Journal Pre-proofs

Quantitative Detection of Adulteration in Brown Sumac (Rhus coriaria) Powder Using Hyperspectral Imaging and Machine Learning

Esmat Kishani Farahani, Seyedehsamaneh Shojaeilangari b, Alireza Basiric

DOI: https://doi.org/ 10.22104/ift.2026.8058.2259

To appear in: Innovative Food Technologies (IFT)

Received Date: 14 December 2025
Revised Date: 5 January 2026
Accepted Date: 26 January 2026

Please cite this article as: Esmat Kishani Farahani, Seyedehsamaneh Shojaeilangari b, Alireza Basiric, Quantitative Detection of Adulteration in Brown Sumac (Rhus coriaria) Powder Using Hyperspectral Imaging and Machine Learning, *Innovative Food Technologies (2026)*,
doi: https://doi.org/ 10.22104/ift.2026.8058.2259

Quantitative Detection of Adulteration in Brown Sumac (*Rhus coriaria)* Powder Using Hyperspectral Imaging and Machine Learning

Esmat Kishani Farahani [a, #, *], Seyedehsamaneh Shojaeilangari [b, #], Alireza Basiri[c]

[a] Information Technology and Intelligent Systems Group, Department of Electrical Engineering and Information Technology, Iranian Research Organization for Science and Technology (IROST), 33535111, Tehran, Iran.
[b] Biomedical Engineering Group, Department of Electrical Engineering and Information Technology, Iranian Research Organization for Science and Technology (IROST), 33535111, Tehran, Iran.
[c] Department of Chemical Technologies, Iranian Research Organization for Science and Technology (IROST), 33535111, Tehran, Iran.

# Authors contributed equally and share first authorship

* Corresponding authors e-mails: e.kishani@irost.ir

Abstract
This study explores the potential of hyperspectral imaging (HSI) combined with machine learning for the non-destructive detection of ghoore (unripe grape) adulteration in brown sumac, a medicinally and economically valuable spice vulnerable to quality degradation. Samples with adulteration levels of 5%, 20%, 35%, 50%, and 100% were analyzed. Hyperspectral images were acquired and processed using spatial segmentation and Savitzky–Golay filtering to extract informative spectral features. Classification models including Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Artificial Neural Network were employed for both binary (pure vs. adulterated) and six-class (specific adulteration levels) classification. The SVM model achieved the highest accuracy, with 99.00% for binary classification and 94.55% for six-class classification. Key discriminative features identified through RF and XGBoost analysis included phase-related components, fractal dimension, and the area under the curve (AUC) in the 700–900 nm spectral range. The results demonstrate that the integration of HSI and machine learning enables a rapid, non-destructive, and reliable method for detecting sumac adulteration, offering significant potential for food quality assurance applications.
Keywords: Adulteration detection, brown sumac, feature extraction, hyperspectral imaging, machine learning, multi-class classification.

1. Introduction

Sumac, primarily derived from *Rhus coriaria L.* of the *Anacardiaceae* family, is a small deciduous tree widely distributed across the Mediterranean region, North Africa, and the Middle East [1-3]. The plant has attracted increasing attention for its culinary, nutritional, and medicinal significance. Sumac has a long history as a condiment, appetizer, and souring agent, particularly in Mediterranean and Middle Eastern cuisines [4-5]. It is widely used with grilled meats and kebabs, imparting a tangy, citrus-like flavor to foods [3], [5].

Sumac contains a rich array of bioactive compounds, including flavonoids, anthocyanins, phenolic acids, and organic acids [6]. The remarkable volatile compounds present in sumac give it a unique aroma that increases consumer acceptance and its potential use in the food industry. Sumac has been evaluated for a broad range of nutritional and pharmacological activities such as antioxidant, antinociceptive, anti-inflammatory, anti-diabetic, hepatoprotective, cardioprotective, anticancer, anti-infertility, and neuroprotective potentials [7].

Spices like sumac are particularly prone to adulteration for several reasons, including complex global supply chains, high demand, limited supply, and high value. Sumac is usually sold in powder form in markets. The powdered form of sumac is often adulterated with cheaper materials that resemble the color

or appearance of the main ingredient, such as ghoore powder and lemon powder, to increase the volume of the lot [8].

Methods based on chromatographic techniques, such as gas chromatography and high-performance liquid chromatography, are the most widely used to detect adulteration [9]. While accurate and highly sensitive, these techniques are destructive, requiring complex sample pre-treatment, skilled professionals, expensive and sophisticated equipment, and long analysis times [10].

Hyperspectral imaging (HSI) has emerged as a non-destructive, rapid, and high-resolution technique for detecting adulteration in the food industry. By providing both spatial and spectral information, it offers a powerful and environmentally friendly alternative to conventional methods used in this field. HSI has been applied across various domains, including agriculture and food quality assessment, and specifically for the evaluation of powdery materials. This method has proven effective in identifying food adulteration in products such as oil, honey, and meat [11].

For example, Nargesi et al. [12] investigated the capability of HSI to estimate three common adulterants—sea foam powder, chickpea flour, and wheat flour—at concentrations ranging from 5% to 50% in cinnamon powder, achieving classification accuracies above 98% for all adulterants. Malavi et al. [13] studied saffron adulteration with Crocus sativus style at concentrations ranging from 20% to 90%, finding that HSI achieved correct classification rates between 95.6% and 100% in discriminating authentic saffron from plant adulterants. Orrillo et al. [14] explored the potential of near-infrared HSI combined with multivariate analysis to identify black pepper adulterated with papaya seeds. Their classification models, based on principal component analysis (PCA) and soft independent modeling of class analogy (SIMCA), reached 100% accuracy for whole berry samples and sensitivities above 90% for ground samples.

Despite these advances, no study has yet addressed the application of HSI for both identifying adulterant types and quantifying adulteration levels in powdered sumac. The current study aims to develop an HSI-based approach combined with machine learning techniques to provide a rapid, non-destructive method for authenticating brown sumac powder and predicting adulteration levels in samples. An overview of the research workflow is illustrated in Figure 1.

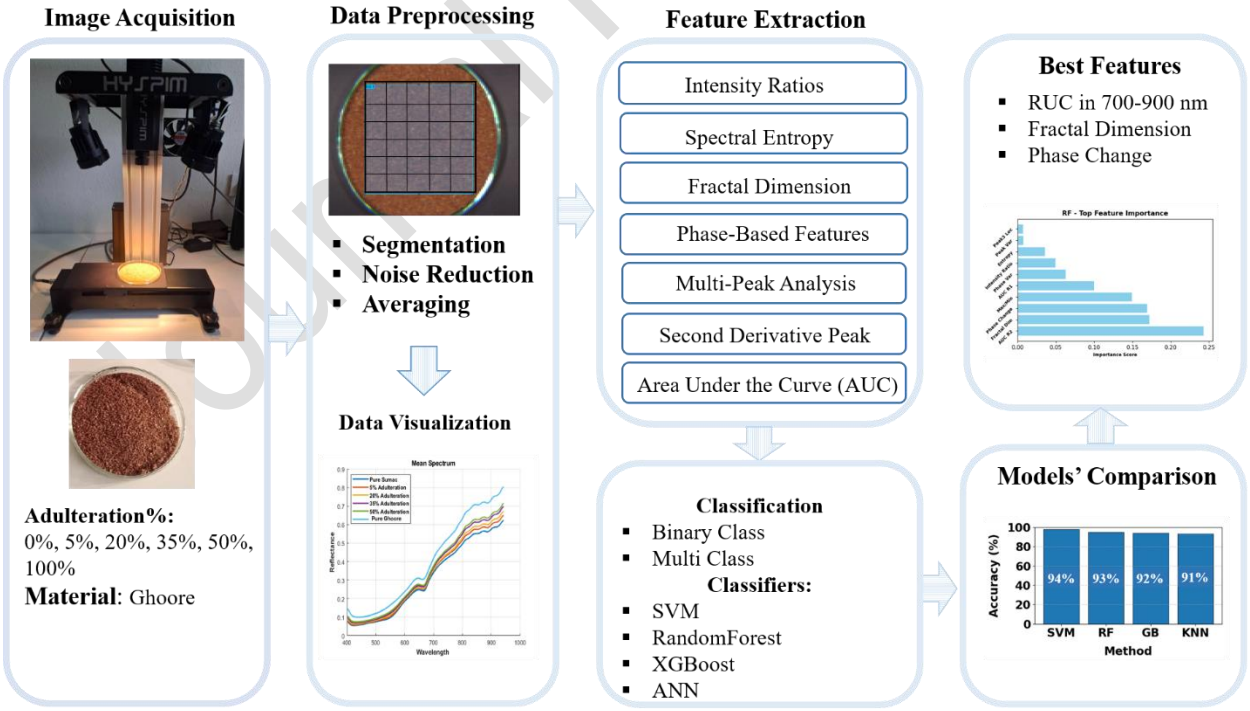


Figure 1. Overview of the research workflow for detecting adulteration in brown sumac using hyperspectral imaging and machine learning algorithms.

2. Materials and Methods

2.1 Sample Preparation

Brown sumac berries and ghoore (unripe grape) samples were obtained from a reliable supplier. After removing impurities and foreign particles, the samples were cleaned and dried in a convection oven at 40°C until they reached an average final moisture content of 8% (wet basis). The dried materials were then ground using an electric grinder (IKA, model A11, Germany) and sieved through a 35-mesh screen. The resulting powders were stored in glass jars, kept in the dark, and maintained at 4°C. Mixtures were prepared by combining 10-gram portions of sumac and ghoore powders in various proportions. The ghoore powder content, used as an index of sumac adulteration, was adjusted to 0%, 5%, 20%, 35%, 50%, and 100%. To account for sample variability, five independent replicates were prepared for each ratio. The pure forms of sumac and ghoore are illustrated in Figure 2.

a) b)

Figure 2. a) Pure sumac. b) Pure ghoore

2.2 Hyperspectral Image Acquisition

A desktop model hyperspectral imaging system (Partoafzar Company, Zanjan, Iran) was used to acquire hyperspectral images (Figure 3). The system consisted of a halogen light source, a line-scan hyperspectral camera covering a wavelength range of 400–950 nm with 993 spectral channels, a motorized linear scanner (20 cm travel length), and dedicated image acquisition software.

Prior to image acquisition, the light source was switched on and allowed to stabilize for approximately 2.5 minutes. The system was then calibrated using white and black reference targets. The scanning length was set to 12 cm for each measurement. Each sample, prepared as described in Section 2.1, was placed in a borosilicate Petri dish with an 8 cm diameter, and hyperspectral images were subsequently captured. The captured hyperspectral data had dimensions of $472 \times 120 \times 993$, where 472×120 represents the spatial resolution (pixels) and 993 represents the spectral resolution (number of wavelength bands).

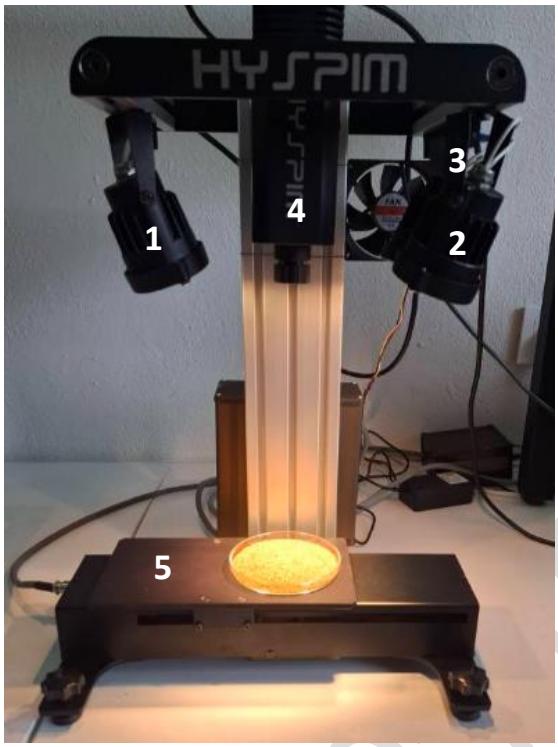


Figure 3. **Desktop hyperspectral imaging system used for sample image acquisition, with components labeled as follows: halogen lamps (1–3), line-scan hyperspectral camera (4), motorized linear scanner (5), and Software.**

2.3 Data Preprocessing

A region of interest (ROI) was spatially selected from each image (Figure 4). To increase dataset size and emphasize localized spectral variations, each ROI was spatially divided into 10 × 10-pixel segments. Each segment was treated as an independent unit for feature extraction, yielding 30 segments per image. Since five images were captured for each of the five samples at every impurity level, a total of 150 image segments (30 segments × 5 images) were generated for each adulteration percentage.

To reduce spectral noise in each 3D image segment, a Savitzky–Golay smoothing filter [15] with a polynomial order of 3 and a window length of 7 was applied. Then, the spectral data of all pixels within each segment were averaged, representing each segment by its mean spectral reflectance. The final data for the extreme classes—pure sumac and pure ghoore—are illustrated in Figure 5 (top). Through this process, a comprehensive dataset of 900 samples (150 segments × 6 adulteration levels) was built, each containing 993 spectral features. The dataset covered six adulteration levels of sumac mixed with ghoore, corresponding to impurity percentages of 0%, 5%, 20%, 35%, 50%, and 100%. Classes 1 through 6 were assigned to these respective impurity levels. Figure 5 (bottom) presents the mean spectral reflectance curves of each class in the dataset.

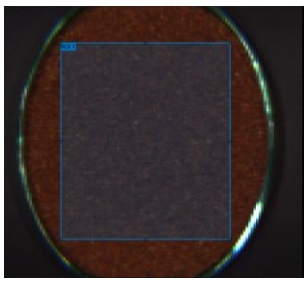


Figure 4. Region of interest (ROI) selected from the hyperspectral image. The ROI was divided into 30 segments of 10 by 10 pixels each for further analysis.

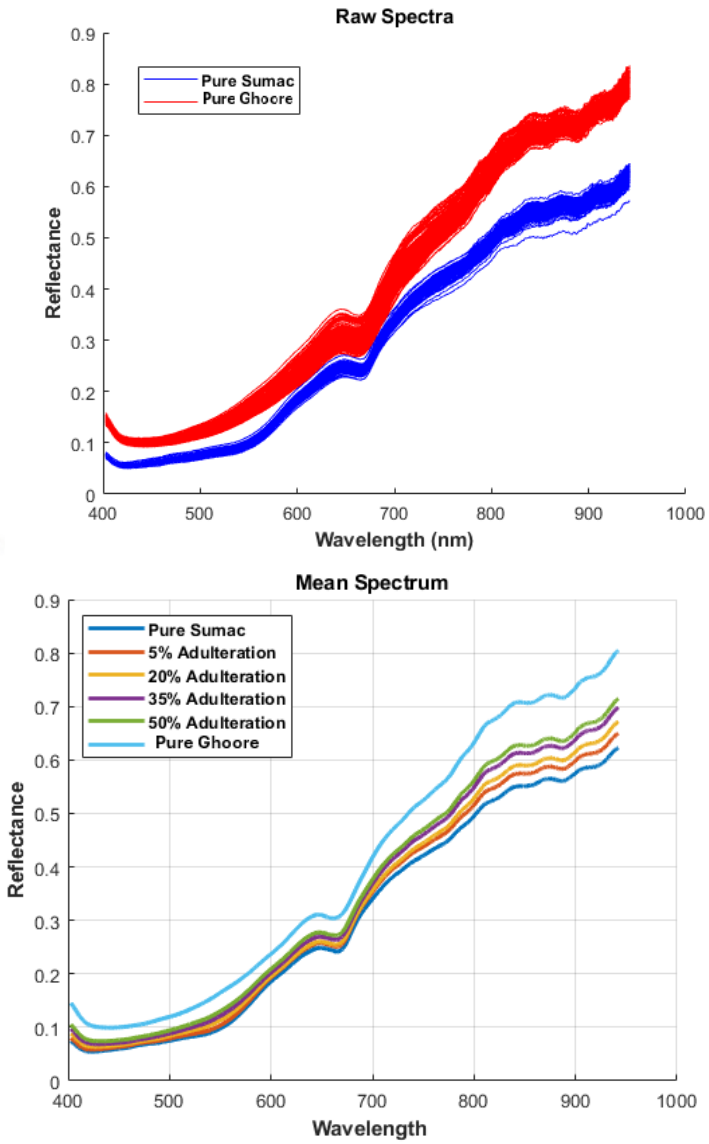


Figure 5. Top) Spectral reflectance of 150 image patches from pure sumac and pure ghoore. Bottom) Mean spectral reflectance for all six classes.

2.4 Feature Extraction

A dataset comprising 900 samples and 993 features was constructed in the previous section. In this section, several types of features are extracted from the reflectance spectra for model training. Because the signal phase plays a crucial role in distinguishing subtle differences between spectral signals, phase-based features were computed using the Hilbert transform, including the mean absolute phase change (*Phase Change*) and the phase standard deviation (*Phase Var*). These features capture local variability in spectral shape and represent overall fluctuations in phase structure, making them effective for identifying fine wavelength shifts.

Additionally, spectral entropy (*Entropy*) was derived by applying the fast Fourier transform (FFT) to the signals and calculating entropy from the resulting power spectrum. Higher entropy values indicate more complex spectra with multiple absorption peaks, whereas lower values correspond to simpler spectra with energy concentrated in fewer regions. This metric aids in distinguishing materials with different absorption characteristics.

Next, the fractal dimension (*Fractal Dim*) was calculated using a simplified box-counting method. This feature quantifies the complexity and self-similarity of spectral curves; in general, chemically complex materials with multiple absorption features exhibit higher fractal dimensions. This parameter assists in identifying irregular reflectance patterns.

Furthermore, the positions of the three main absorption peaks (*Peak loc1*, *Peak loc2*, and *Peak loc3*) and their standard deviation (*Peak Var*) were calculated as additional features. These metrics enable precise measurement of wavelength shifts among samples, which is essential for differentiating materials whose absorption peaks occur at varying wavelengths. The standard deviation further quantifies the variability of absorption patterns across samples.

The area under the curve (AUC) was extracted for two wavelength ranges, 500–700 nm and 700–900 nm (*AUC R1* and *AUC R2*). The AUC represents the total reflected energy within these spectral regions and facilitates differentiation of overall reflectance characteristics among samples, particularly when distinct spectral behaviors are present.

Additionally, the ratio between the average reflectance in the 600–700 nm and 800–900 nm bands (*Intensity Ratio*), as well as the ratio of maximum to minimum reflectance (*Max/Min*), were calculated. These features preserve relative reflectance patterns and remain robust against uniform intensity variations, making them useful for distinguishing samples measured under varying illumination conditions. Finally, peaks in the second-derivative spectra (*Deriv2 Peak*), which are sensitive to changes in spectral curvature, were identified and quantified.

This comprehensive feature set enables reliable discrimination among samples. In total, these thirteen features provide a compact yet informative representation of the spectral data for subsequent analysis. For clarity and brevity in the results and discussion sections, these features are referred to by their abbreviated names (as shown in parentheses) hereafter.

2.5 Classification Methods

Four classification algorithms were applied to the thirteen extracted features from hyperspectral images to detect adulteration in brown sumac samples. The employed methods include Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Network (ANN). Each algorithm was selected for its specific strengths in modeling complex data structures, and their parameters were carefully optimized to achieve the best classification performance.

2.5.1 SVM

SVM is a supervised learning algorithm widely recognized for its classification performance [16-17]. It determines an optimal hyperplane by maximizing the margin between different classes in the feature space. SVM projects input features into a higher-dimensional space using kernel functions, allowing non-linearly separable data to become linearly separable. This ability to model complex, non-linear decision boundaries makes SVM well-suited for detecting subtle spectral variations between pure and adulterated sumac samples. The model constructs a decision function based on support vectors, the critical training samples that define class boundaries, ensuring robust generalization even with limited data.

SVM was implemented using a radial basis function (RBF) kernel to capture non-linear relationships inherent in hyperspectral data. The regularization parameter (C) was set to 10 to balance margin maximization and classification error tolerance, while the gamma parameter was specified using the 'scale' option to adapt kernel coefficients based on feature variance.

2.5.2 RF

RF algorithm is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to enhance classification robustness [18]. Each tree is trained on randomly selected subsets of both features and samples, promoting diversity that reduces variance and mitigates overfitting. This approach makes RF particularly effective for hyperspectral data, as it captures complex interactions among spectral features relevant to adulteration detection. Additionally, RF provides inherent feature importance metrics that help identify the most discriminative spectral characteristics.

In this study, an ensemble of 200 trees was implemented to balance predictive stability and computational efficiency. The maximum tree depth was limited to 15 to avoid over-complex models. The minimum

number of samples required to split an internal node and to form a leaf node were set to 5 and 2, respectively, ensuring appropriate generalization. These hyperparameters were optimized through preliminary validation experiments.

2.5.3 XGBoost

XGBoost is an efficient gradient boosting framework that sequentially constructs an ensemble of weak learners, with each iteration correcting the residual errors of the previous ones [19]. It incorporates regularization terms in its objective function to mitigate overfitting and enhance generalization performance. XGBoost is particularly suited for hyperspectral data analysis due to its ability to manage heterogeneous features, missing values, and nonlinear interactions among spectral variables.

In this study, XGBoost was trained for 200 boosting rounds with a learning rate of 0.05 to ensure stable convergence. The maximum tree depth was limited to 10, while both sample and feature subsampling ratios were set to 0.8 to introduce stochasticity and improve model robustness. This configuration enables effective learning of complex spectral patterns while maintaining strong resistance to overfitting.

2.5.4 ANN

ANN is a multi-layer perceptron [20] designed to learn hierarchical feature representations through nonlinear transformations. Lower layers captured fundamental spectral patterns, which were progressively combined in higher **layers** to construct more complex representations. This hierarchical learning approach is particularly advantageous for hyperspectral data, where adulteration signatures often manifest as subtle, nonlinear interactions among spectral features across wavelengths. The ANN's distributed representation and capability for deep feature extraction allow it to model intricate relationships beyond the capacity of conventional algorithms.

The network architecture comprised three hidden layers with 128, 64, and 32 neurons, respectively, forming a tapered structure. The ReLU activation function was used to introduce nonlinearity, and the Adam optimizer adaptively adjusted learning rates during training. L2 regularization (alpha = 0.001) was applied to reduce overfitting, and early stopping with a patience of 20 iterations was employed based on validation performance. The initial learning rate was set to 0.001 with adaptive adjustment throughout training. These hyperparameters were tuned to optimize performance in the six-class classification task.

2.6 Experimental Setup

To ensure a rigorous and unbiased evaluation, a stratified train–validation–test split of 80%, 10%, and 10% was employed. This approach maintained consistent class distribution across all subsets. Hyper parameter tuning was conducted using an exhaustive grid search, and model performance was monitored on the validation set to prevent overfitting.

The finalized models were subsequently evaluated on an independent test set. For the multi-class classification task, model performance was quantified using Accuracy, Precision, Recall, and F1-score. These metrics were computed using macro-averaging, ensuring equal contribution from each class. The metrics were calculated based on the confusion matrix elements—true positives (TP_i), false positives (FP_i), and false negatives (FN_i) for each class i—as defined in Equations (1)–(5); where C denotes the total number of classes.

To ensure statistical reliability and account for

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} T\,P_i}{\sum_{i=1}^{C} (T\,P_i + FP_i + FN_i)} \qquad \text{Eq. (1)}$$

$$\text{Precision} = \frac{1}{C} \sum_{i=1}^{C} \frac{TP_i}{TP_i + FP_i} \qquad \text{Eq. (2)}$$

$$\text{Recall} = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FN_i} \quad \text{Eq. (3)}$$

$$\text{F1-score} = \frac{1}{C}\sum_{i=1}^{C}\left(2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}\right) \quad \text{Eq. (4)}$$

3. Results & Discussion

This investigation employed thirteen spectral features in conjunction with four distinct classification algorithms to differentiate sumac samples across varying degrees of adulteration. The analytical framework incorporated both binary and multi-class classification strategies, reflecting the practical requirements of industrial food authentication systems.

The binary classification task represents a critical preliminary objective aligned with industrial quality control priorities. In real-world food safety applications, the primary concern is often the rapid detection and discrimination between genuine, unadulterated products and any compromised material, regardless of adulterant concentration. This binary distinction serves as the fundamental step in establishing product authenticity before more granular adulteration level characterization.

The binary classification framework was structured as follows: pure sumac samples were designated as Class 1, while all adulterated sumac specimens—regardless of adulteration concentration—were collectively assigned to Class 2. To ensure balanced class representation and mitigate potential bias inherent in imbalanced datasets, an equivalent number of samples from the adulterated specimens in Class 2 was randomly selected to match the total count of pure sumac category (Class 1). This balanced approach is particularly important for obtaining reliable performance metrics, as classification algorithms typically exhibit biased accuracy measurements when trained on imbalanced datasets. Performance metrics of various classifiers for distinguishing pure from adulterated sumac (binary classification) are compared in Table 1.

Following the foundational binary classification analysis, a more granular six-class classification scheme was implemented to distinguish sumac samples across distinct adulteration concentration levels. This multi-class approach extends the methodology beyond simple authenticity verification to enable quantitative characterization of adulteration severity, which provides additional information for quality control and regulatory compliance purposes.

The six-class classification distinguished the following adulteration levels: 0% (pure sumac), 5%, 20%, 35%, 50%, and 100% adulterant mixed with ghoore. This graduated scale allows for more nuanced assessment of sample integrity and can inform decisions regarding product disposition, remediation, or market classification. Table 2 presents the corresponding performance metrics for all classifiers evaluated in this multi-class scenario.

The classification models exhibited strong performance in both binary and multi-class tasks. The best performing method was SVM, achieving accuracies of 99.00% and 94.55% for binary and six-class classification, respectively. RF and XGBoost also demonstrated high performance, achieving accuracies of 97.33% and 98.00% for binary classification and 93.66% and 92.48% for six-class classification, respectively. Among the classifiers, ANN showed the weakest performance, with accuracies of 89.00% for binary classification and 91.89% for six-class classification.

Furthermore, confusion matrices were computed to evaluate the performance of each method in the six-class classification task, as shown in Figure 6. These matrices illustrate classification accuracy and misclassification patterns, providing detailed insights into the models' capabilities to discriminate pure sumac from samples with varying adulteration degrees. The examination of confusion matrices across the evaluated models revealed a distinctive error distribution pattern, with the majority of classification errors occurring between adjacent categorical classes. These misclassifications at class boundaries indicates that

the models encountered difficulties in differentiating between samples that were either pure or only slightly adulterated, a challenge that can be attributed to their remarkably analogous spectral signatures or chemical compositions. However, the classification performance improved substantially when distinguishing samples with more pronounced levels of adulteration. The models demonstrated robust discriminative capabilities for higher adulteration concentrations, exhibiting strong confidence and accuracy when classifying genuinely pure sumac specimens alongside samples containing 15% adulterant levels or beyond.

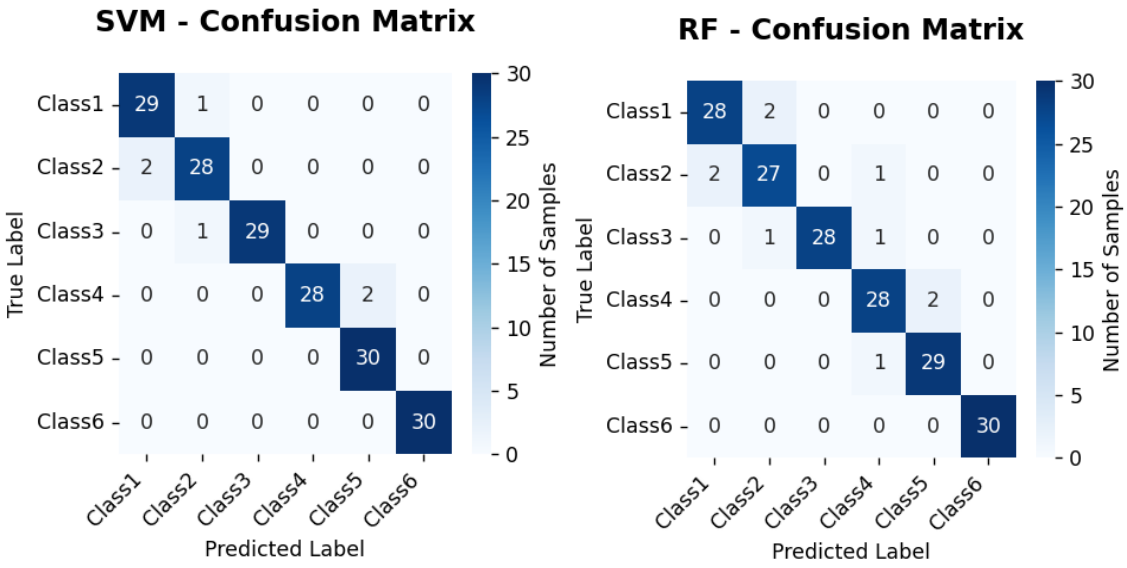
The overall findings suggest that although the models exhibited a degree of uncertainty when attempting to differentiate between spectrally similar classes, their general classification reliability remains sufficiently reliable for real-world industrial applications involving adulteration detection. This performance profile is particularly suitable for food quality control systems, where accurate identification of substantial contamination is more critical than distinguishing minimal or negligible adulteration.

Table 1: Performance metrics of binary classification of pure sumac and adulterated sumac.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	99.00±1.49	99.01±1.48	99.00±1.49	99.00±1.49
RF	97.33±1.49	97.38±1.50	97.33±1.49	98.00±1.49
XGBoost	98.00±0.75	98.04±0.77	98.00±0.75	98.00±0.75
ANN	96.67±1.90	96.87 ±1.60	96.66±1.90	96.66±1.96

Table 2: Performance metrics of different classifiers for six-class classification.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	94.55±1.29	94.82±1.24	94.55±1.29	94.55±1.29
RF	93.66±1.29	93.86±1.24	93.66±1.29	93.68±1.31
XGBoost	92.48±1.39	93.11±1.21	92.88±1.29	92.89±1.27
ANN	91.89±0.97	92.15±1.05	91.89±0.97	91.87±0.96



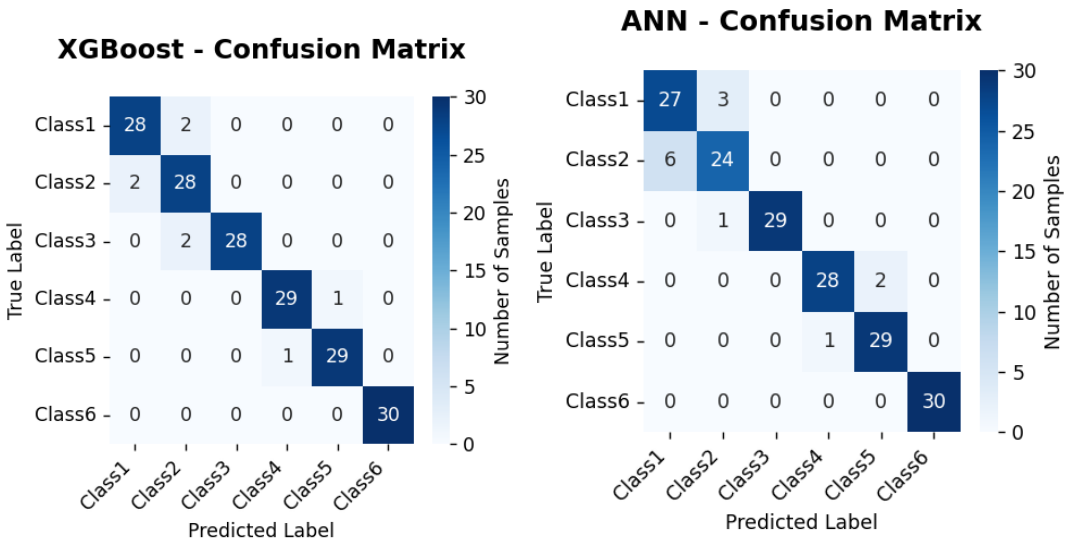


Figure 6. Confusion matrix illustrating the performance of six-class classification using SVM, RF, XGBoost, and ANN classifiers.

To better understand the role of each predictor in the classification outcomes, we conducted a feature importance evaluation leveraging two ensemble tree-based methods: RF and XGBoost. These algorithms are particularly effective for this purpose, as they inherently assess the contribution of individual features to enhancing decision tree accuracy during model training.

For the RF approach, feature importance was calculated based on the mean reduction in node impurity—using Gini or entropy metrics—whenever a feature was selected for data splitting. Features that consistently lowered impurity across the tree ensemble received higher importance rankings. In contrast, XGBoost derives importance from "gain," which quantifies the average enhancement in the loss function (i.e., error minimization) achieved through splits involving a given feature throughout the sequential boosting iterations. These scores were then aggregated and scaled for straightforward cross-feature comparisons.

The feature importance visualizations (Figure 7) indicated strong alignment between the two models in identifying core predictors, albeit with minor variations in their ordering. In RF, *Fractal Dim*, *AUC R2*, and *Phase Change* ranked as the top three influencers, with *Max/Min* ratio and *AUC R1* following suit. These attributes appear to encapsulate intricate signal patterns that effectively distinguish between classes. Conversely, metrics like *Entropy*, *Peak Var*, and *Peak3 Loc* exhibited notably lower scores, implying limited impact on the model's predictive decisions.

XGBoost displayed a comparable hierarchy, positioning *AUC R2* as the paramount feature, trailed by *Phase Change* and *Fractal Dim*—reinforcing their cross-model significance. This model's focus on a concise set of high-impact variables stems from its gradient-based, iterative refinement, resulting in more pronounced emphasis on leading features. Metrics such as *Entropy*, *Phase Variance*, and various *Peak3 Loc* indicators again showed negligible influence, highlighting their peripheral role in discrimination.

From a spectrochemical and structural viewpoint, these leading features align with tangible alterations in the optical and morphological properties distinguishing pure from adulterated sumac. *Fractal Dim* measures the irregularity and intricacy of the spectral profile, capturing fluctuations in reflectance that arise from changes in particle texture, shape, and light scattering—effects exacerbated by adulterant incorporation. Its elevated ranking signals that adulteration disrupts spectral roughness in detectable ways, often due to inconsistent particle blending and composition heterogeneity.

AUC R2, which integrates reflectance over the 700–900 nm range, aggregates spectral energy in a band sensitive to pigments like anthocyanins, tannins, and polyphenols prevalent in spices. Adulterants, differing in molecular makeup, modify absorption profiles here, explaining *AUC R2*'s consistent dominance in both models as a marker of chemical adulteration. Likewise, *AUC R1*, covering the 500–

700 nm visible spectrum, detects color shifts tied to pigment content, aiding differentiation of genuine sumac from contaminated variants.

Phase Change, derived from phase shifts in Fourier-transformed spectra, detects subtle frequency alterations linked to microstructural disruptions or scattering modifications induced by adulterants. Its prominence across models underscores its utility in pinpointing purity via consistent spectral phase anomalies. The *Max/Min* ratio describes the overall contrast or dynamic range of the reflectance spectrum. Adulterated samples often exhibit diminished contrast as a result of increased scattering and blending effects, whereas pure samples retain sharper spectral peaks. Hence, this ratio provides an additional indicator of spectral dilution or compositional blending. The *Intensity Ratio*, while of moderate significance, bolsters this by highlighting inter-wavelength absorption disparities reflective of compositional variances.

In opposition, features like *Entropy*, *Peak Var*, and *Peak Loc* metrics received persistently low rankings, as they primarily gauge stochasticity or pinpoint spectral events that are insufficiently attuned to adulteration's overarching effects. This pattern suggests adulteration impacts holistic spectral architecture—encompassing form, integrated intensity, and phase—more than discrete peak anomalies.

Both RF and XGBoost concur on the primacy of *AUC R2*, *Phase Change*, and *Fractal Dim*, affirming their reliability in extracting discriminatory signals and bolstering the analysis's validity. The marginal scores for peak- and entropy-derived features point to potential redundancy, favoring more holistic descriptors for model efficiency. Collectively, this evaluation not only pinpoints pivotal predictors but also bridges machine learning outputs to the physicochemical underpinnings of the samples, elucidating how hyperspectral signatures reveal adulteration's optical footprints. Such insights pave the way for targeted feature engineering in subsequent models and streamlined adulteration quantification.

In essence, fusing hyperspectral imaging with these machine learning techniques yields a dependable, swift, and contactless strategy for sumac adulteration identification, holding substantial potential as a screening mechanism in food safety protocols.

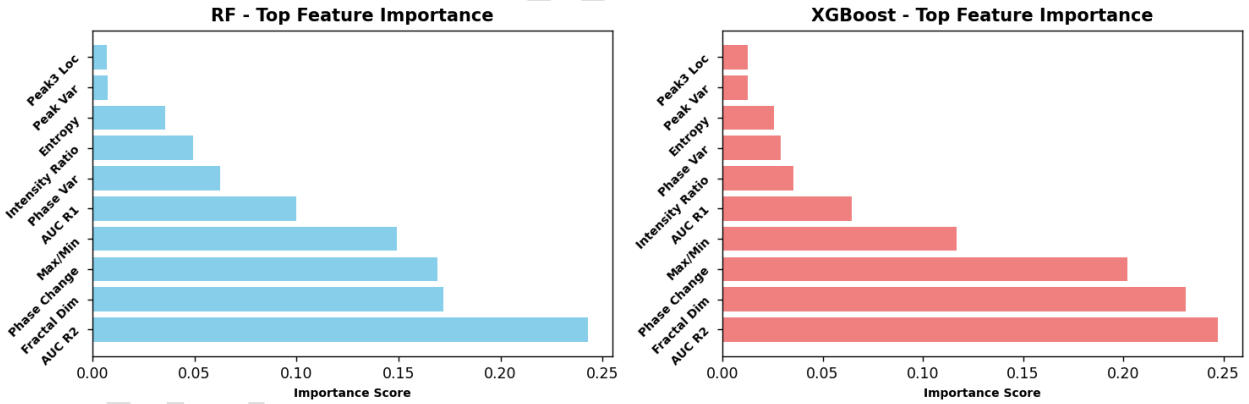


Figure 7. Feature importance chart illustrating the contribution of each feature in the classification of sumac samples with different levels of adulteration.

4. Conclusion

This study explored the application of hyperspectral imaging (HSI) for detecting adulteration in brown sumac powder, a spice prone to fraudulent practices such as dilution with extraneous materials, as commonly observed in global spice trade. A comprehensive dataset of HSI images was acquired and subjected to preprocessing, yielding thirteen key features that capture spectral and spatial variations indicative of adulteration. These features were subsequently classified using four machine learning algorithms—SVM, RF, XGBoost, and ANN—for both binary (pure vs. adulterated) and six-class (specific adulteration levels from 0% to 50%) tasks. Binary classification achieved accuracies ranging from 96.67% to 99.00%, while six-class classification ranged from 91.89% to 94.55%, with SVM

demonstrating superior performance across both scenarios, consistent with its efficacy in handling high-dimensional spectral data for food authentication.

Feature importance analysis revealed that AUC R2, Fractal Dim, and Phase Change metrics were the most influential in discriminating sumac samples at varying adulteration levels, enabling reliable detection even at low concentrations (e.g., 5%), and thereby addressing challenges in non-destructive quality assessment.

This HSI-machine learning framework offers a rapid, non-invasive alternative to traditional methods, aligning with broader trends in spectroscopic techniques for spice adulteration detection, including black pepper, turmeric, and saffron, where similar approaches have yielded high accuracies through chemometric integration. By mitigating economic and health risks associated with adulteration—such as substitution with inferior flours or hazardous dyes—the proposed method underscores the transformative potential of HSI coupled with advanced analytics for enhancing food safety, regulatory compliance, and supply chain integrity in the spice industry.

Funding

No funding was received to assist with the preparation of this manuscript.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used DeepSeek, OpenAI's GPT-4o, and Perplexity (Free version) to refine the manuscript's language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

[1]. Batiha, G. E.-S., Ogunyemi, O. M., Shaheen, H. M., Kaczmarek, M. J., Martorell, M., Ozdemir, M., ... & Sur, D. (2022). Rhus coriaria* L. (Sumac), a versatile and resourceful food spice with cornucopia of polyphenols. Molecules, 27(16), 5179. https://doi.org/10.3390/molecules27165179

[2]. Perrone, A., Yousefi, S., Basile, B., Papini, A., Nocera, P., & Covacci, V. (2022). Phytochemical, antioxidant, anti-microbial, and pharmaceutical properties of sumac (*Rhus coriaria* L.) and its genetic diversity. Horticulturae, 8(12), 1168. https://doi.org/10.3390/horticulturae8121168

[3]. Singletary, K. W. (2023). Sumac potential health benefits. Nutrition Today, 58(2), 77–83. https://doi.org/10.1097/NT.0000000000000595

[4]. Elagbar, Z. A., Shakya, A. K., Barhoumi, L. M., & Al-Jaber, H. I. (2020). Phytochemical diversity and pharmacological properties of Rhus coriaria. Chemistry & Biodiversity, 17(4), e1900561. https://doi.org/10.1002/cbdv.201900561

[5]. Shahrajabian, M. H., & Sun, W. (2022). Using sumac (*Rhus coriaria* L.), as a miraculous spice with outstanding pharmacological activities. Notulae Scientia Biologicae, 14(1), 11118. https://doi.org/10.15835/nsb14111118

[6]. Reidel, R. V. B., Cioni, P. L., Majo, L., Pistelli, L., & Pistelli, L. (2017). Evolution of volatile emission in Rhus coriaria organs during different stages of growth and evaluation of the essential oil composition. Chemistry & Biodiversity, 14(11), e1700270. https://doi.org/10.1002/cbdv.201700270

[7]. Zannou, O., Oussou, K. F., Chabi, I. B., Awad, N. M., Fattah, A., Pashazadeh, H., ... & Koca, I. (2025). Phytochemical and nutritional properties of sumac (*Rhus coriaria*): A potential ingredient for developing functional foods. Journal of Future Foods, 5(1), 21–35. https://doi.org/10.1016/j.jfutfo.2024.01.002

[8]. Momtaz, M., Bubli, S. Y., & Khan, M. S. (2023). Mechanisms and health aspects of food adulteration: A comprehensive review. Foods, 12(1), 199. https://doi.org/10.3390/foods12010199

[9]. Khan, M. H., Saleem, Z., Ahmad, M., Sohaib, M., & Ayaz, H. (2020). Hyperspectral imaging for color adulteration detection in red chili. Applied Sciences, 10(17), 5955. https://doi.org/10.3390/app10175955

[10]. Temiz, H. T., & Ulaş, B. (2021). A review of recent studies employing hyperspectral imaging for the determination of food adulteration. Photochem, 1(2), 125–146. https://doi.org/10.3390/photochem1020008

[11]. Medina-García, M., Amigo, J. M., Martínez-Domingo, M. A., & Sánchez, A. (2025). Strategies for analysing hyperspectral imaging data for food quality and safety issues – A critical review of the last 5 years. Microchemical Journal, 214, 113994. https://doi.org/10.1016/j.microc.2025.113994
[12]. Nargesi, M. H., Amiriparian, J., Bagherpour, H., & Kheiralipour, K. (2024). Detection of different adulteration in cinnamon powder using hyperspectral imaging and artificial neural network method. Results in Chemistry, 9, 101644. https://doi.org/10.1016/j.rechem.2024.101644
[13]. Malavi, D., Nikkhah, A., Alighaleh, P., & Mousavi, S. A. (2024). Detection of saffron adulteration with Crocus sativus style using NIR-hyperspectral imaging and chemometrics. Food Control, 157, 110189. https://doi.org/10.1016/j.foodcont.2023.110189
[14]. Orrillo, I., Cruz-Tirado, J. P., Cardenas, A., Carnero, A., & Barbin, D. F. (2019). Hyperspectral imaging as a powerful tool for identification of papaya seeds in black pepper. Food Control, 101, 45–52. https://doi.org/10.1016/j.foodcont.2019.02.036
[15]. Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry, 36(8), 1627–1639. https://doi.org/10.1021/ac60214a047
[16]. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
[17]. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7
[18]. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
[19]. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785
[20]. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536. https://doi.org/10.1038/323533a0