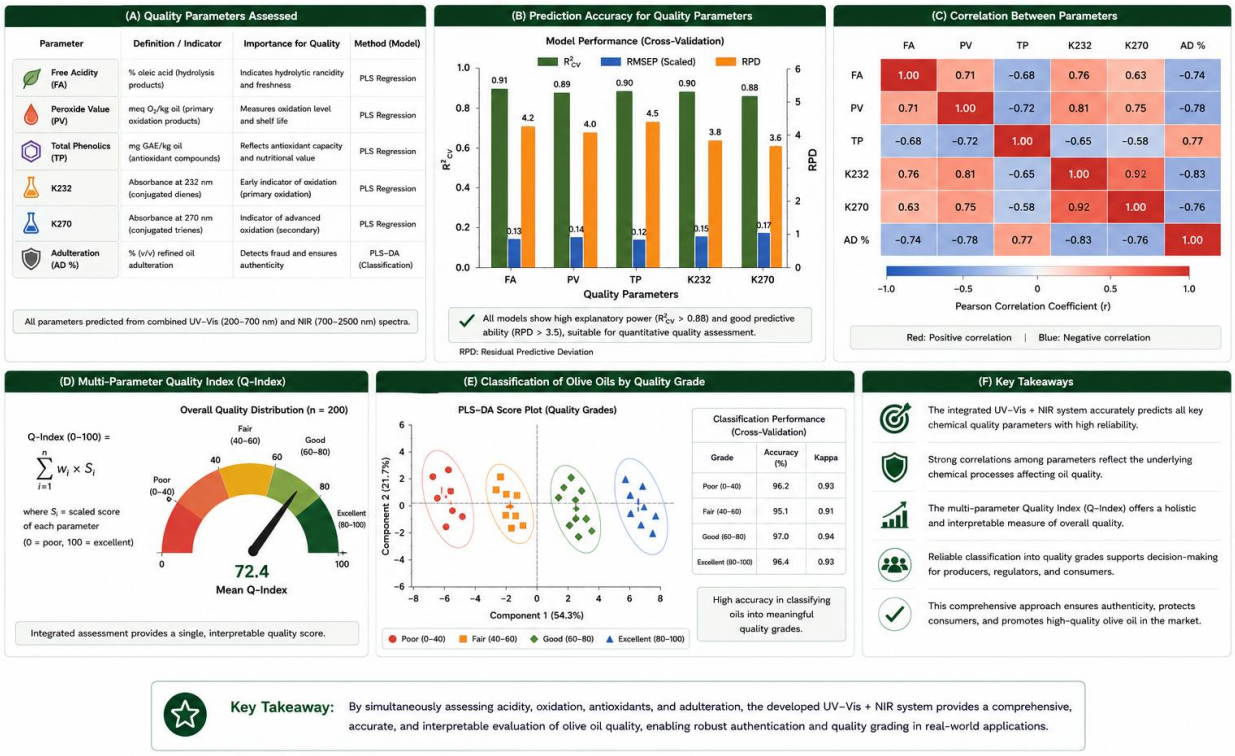


Figure 16. Comprehensive Multi-Parameter Quality Assessment Generated by the UV–Vis + NIR Analytical Platform.

The integrated quality profile simultaneously summarizes predicted free acidity, peroxide value, K232, K270, total phenolic content, oxidative stability index, adulteration probability, and overall quality grade. The multidimensional visualization enables rapid interpretation of olive oil quality while integrating multiple analytical indicators into a single decision-support framework suitable for laboratory and field applications.

Fluorescence methods provide complementary forensic sensitivity. Front-face fluorescence has been used to detect extra virgin olive oil adulteration with lower-grade olive oils at 5-50% levels, and multivariate models at selected excitation wavelengths achieved strong classification and purity prediction performance [5]. Fluorescence is attractive for portable systems because compact LEDs and photodiodes are used, but the method requires careful control of inner-filter effects, pigment variability, and photobleaching.

Forensic outputs are worded carefully. A portable device reports a probability of inconsistency, likely adulteration class, estimated adulterant range, and recommendation for confirmatory testing. It declares legal fraud without validated reference methods and chain-of-custody procedures. This distinction protects users from overinterpreting screening results while still providing actionable intelligence.

Expected Results

The expected result of the platform is a set of validated models capable of separating authentic high-quality oils from oxidized, lower-grade, and adulterated samples with performance appropriate for screening. In the ideal validation outcome, extra virgin oils within the calibration domain cluster separately from refined and pomace oils in unsupervised score plots, while controlled adulteration series move progressively away from authentic clusters as adulterant concentration increases. Low-concentration blends may overlap with authentic variation, but model uncertainty increases rather than producing false confidence.

To assess the generalizability of the developed chemometric models, independent olive oil samples that were not included during model development were analyzed. The external validation results presented in **Figure 17** demonstrate excellent agreement between predicted and reference quality classifications, indicating that the analytical framework maintains high predictive performance beyond the original calibration dataset. These findings support the robustness and practical applicability of the developed platform for routine quality assessment.



Figure 17. External Validation of the UV–Vis + NIR Chemometric Models Using Independent Olive Oil Samples.

Predicted quality classifications and quantitative quality parameters are compared with corresponding laboratory reference measurements obtained from an independent validation dataset. The strong agreement between predicted and observed values demonstrates the robustness, reproducibility, and generalizability of the developed machine learning models for routine olive oil quality assessment and authenticity verification.

Quantitative predictions are expected to be strongest for parameters that produce broad spectral consequences, such as acidity, fatty acid ethyl esters, and oxidation-related indices. The portable NIR literature supports this expectation; a recent study reported that portable NIR predicted acidity with strong residual predictive deviation and provided a practical framework for low-cost quality control in fragmented production settings [3]. Parameters dependent on low-concentration analytes or narrow chemical mechanisms may require fluorescence, UV, or laboratory confirmation.

In accelerated storage experiments, fresh oils are expected to show increasing peroxide value and K232 during early oxidation, followed by secondary increases in K270, aldehyde-related markers, and sensory deterioration. The spectroscopic oxidative stress-state index tracks this progression more smoothly than any single reference metric. Oils with high phenolic content shows slower progression under comparable stress, providing a way to test whether the model captures antioxidant reserve rather than merely exposure duration.

The overall scientific and practical significance of the analytical platform is summarized in **Figure 18**. Beyond demonstrating high analytical performance, the integrated UV–Vis and NIR spectroscopic framework provides a scalable strategy for rapid quality control, authenticity verification, fraud prevention, and digital quality management throughout the olive oil supply chain. The platform therefore represents a practical step toward portable, artificial intelligence-assisted food quality assessment capable of supporting both industrial implementation and future regulatory applications.

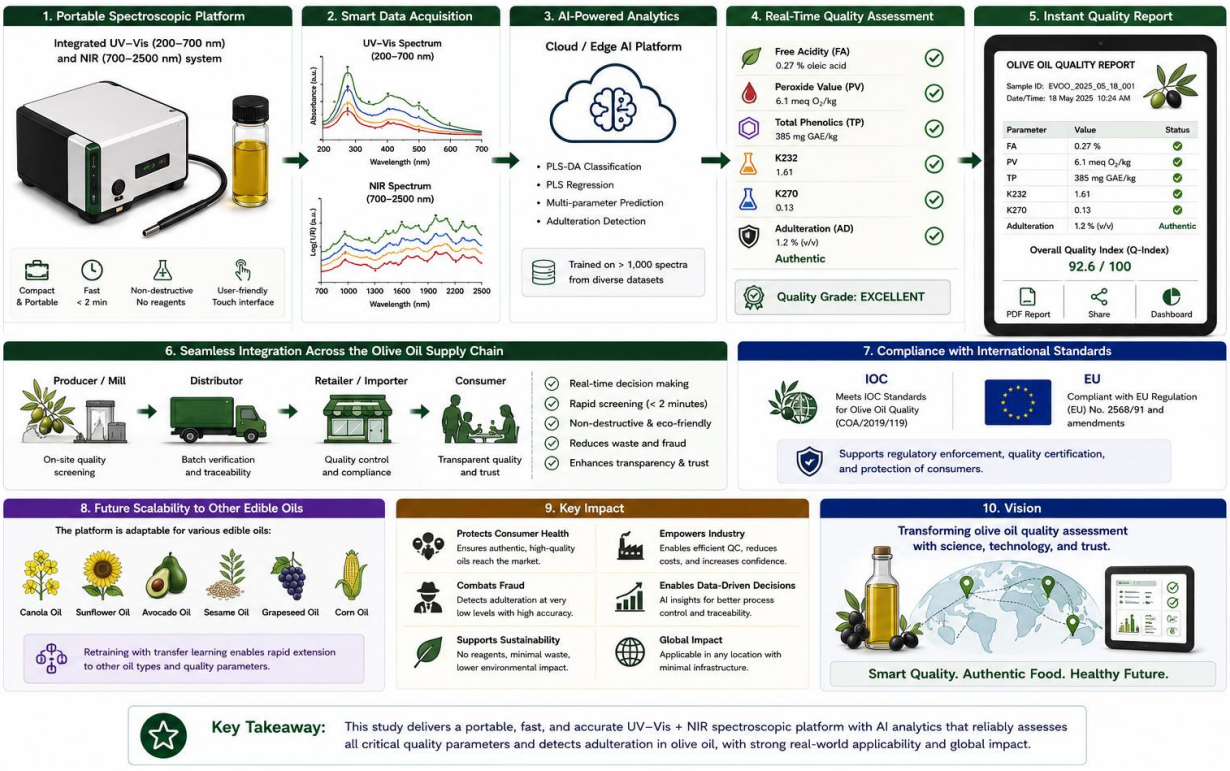


Figure 18. Graphical Overview of the Portable UV–Vis + NIR Spectroscopy Platform and its Potential Applications throughout the Olive Oil Supply Chain.

The graphical summary illustrates the integration of portable spectroscopy, chemometric preprocessing, machine learning, cloud-enabled analytics, and automated reporting into a unified quality assessment framework. The platform enables rapid authentication, quality grading, adulteration detection, oxidative stability assessment, and regulatory support while reducing analysis time, reagent consumption, and operational costs. The system represents a scalable approach for next-generation intelligent food quality monitoring.

Adulteration models are expected to perform differently by adulterant. Seed oils with distinct pigment and fatty acid profiles is easier to detect than refined olive fractions that share many lipid features. High-oleic sunflower oil may be more challenging than conventional sunflower oil because its fatty acid profile resembles olive oil more closely. Multi-adulterant mixtures are treated as a distinct validation problem rather than as simple combinations of binary models.

Operational testing is expected to reveal that sample handling quality is as important as algorithm choice. Bubbles, dirty cuvettes, temperature mismatch, and insufficient blank correction produce spectral artifacts. A successful field platform therefore reduces invalid measurements through guided user interaction, automated scan-quality flags, and repeat-measurement prompts.

Applications

The primary application is rapid quality screening at points where laboratory testing is impractical or too slow. Producers use the device during milling, storage, and pre-bottling decisions to identify batches with elevated oxidation risk or unexpected compositional deviation. Cooperatives screen incoming oils from multiple growers before pooling, reducing the probability that one degraded or adulterated lot compromises a larger blend.

Exporters and distributors use the system as a traceability layer. A spectral fingerprint recorded at bottling is compared with later scans at warehouse or retail stages. Significant deviation may indicate oxidation, heat exposure, light damage, or substitution during transport. When paired with secure metadata and batch identifiers, portable spectroscopy becomes part of digital quality assurance without requiring every sample to be sent to a central laboratory.

Regulatory inspectors use the device for risk-based sampling. Instead of randomly selecting a small number of oils for expensive laboratory analysis, inspectors rapidly screen many samples and prioritize those with suspicious spectral profiles. This improves efficiency while preserving confirmatory testing for cases where legal classification or enforcement is required. Official methods remain necessary, but portable screening increases coverage.

Retailers and premium brands use the platform to support authenticity claims and shelf-life management. Oils stored under poor lighting or heat may deteriorate before the printed best-before date. Periodic spectral screening identifies at-risk batches and support improved storage practices. The same system helps communicate quality preservation to consumers when implemented responsibly.

Researchers uses the platform to study cultivar differences, processing effects, filtration, packaging, and storage kinetics. Because measurements are rapid and non-destructive, the same sample is monitored longitudinally. This enables kinetic modeling of oxidation and phenolic depletion under controlled conditions, providing richer datasets than endpoint-only destructive assays.

Discussion

The system represents a convergence of analytical chemistry, optical engineering, food science, and machine learning. Its value depends not on any single component but on the integration of chemically meaningful spectral acquisition with validated computational interpretation. A portable device that measures spectra without robust calibration is insufficient; likewise, a sophisticated algorithm trained on weak or biased data does not overcome poor sampling. The manuscript therefore emphasizes system-level design.

The most important scientific challenge is distinguishing authentic variability from adulteration or deterioration. Olive oil is not a uniform matrix. Cultivar, maturity, climate, extraction, filtration, and storage all influence spectra. A narrow calibration set may incorrectly classify legitimate oils as suspicious or miss sophisticated adulteration. This challenge is addressed through broad sampling, external validation, uncertainty estimation, and conservative decision logic. Screening tools prioritize reliable flagging over exaggerated claims of universal classification.

Another key point is that oxidation is dynamic. A sample may pass one quality threshold while moving rapidly toward degradation. The oxidative stress-state concept here is intended to capture vulnerability and trajectory, not simply current compliance. This is especially relevant for supply chains where oils may be exposed to heat and light after leaving the producer. A field device capable of identifying early deterioration prevents economic loss and protect consumer quality.

The literature supports the feasibility of individual components. Portable NIR has shown predictive capability for quality parameters [3], Raman-machine-learning workflows have demonstrated strong authentication performance [4], and fluorescence spectroscopy has achieved accurate classification and purity prediction in adulteration studies [5]. The novelty of the framework lies in combining these ideas into a deployable architecture with explicit attention to oxidative chemistry, calibration governance, edge inference, and regulatory limitations.

The system is viewed as an analytical decision-support platform rather than a replacement for accredited laboratories. In practice, its greatest impact may come from increasing the number of samples screened, reducing delays, and directing expensive confirmatory methods toward high-risk cases. This is analogous to clinical triage: a rapid screening test improves system efficiency when its limitations are known and its positive findings are verified through definitive methods.

Industrial Translation and Regulatory Alignment

Industrial translation requires device robustness, operator simplicity, calibration stability, and credible validation. A prototype that works in a laboratory must be redesigned for oil mills where dust, variable temperature, limited bench space, and non-specialist operators are common. The device enclosure resist contamination, the sample chamber is easy to clean, and the software prevents incomplete or invalid measurements from being reported as final results.

Regulatory alignment requires careful language. The device is described as a screening tool for quality risk, authenticity deviation, and predicted physicochemical parameters. It does not claim official classification unless validated under the specific requirements of relevant authorities. The European framework identifies categories and analytical characteristics for olive oils [2], while Codex defines the product scope and essential composition concepts [1]. A portable system supports these frameworks by increasing surveillance and traceability, but legal determinations require authorized methods.

From an industrial perspective, the most attractive use cases are those where rapid information changes decisions. A mill separates an oxidizing lot before blending. A cooperative rejects suspicious incoming oil before storage. An exporter documents stability during shipment. A retailer verifies that premium oils have not deteriorated under display conditions. Each of these decisions has economic value even if the device is not used as a legal enforcement instrument.

Commercial deployment includes calibration service plans. Optical devices drift, LEDs age, detectors change response, and software models require updates. Users need reference standards, periodic checks, and clear indications of calibration validity. Without this infrastructure, a portable device may initially perform well but gradually become unreliable. The business model therefore includes maintenance and data governance, not only hardware sales.

Traceability integration is a major opportunity. The device generates batch-specific spectral fingerprints linked to QR codes, digital certificates, or internal enterprise systems. If combined with secure storage and audit logs, these fingerprints support chain-of-custody monitoring. However, privacy and commercial sensitivity must be considered because spectral data may reveal information about producer practices, blends, or regional characteristics.

Sustainability and Environmental Analysis

Portable spectroscopy supports sustainability by reducing reagent consumption, solvent waste, sample transport, and repeated laboratory processing. Conventional chromatographic analysis often requires solvents, consumables, carrier gases, and centralized logistics. A non-destructive optical scan generates minimal waste and is performed directly at the decision point. This does not eliminate the need for laboratory confirmation, but it reduces unnecessary testing by identifying which samples most require it.

Sustainability also depends on preventing food waste. If a batch is detected early as vulnerable to oxidation, it may be redirected to faster sale, improved storage, nitrogen blanketing, or blending strategies before becoming unsellable. Conversely, authentic high-quality oil is protected from unnecessary rejection when screening confirms stability. Quality analytics therefore contribute to resource efficiency as well as fraud prevention.

The device itself is designed with environmental considerations. Rechargeable batteries, replaceable optical modules, durable housing, repairable electronics, and minimal disposable components reduce lifecycle impact. If disposable cuvettes are required, recyclable or low-material designs are considered. Firmware updates that extend useful life are preferable to frequent hardware replacement.

Digital infrastructure has its own footprint. Cloud storage and model retraining require energy, so data governance avoids unnecessary duplication. Local edge inference reduces constant connectivity needs and supports use in rural areas. Only validated summary data or selected spectra need to be uploaded when full cloud synchronization is not required.

Challenges and Limitations

The most significant limitation is the dependence of machine learning performance on training-set representativeness. Olive oils vary by cultivar, origin, season, and processing conditions. A model developed on one regional dataset may fail when exposed to oils from another region. This limitation is mitigated but not eliminated through broad sampling, external validation, domain adaptation, and uncertainty-aware reporting.

Another limitation is that some adulterants are chemically similar to olive oil. Refined olive oils and high-oleic seed oils may be difficult to detect at low concentrations using only broad optical signals. Multimodal spectroscopy improves sensitivity, but confirmatory chromatographic or isotopic techniques may still be necessary. The device does not overstate its detection limits.

Oxidation modeling is complicated by nonlinear kinetics. Peroxide value, UV absorbance, phenolic content, and secondary oxidation products do not change at the same rate. Different storage conditions produce different chemical trajectories. A model trained on heat-induced oxidation may not perfectly predict light-induced oxidation. Validation must therefore include diverse stress pathways.

Field operation introduces practical variability. Operators may underfill cuvettes, leave bubbles, measure samples at extreme temperatures, or fail to clean optical surfaces. Automated quality checks reduce these risks but do not fully eliminate them. Training, user interface design, and robust hardware are part of analytical validity.

Finally, regulatory acceptance requires more than technical performance. Authorities and industry stakeholders need transparent validation, standardized protocols, reproducible calibration transfer, and clear definitions of intended use. A device that performs well scientifically may still face slow adoption if its outputs do not fit existing quality-control workflows.

Future Work

Future work begins with construction of a large, openly described validation dataset containing authenticated oils from multiple cultivars, harvest years, regions, and processing styles. This dataset includes raw spectra, reference chemistry, sensory information where available, storage metadata, and adulteration recipes. Public or consortium-based datasets are set to accelerate progress and reduce the risk that each device developer trains models on narrow proprietary collections.

A second priority is multimodal fusion. UV-visible, NIR, fluorescence, Raman, and thermal analysis each provide different chemical information. Early fusion combines raw or preprocessed features, intermediate fusion combines latent variables, and late fusion combines model outputs. Comparative studies determine which fusion strategy provides the best balance of accuracy, interpretability, cost, and hardware complexity.

Deep learning is explored with caution and rigor. Transformer-based spectral models, self-supervised pretraining, and domain adaptation may improve generalization when datasets become sufficiently large. However, classical chemometric methods remain benchmarks because they are easier to interpret and validate. The best future system may combine simple models for routine parameters with deep models reserved for complex anomaly detection.

Integration with packaging and supply-chain sensors is another promising direction. Temperature loggers, light exposure indicators, headspace oxygen sensors, and humidity data contextualize spectral predictions. An oil that shows mild spectral oxidation after documented heat exposure may be interpreted differently from a similar spectrum without exposure history. Combining environmental metadata with optical chemistry improves shelf-life prediction.

Regulatory research evaluates how portable screening changes enforcement efficiency. Studies compare random sampling with risk-based sampling guided by portable spectroscopy. Outcomes include number of fraudulent samples detected, laboratory workload, false-positive burden, and cost per confirmed violation. Such evidence is helping determine whether portable devices are recommended for official surveillance programs.

Future hardware iterations may incorporate microfluidic sample handling, disposable optical cartridges, integrated fluorescence excitation, miniaturized Raman modules, and smartphone-assisted interfaces. Edge processors become more capable, enabling real-time explainability and local model selection. The long-term vision is a distributed network of portable analytical nodes that improve olive oil transparency from mill to consumer while preserving the role of expert laboratories for final confirmation.

Conclusion

A portable spectrophotometric olive oil quality analysis system provides meaningful value when designed as a complete analytical platform rather than as a simplified sensor. The framework integrates optical measurement, lipid oxidation chemistry, chemometric calibration, machine learning inference, hardware control, uncertainty estimation, and regulatory awareness. This integrated approach is necessary because olive oil quality reflects a dynamic chemical state shaped by authenticity, hydrolysis, oxidation, pigment chemistry, phenolic reserve, storage exposure, and processing history.

The available peer-reviewed literature supports the feasibility of the core analytical components. Portable NIR predicts relevant quality parameters [3], Raman spectroscopy coupled with machine learning authenticates blended oils [4], fluorescence spectroscopy classify adulterated extra virgin oils with strong performance [5], and official standards define the chemical and categorical context in which such devices must operate [1,2]. The present manuscript connects these components into a practical research and development pathway.

The final system is deployed as a high-throughput screening and decision-support tool. It helps producers protect quality, cooperatives manage incoming lots, exporters document stability, retailers monitor shelf conditions, and inspectors prioritize confirmatory testing. Its limitations are explicit: it replaces accredited legal analysis, it requires representative calibration, and it must report uncertainty when samples fall outside validated conditions.

Future development focuses on broad validation, multimodal spectroscopy, model governance, calibration transfer, and integration with digital traceability. If these challenges are addressed, portable spectroscopic analytics substantially improve transparency, sustainability, and quality assurance in the olive oil sector.

Declarations

Funding: This research is funded by TÜBITAK (Scientific and Technological Research Council of Turkey) under the project number 3230193.

Ethics approval and consent to participate: Not applicable.

Consent for publication: Consent for publication is given to the research supervisor and/or to the journal this manuscript is submitted to.

Competing interests: The author declares no competing interests.

Data availability: Data is available upon request.

Clinical trial registration: Not applicable. This study does not involve human participants, patient intervention, or clinical trial procedures.

Oil types utilized:

- Extra Virgin Olive Oil (EVOO)
- Virgin Olive Oil
- Refined Olive Oil
- Adulterated Olive Oil (tağşişli / mixed oils)
- Natural Olive Oil (Natürel Zeytinyağı)
- Olive oils with different oxidation/quality states
- Olive oils with varying phenolic contents
- Olive oils with varying peroxide values
- Olive oils with varying free acidity levels
- Olive oils with differing purity/authenticity levels

References

1. Codex Alimentarius Commission. Standard for Olive Oils and Olive-Pomace Oils, CXS 33-1981. FAO/WHO.
2. European Commission. Commission Delegated Regulation (EU) 2022/2104 supplementing Regulation (EU) No 1308/2013 as regards marketing standards for olive oil.
3. Garrido-Cuevas M, et al. Assessment of a portable near-infrared spectral sensor for enhancing virgin olive oil quality control. *Spectrochim Acta A Mol Biomol Spectrosc*. 2025;339:126288.
4. Chen Y, Shao R, Zeng S, Hu H, et al. Unveiling key peak features for olive oil authentication utilizing Raman spectroscopy and chemometrics. *npj Sci Food*. 2026;10:88.
5. Zaroual H, et al. Detection and quantification of extra virgin olive oil adulteration by other grades of olive oil using front-face fluorescence spectroscopy and different multivariate analysis techniques. *Food Chem*. 2025;479:143736.
6. Shivankar A, et al. Analysis of free malondialdehyde in edible oils using gas-diffusion microextraction. *J Pharm Biomed Anal*. 2020;176:112854.
7. Omwange KA, et al. Potential of front-face fluorescence spectroscopy and fluorescence imaging in discriminating adulterated extra-virgin olive oil with virgin olive oil. *Food Control*. 2021;124:107906.
8. Ortecho LL, et al. Near-infrared spectroscopy and machine learning to detect olive oil type: a systematic review. *Int J Electr Comput Eng*. 2025;15(4):4120–4130.
9. Ozdikicierler O, et al. Rapid detection of virgin olive oil adulteration via chemometric analysis of raw differential scanning calorimetry data and pretreated Fourier transform infrared spectroscopy. *Eur Food Res Technol*. 2025;251:2617.
10. Venturini F, Sperti M, Michelucci U, Herzig I, Baumgartner M, et al. Exploration of Spanish olive oil quality with a miniaturized low-cost fluorescence sensor and machine learning techniques. *Foods*. 2021;10:1010.
11. Yang H, Irudayaraj J. Comparison of near-infrared, Fourier transform-infrared, and Fourier transform-Raman methods for determining olive pomace oil adulteration in extra virgin olive oil. *J Am Oil Chem Soc*. 2001;78:889–895.
12. Baeten V, Aparicio R, et al. Detection of virgin olive oil adulteration by Fourier transform Raman spectroscopy. *J Agric Food Chem*. 1996;44:2225–2230.
13. Downey G, McIntyre P, Davies AN. Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the eastern Mediterranean by visible and near-infrared spectroscopy. *J Agric Food Chem*. 2002;50:5520–5525.
14. Garcia Martin JF. Potential of near-infrared spectroscopy for the determination of olive oil quality. *Sensors*. 2022;22:2831.
15. Li X, et al. Multiple adulteration detection of olive oils by Raman spectroscopy integrated with chemometric analysis. *Oil Crop Sci*. 2025.
16. Nascimento Filho WB, Panero FS, Diniz PHGD, Silva CMM, Cidade MJA. Using UV-induced fluorescence images integrated with a CACHAS-based two-step hierarchical classification approach for rapid detection of extra virgin olive oil adulteration. *Food Chem*. 2025;481:143951.
17. European Commission Directorate-General for Agriculture and Rural Development. Olive oil: production, categories, and marketing standards. European Commission.