

Article

NIR and MIR Spectroscopy for the Detection of Adulteration of Smoking Products

Zeb Akhtar ^{1,2} , Ihtesham ur Rehman ² , Cédric Delporte ^{3,4} , Erwin Adams ²  and Eric Deconinck ^{1,3,*} 

- ¹ Scientific Direction Chemical and Physical Health Risks, Service of Medicines and Health Products, Sciensano, Rue Juliette Wytsmanstraat 14, 1050 Brussels, Belgium; zeb.akhtar@sciensano.be
- ² Department of Pharmaceutical and Pharmacological Sciences, Faculty of Pharmaceutical Sciences, Pharmaceutical Analysis, KU Leuven, Herestraat 49, O&N2, PB 923, 3000 Leuven, Belgium; ihtesham.rehman@kuleuven.be (I.u.R.); erwin.adams@kuleuven.be (E.A.)
- ³ Pharmacognosy, Bioanalysis and Drug Discovery Unit, RD3, Faculty of Pharmacy, Université Libre de Bruxelles (ULB), Campus Plaine, Boulevard du Triomphe, CP 205/5, 1050 Brussels, Belgium; cedric.delporte@ulb.be
- ⁴ Analytical Platform, Faculty of Pharmacy, Université Libre de Bruxelles (ULB), Campus Plaine, Boulevard du Triomphe, CP 205/5, 1050 Brussels, Belgium
- * Correspondence: eric.deconinck@sciensano.be

Abstract

This study explores the application of Mid-Infrared (MIR) and Near-Infrared (NIR) spectroscopy combined with various multivariate calibration techniques to detect the presence of cannabis in tobacco samples and tobacco in herbal smoking products. Both MIR and NIR spectra were recorded for self-prepared samples, followed by data exploration using Principal Component Analysis (PCA) and Hierarchical Clustering Analysis (HCA), and the calculation of binary classification models with Soft Independent Modelling of Class Analogy (SIMCA) and Partial Least Squares-Discriminant Analysis (PLS-DA). PCA demonstrated a clear differentiation between tobacco samples containing and not containing cannabis. On the other hand, based on PCA, only NIR was able to distinguish herbal smoking products adulterated and not adulterated with tobacco. HCA further clarified these results by revealing distinct clusters within the data. Modelling results indicated that MIR and NIR spectroscopy, particularly when paired with preprocessing techniques like Standard Normal Variate (SNV) and autoscaling, demonstrated high classification accuracy in SIMCA and PLS-DA, achieving correct classification rates of 90% to 100% for external test sets. Comparison of MIR and NIR revealed that NIR spectroscopy resulted in slightly more accurate models for the screening of tobacco samples for cannabis and herbal smoking products for tobacco. The developed approach could be useful for the initial screening of tobacco samples for cannabis, e.g., in a night life setting by law enforcement, but also for inspectors visiting shops selling tobacco and/or herbal smoking products.

Keywords: cannabis; herbal smoking products; tobacco; (N)IR spectroscopy; adulteration; multivariate calibration



Received: 30 July 2025
Revised: 13 October 2025
Accepted: 14 October 2025
Published: 16 October 2025

Citation: Akhtar, Z.; Rehman, I.u.; Delporte, C.; Adams, E.; Deconinck, E. NIR and MIR Spectroscopy for the Detection of Adulteration of Smoking Products. *Chemosensors* **2025**, *13*, 370. <https://doi.org/10.3390/chemosensors13100370>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Smoking consists of inhaling and exhaling the smoke produced by burning plant material. The most common is the smoking of tobacco products in the form of cigarettes, cigars, and pipes [1]. As is generally known, smoking tobacco is one of the major public health challenges in the world and is associated with significantly increased risks for numerous health issues, including cardiovascular diseases, various cancers, diabetes, and

strokes. Next to tobacco, other products are also smoked [2–5]. An example is herbal smoking products, also marketed as tobacco-free smoking products. These herbal products are positioned in the market as a healthier or safer alternative to tobacco. However, there is not a lot of solid evidence to back up these claims. In fact, there are concerns that people might misunderstand these herbs as being completely safe, even though they might still pose health risks. Moreover, although nicotine is the compound in tobacco causing habituation and addiction, the most important compounds generating the health concerns are from the burning process itself, which is similar to the one of the herbal smoking products [6–8].

Through definition, herbal smoking products may not contain tobacco, since otherwise they would fall under the tobacco products regulation [9]. Nevertheless, some studies show that herbal smoking products are marketed as combinations of herbs and tobacco, misleading consumers and causing a possible nicotine addiction. For instance, a study in tobacco control found that Asian herbal-tobacco cigarettes often contain both herbs and tobacco, creating a false impression of reduced harm. Another study revealed that Chinese “herbal” cigarettes are just as carcinogenic and addictive as regular cigarettes due to their tobacco content [10,11].

Another product that is often smoked is cannabis, often referred to as marijuana. It is the most widely used illicit drug in both the United States of America and in Europe. The health effects of regular cannabis use are a subject of ongoing debate, but they can include several potential risks. These risks range from reduced educational achievement, particularly among teenagers due to its impact on cognitive development, and among young adults due to potential motivational and cognitive impairments and increased likelihood of driving-related injuries, to elevated respiratory issues. Long-term health concerns associated with cannabis use may include cancer, chronic obstructive pulmonary disease (COPD), heart disease, addiction, and an increased risk of psychosis, particularly in susceptible individuals [12,13].

Although it is generally known that cannabis users often also consume tobacco, there is also the phenomenon where cannabis and tobacco are mixed, e.g., in so-called “spliffs” (joints that contain both cannabis and tobacco). This practice, sometimes called “mulling”, is widespread in various regions, including some European countries where spliffs are a popular method for consuming marijuana. These practices not only combine the substances but also expose users to the risks associated with both, such as nicotine addiction and respiratory issues. Mixing cannabis and tobacco is often motivated by reasons such as cost-saving (for users or sellers (fraud)), smoother intake, and improved combustion properties. Qualitative studies have highlighted that spliffs can contain substantial amounts of tobacco, and this presence may not always be known to the user, unknowingly increasing their risk of nicotine intake and dependence. Given that nicotine is generally considered more addictive than Delta-9-tetrahydrocannabinol, this hidden exposure may pose a greater risk of addiction than cannabis alone [14–16].

Several countries already have legislation allowing the use of marijuana and even more have a so-called tolerance policy, though cannabis is still considered an illicit drug by a lot of countries and by the international legislation, like, for example, the United Nations Convention against illicit traffic in narcotic drugs and psychotropic substances [17]. Despite the different legislations, users have the right to know if tobacco is present in the cannabis products they purchase, and otherwise law enforcers should be able to recognize the presence of cannabis in tobacco products, especially if they consider cannabis as an illicit drug [18].

This paper concerns both issues, i.e., the presence of tobacco in herbal smoking products and the presence of cannabis in tobacco. In fact, there is a need for fast screening

of products for the presence of adulterants, tobacco, and/or cannabis, both for inspectors to decide which products to seize, e.g., during inspection of shops or suspicious wholesalers, and for law enforcement, e.g., to check products at festivals and night life settings. In this context, spectroscopic techniques, Mid- and Near-Infrared spectroscopy (MIR and NIR) were evaluated for their performance to detect tobacco in herbal smoking products as well as to detect cannabis in tobacco. Spectra were recorded with both techniques for different mixtures, followed by the creation of binary classification models, using classical multivariate calibration techniques. The created models were evaluated and validated for their predictive performance and the performance of both modelling techniques for the two spectroscopic techniques was compared.

2. Materials and Methods

2.1. Sample Preparation

For this study, three types of samples including cannabis, tobacco, and herbal smoking products were used. Twelve cannabis samples were confiscated by Belgian authorities at festivals and night life events. These samples were previously analyzed in the laboratory for their content of tetrahydrocannabinol (THC) and tested negative for nicotine and therefore were considered as not adulterated with tobacco. A representative selection of 12 tobacco products was chosen and purchased from the market. The selection was based on the feedback of vendors, as to what were their most popular products. All samples were purchased from vendors in the city of Leuven, Belgium. Seven herbal smoking products were either seized by the authorities, in the context of market surveillance, or donated by distributors. These samples were analyzed previously and found negative for “tobacco” (nicotine content) and “cannabis” (THC content).

2.1.1. Mixing

All samples (tobacco, cannabis, and herbal smoking products) were initially homogenized by manual grinding using a mortar and pestle. The ground materials were then transferred into labelled vials and stored at controlled temperature (22 ± 2 °C), relative humidity ($50 \pm 5\%$), under dark conditions to avoid degradation, and in sealed containers to minimize moisture exchange prior to further processing. To create representative mixture sets, Sample A was prepared by thoroughly mixing all individual tobacco samples, Sample B was formed by combining all cannabis samples, and Sample C consisted of a uniform mixture of all herbal smoking products.

2.1.2. Design of Experiment

To simulate real-world variability and complexity, each tobacco sample was individually mixed with various cannabis samples. Then, Sample A (tobacco mix) was spiked with Sample B (cannabis mix) at concentrations ranging from 1% to 50%. The total number of samples obtained was 110. Each herbal smoking sample was mixed with various tobacco samples. Sample C (herbal mix) was then spiked with Sample A (tobacco mix) at the same concentration gradient (1–50%). The total number of samples obtained was 56. Table 1 summarizes the process. Next to the mixtures, all individual tobacco samples and herbal smoking samples were also analyzed, featuring as negative samples. A detailed overview of the sample concentration levels is provided in Tables S1–S3 (Supplementary Materials). Due to limited quantities, not all concentrations were prepared for all samples, resulting in a lower number of mixtures prepared than would have been the case in a full design. For each group, the limit of detection (LOD) was established at 1% (m/m), which corresponds to the lowest concentration tested within the 1–50% range. The developed model did not exhibit systematic errors at low concentrations, and most samples with 1%

adulteration were correctly identified. Therefore, this concentration was considered the LOD. As this study did not include quantitative analysis, no limit of quantification (LOQ) was determined or reported.

Table 1. Summary of the sample preparation and mixing process.

Sample Type	Mixing Process	Number of Samples	Concentration Range
Tobacco–Cannabis Individual Mix	Each tobacco sample mixed with the different cannabis samples, respectively	102	1–50%
Tobacco sample set (Sample A)	All tobacco samples mixed together	1 (not included in data set)	1–50%
Cannabis sample set (Sample B)	All cannabis samples mixed together	1 (not included in data set)	1–50%
Tobacco–Cannabis Spiked Mix	Sample A spiked with Sample B	8	1–50%
Herbal–Tobacco Individual Mix	Each herbal smoking sample mixed with the different tobacco samples, respectively	48	1–50%
Herbal Smoking sample set (Sample C)	All herbal smoking samples mixed together	1 (not included in data set)	1–50%
Herbal–Tobacco Spiked Mix	Sample C spiked with Sample A	8	1–50%

2.2. Data Acquisition

2.2.1. Mid-Infrared (MIR)

MIR spectra were collected using a Nicolet iS10 FTIR spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), equipped with a Smart iTR accessory and a deuterated triglycine sulphate (DTGS) detector. The Smart iTR accessory features a single-bounce diamond crystal, which was calibrated weekly with a polystyrene film standard. The IR spectra were recorded over a range from 400 to 3400 cm^{-1} , with each spectrum consisting of 32 scans at a resolution of 4 cm^{-1} . The data were processed with OMNIC software version 8.3 (ThermoFisher Scientific). After collecting the data, the diamond crystal was cleaned meticulously with a methanol-soaked tissue followed by air drying. A blank measurement was performed before each analysis to check for contamination and carry-over. The contamination guidelines from the European Directorate for the Quality of Medicines and HealthCare (EDQM, 2007) were followed [19]. To maintain accuracy, background spectra of ambient air were recorded every hour. At the end, a data matrix with dimensions of 110×6949 was obtained for tobacco samples spiked with cannabis, and 56×6949 was obtained for herbal samples spiked with tobacco, where 110 and 56 is the total number of tobacco and herbal samples, respectively, and 6949 is the number of included wavenumbers for chemometric analysis. Triplicate spectral acquisitions were performed for each sample.

For sample recording, a small amount was deposited directly onto the crystal, followed by the application of pressure to ensure complete homogenic coverage of the crystal surface by the sample. For illustration, Figure 1 shows the MIR spectra of the cannabis mix sample, the tobacco mix sample, and the herbal mix sample.

In the MIR spectra of the herbal, cannabis, and tobacco samples, several characteristic absorption bands were observed. The region from 400 to 1500 cm^{-1} , known as the fingerprint region, is highly complex and holds key vibrational features that are useful for distinguishing different plant-based materials. Specifically, 600–900 cm^{-1} corresponds to out-of-plane C–H bending, typically associated with aromatic rings, which are common in phenolic compounds [20]. Moving to 1000–1300 cm^{-1} , C–O stretching vibrations were observed, which are characteristic of alcohols, ethers, and esters—functional groups commonly found in secondary metabolites like terpenes and flavonoids [21]. The region between 1800 and 2100 cm^{-1} suggests the presence of triple bonds, particularly $\text{C}\equiv\text{C}$ or $\text{C}\equiv\text{N}$ [20,22]. Lastly, the 3100–3400 cm^{-1} region indicates the presence of phenolic OH groups or residual moisture, common in plant materials, with the broadness of the peak suggesting phenolic compounds or water [20].

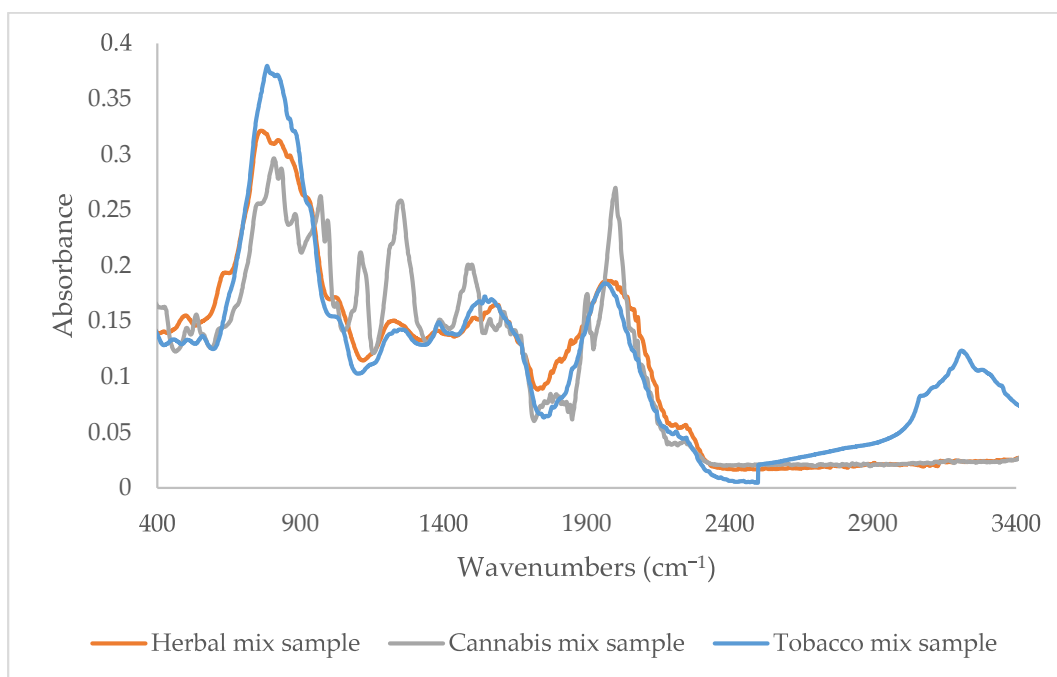


Figure 1. MIR spectra of cannabis mix sample, tobacco mix sample, and herbal mix sample.

2.2.2. Near-Infrared (NIR)

For NIR analysis, all samples were scanned using a Frontier MIR/NIR spectrometer (PerkinElmer, Waltham, MA, USA). The NIR spectra were collected in reflectance mode with an 8 cm^{-1} resolution over the range of 700 to 2300 nm. Each spectrum was generated by averaging 16 scans. Background spectra were recorded with a diffuse reflector from PerkinElmer between individual samples. Background subtraction and arithmetic corrections were applied to account for any background effects. At the end, a data matrix with dimensions of 110×3001 was obtained for tobacco samples spiked with cannabis, and 56×3001 was obtained for herbal samples spiked with tobacco, where 110 and 56 are the total number of tobacco and herbal samples, respectively, and 3001 is the number of included wavenumbers for chemometric analysis. Triplicate spectral acquisitions were performed for each sample.

For this analysis, each sample was split into three glass vials. The plant material in each vial was kept at a height of approximately 1.5–2 cm to optimize the analysis in diffuse reflectance mode. This mode allows the instrument to capture the diffusely reflected light from the sample and direct it to the detector [23]. Figure 2 shows the obtained NIR spectra of the cannabis mix sample, the tobacco mix sample, and the herbal mix sample. In the NIR spectra, which primarily represent overtones and combination bands, a broad absorption near 1400 nm in the spectrum of the herbal mix sample was assigned to the first overtone of O–H stretching [24], and the distinct feature around 1900 nm in the cannabis and tobacco mix samples was attributed to a combination of O–H stretching and bending, likely from water or alcohols [25]. Peaks observed in the herbal mix sample between 2150 and 2300 nm corresponded to C–H combination bands and overtones, particularly from CH_3 and CH_2 groups and C–N stretching vibrations [26]. These spectral features reflect the chemical complexity and compositional variation among the three sample types.

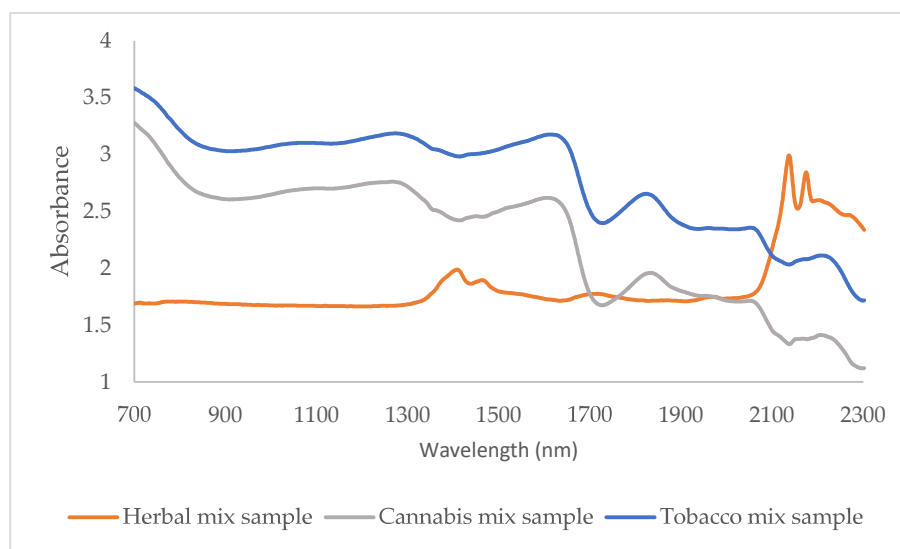


Figure 2. NIR spectra of cannabis mix sample, tobacco mix sample, and herbal mix sample.

2.3. Data Preprocessing

In multivariate modelling for IR spectroscopy, data preprocessing is crucial for improving the quality and reliability of spectral data. Baseline correction was applied using the MIR and NIR vendor's software (OMNIC software version 8.3 (ThermoFisher Scientific)) to address any baseline shifts or curvatures caused by instrumental or environmental factors, ensuring accurate spectral representation. In MIR spectroscopy, a wavenumber range from 650 to 2000 cm^{-1} was selected, focusing on the most informative regions of the spectrum. Further preprocessing steps were performed using Matlab. Normalization techniques were used to correct variations in intensity due to sample-related or instrument factors, allowing for accurate comparison between spectra. In this study, autoscaling and Standard Normal Variate (SNV) were explored. Additionally, derivative transformations were used to enhance spectral features and resolve overlapping peaks for better analysis.

2.4. Principal Component Analysis (PCA)

PCA was employed as an exploratory, unsupervised method to investigate the tendency of samples to form clusters. It helps to reduce the complexity of the data by projecting the original data into a smaller space of latent variables, while retaining the most significant variance in the spectral data. These latent variables are linear combinations of the original variables (such as signal intensity at specific wavelengths) that capture the most variance in PC1 and the remaining variance in the other PCs (PC2, PC3, etc.).

2.5. Hierarchical Clustering Analysis (HCA)

Hierarchical clustering was used to group samples based on the similarity of their spectra. The Euclidean distance metric was selected, and Ward's method was used as the clustering approach. Ward's algorithm follows a minimum variance criterion, aiming to minimize the increase in the sum of squared differences (or within-cluster variance) at each step of clustering [27].

2.6. Selection of Training and Test Sets

To validate the models, an external test set was selected to evaluate their performance. The Duplex algorithm was used for this purpose. This method ensures that the test set represents the original data set well and is evenly distributed across the data space. The Duplex algorithm selects sample pairs based on Euclidean distances. It starts by picking

the two samples with the greatest distance between them for the training set, then the second pair for the test set, and so it continues iteratively to add more pairs with the highest distances until the test set reaches the desired size. The remaining samples are added to the training set. About 20% of the total samples were used for the test set, while the remaining 80% were used for model training [27,28].

2.7. Soft Independent Modelling of Class Analogy (SIMCA)

SIMCA is a classification technique that focuses on modelling the similarity within each class, rather than distinguishing between classes. It uses Principal Component Analysis (PCA) to create a model for each class separately. The model defines a space around the training samples of each class using two distance metrics: Euclidean distance to the SIMCA model and Mahalanobis distance within the class's score space. The Euclidean distance measures how close a new sample's projection is to the class model, while the Mahalanobis distance accounts for variable correlations and the class's covariance structure. New samples are classified based on whether their projections fall within the defined space for each class. SIMCA is effective for complex data sets with multiple classes, as it creates separate models for each class, capturing variability and aiding in reliable classification [29–31].

2.8. Partial Least Squares (PLS)

PLS is a supervised technique similar to PCA, but it focuses on maximizing covariance with a response variable. It is commonly used for regression, where the response is continuous, such as dosage or concentration. PLS-Discriminant Analysis (PLS-DA) is a variant designed for classification tasks and can handle categorical response variables. PLS-DA is often used for pattern recognition and classification tasks [31–34].

2.9. Software

All data processing in this study was carried out using Matlab version R2019b (The Mathworks, Natick, MA, USA), a tool for scientific and numerical analysis. The SIMCA and PLS algorithms were implemented using the ChemoAC toolbox version 4.1 from the ChemoAC Consortium in Brussels, Belgium.

3. Results and Discussion

3.1. PCA

Score plots, produced using various data preprocessing methods, were compared, and the best distinction between spiked and non-spiked samples was achieved for MIR for detecting the presence of cannabis in tobacco samples, with first derivative as the preprocessing technique for the data (see Figure 3). The first two PCs accounted for over 99% of the total variance in the data, with PC1 explaining 99.08%, and PC2 0.85%. Also, for NIR, a clear separation of positive and negative samples could be obtained using the first derivative to pretreat the spectra. Here, 99% of the variance in the data was explained by the first two PCs (PC1: 88.63%, PC2: 10.22%). After analyzing the loadings on PC1, no particular wavelengths or wavelength ranges were found to be distinctly responsible for the observed clustering, which is to be expected, given the complexity of the matrix.

For the detection of tobacco in herbal smoking products, MIR did not show promising results. However, for NIR with SNV pretreatment, a clear separation of the positive and negative samples was observed, capturing 95% of the variance in the data in the first two PCs (PC1: 85.47%, PC2: 9.48%), as shown in Figure 4. After examining the loadings on PC1, no particular wavelengths or wavelength ranges could be clearly identified as contributing to the clustering, most probably due to the complexity of the herbal matrices.

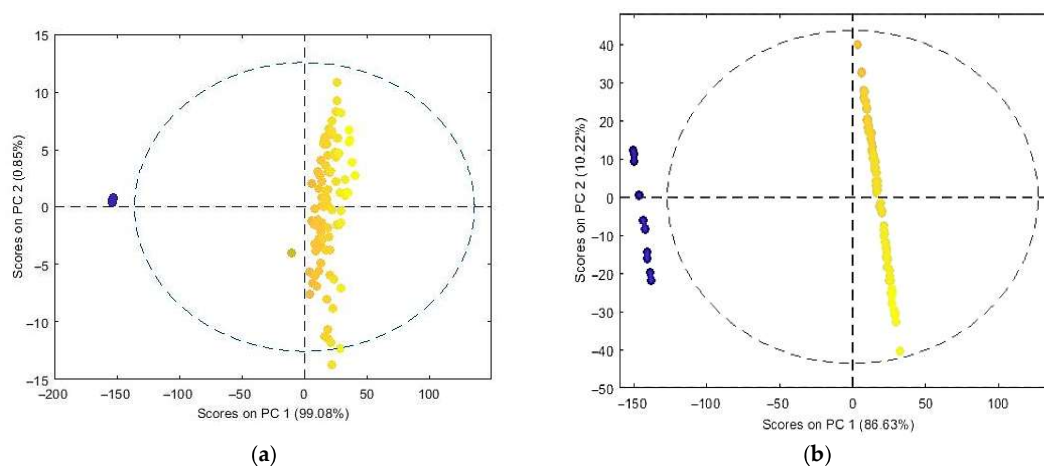


Figure 3. PCA plot obtained with (a) the MIR spectra and (b) the NIR spectra for cannabis detection in tobacco, using SNV as preprocessing technique. Yellow dots: spiked samples, blue dots: non-spiked samples, dashed oval represents confidence ellipse.

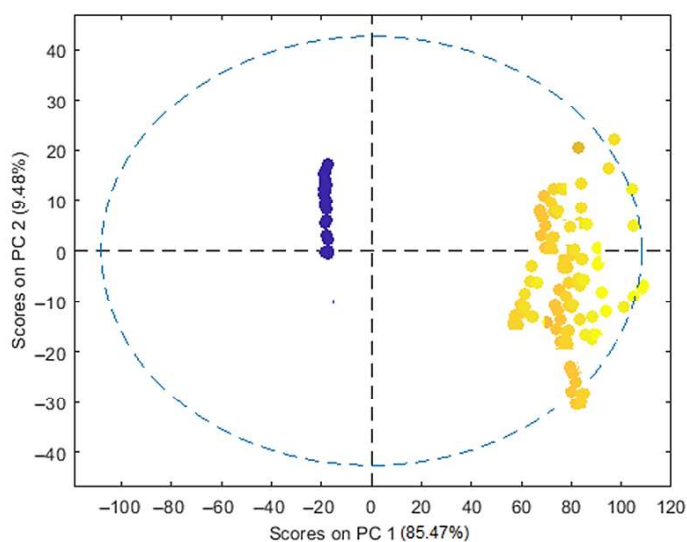


Figure 4. PCA plot obtained with the NIR spectra using the first derivative as preprocessing technique for tobacco detection in herbal smoking products. Yellow dots: spiked samples, blue dots: non-spiked samples, dashed oval represents confidence ellipse.

3.2. HCA

The HCA dendrogram based on MIR (Figure 5a) for cannabis detection in tobacco, showed three major clusters: the first cluster (Blue) corresponded to the mixture of tobacco samples spiked with the mixture of cannabis samples; the second cluster (Green) corresponded to tobacco samples not spiked with cannabis; and the third cluster (Red) were tobacco samples which were spiked with cannabis. The corresponding HCA dendrogram for NIR (Figure 5b) showed two major clusters: the first cluster (Green) corresponded to the non-spiked samples, and the second (Red) to the spiked ones.

The HCA dendrogram obtained using MIR spectra (Figure 6a) for tobacco detection in herbal smoking products showed three clusters: the first cluster (Blue) corresponded to samples which were all prepared with the mixture of herbal samples, spiked with the mixture of tobacco samples; the second cluster (Red) are the herbal samples spiked with tobacco samples; and the third cluster (Green) were samples which were not spiked. For

NIR (Figure 6b), a similar HCA dendrogram was obtained, distinguishing the same three major clusters. One sample was considered an outlier because the HCA dendrogram showed high dissimilarity for this sample towards the rest of the sample set and this sample was spiked with 20% of tobacco mixture.

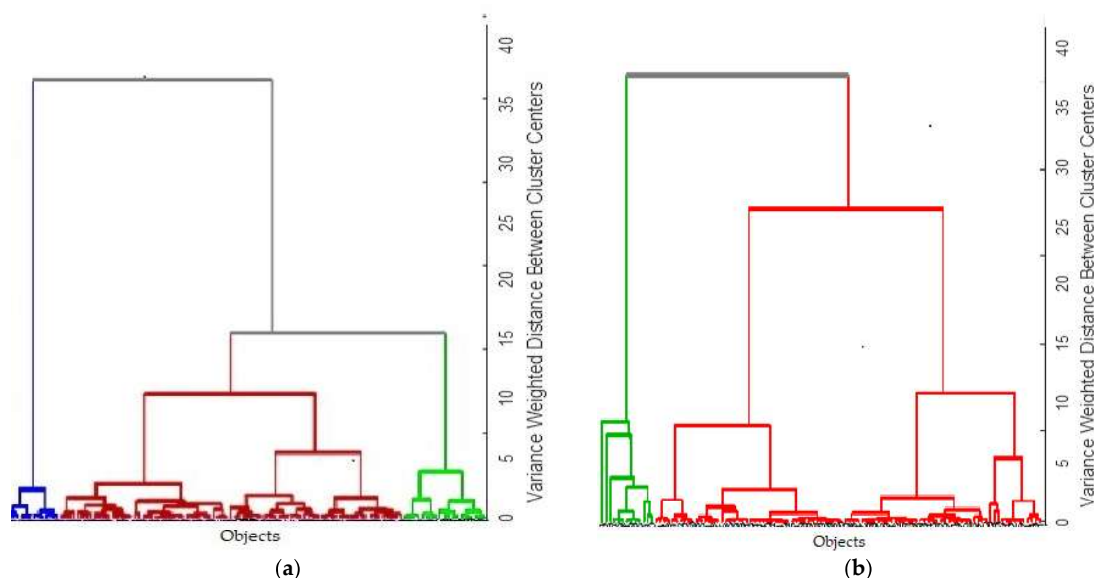


Figure 5. Hierarchical Cluster Analysis (HCA) dendrograms based on (a) MIR data for detecting cannabis in tobacco samples, Blue cluster: samples prepared with the mixture of tobacco samples and spiked with the mixture of cannabis samples, Red cluster: tobacco samples spiked with cannabis, and Green cluster: non-spiked samples; (b) NIR data, Green cluster: non-spiked samples, Red cluster: tobacco samples spiked with cannabis.

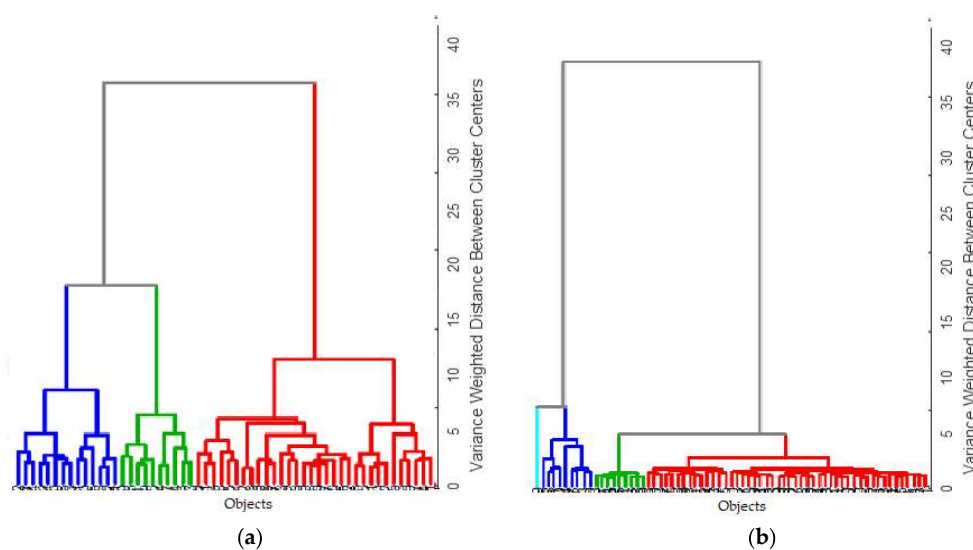


Figure 6. Hierarchical Cluster Analysis (HCA) dendrograms based on (a) MIR and (b) NIR spectroscopy for detecting tobacco in herbal smoking products. Blue cluster: samples prepared using the mixture of herbal smoking samples and spiked with the mixture of tobacco samples, Red cluster: herbal samples, spiked with tobacco samples, and Green cluster: non-spiked samples; Turquoise coloured sample: outlier.

3.3. SIMCA

MIR and NIR spectra were obtained for all samples. Before modelling, several preprocessing methods, including autoscaling, Standard Normal Variate (SNV), and first and second derivatives, were evaluated to optimize data interpretation by correcting baseline shifts, scaling differences, and resolving overlapping peaks in the spectra of the complex matrices. The study focused on detecting cannabis and tobacco adulteration, with 110 samples for cannabis-adulterated tobacco and 56 for tobacco-adulterated herbal smoking products. After preprocessing, SIMCA was used for binary modelling on MIR and NIR data separately. The data sets were split into training and test sets, with the training set used for model construction and internal validation via 10-fold cross-validation. External validation on the test set ensured unbiased performance assessment. The models effectively classified unknown samples by identifying spectral similarities and differences. Globally, SIMCA, with various data pretreatment methods, achieved correct classification rates ranging from 90% to 100% for the external test set and from 98% to 100% for cross-validation, as shown in Table 2. Table 3 shows the calculated values for specificity, accuracy, and sensitivity for all the selected models.

Table 2. Comprehensive overview of the performance of MIR and NIR and chemometric techniques in the classification of various target molecules.

Spectroscopic Technique	Data Pretreatment	Chemometric Technique	Target Adulterant	No. of PCs	No. of Samples in External Test Set	Correct Classification Rate (External Test Set) [No. of Negative Samples]	No. of Samples in the Training Set	Correct Classification Rate (Cross-Validation) [No. of Negative Samples]
MIR	Autoscaling	SIMCA	Cannabis	1-1	21	90% (19/21) [4/21]	89	98% (87/89) [8/89]
NIR	SNV	SIMCA	Cannabis	1-1	21	100% [3/21]	89	100% [9/89]
MIR	Autoscaling	SIMCA	Tobacco	2-1	13	92% (12/13) [3/13]	43	98% (42/43) [6/43]
NIR	Autoscaling	SIMCA	Tobacco	1-1	13	100% [3/13]	43	98% (42/43) [6/43]
Spectroscopic Technique	Data Pretreatment	Chemometric Technique	Target Adulterant	Latent Variables	No. of Samples in External Test Set	Correct Classification Rate (External Test Set) [No. of Negative Samples]	No. of Samples in the Training Set	Correct Classification Rate (Cross-validation) [No. of Negative Samples]
MIR	Autoscaling	PLS-DA	Cannabis	4	21	100% [4/21]	89	91% (81/89) [8/89]
NIR	Autoscaling	PLS-DA	Cannabis	4	21	100% [3/21]	89	100% [9/89]
MIR	Autoscaling	PLS-DA	Tobacco	2	13	100% [4/13]	43	100% [5/43]
NIR	2nd derivative	PLS-DA	Tobacco	7	13	92% (12/13) [3/13]	43	98% (42/43) [6/43]

For detecting the presence of cannabis in tobacco samples, the most effective model based on MIR was achieved using autoscaled spectra. Autoscaling normalizes variance across spectral variables, which seemed critical in MIR spectra where peak intensities varied significantly due to matrix effects. This model yielded a 90% correct classification rate on the external test set, with one sample classified as false positive and one as false negative. There was a cross-validation correct classification rate (CCR) of 98%, with two samples misclassified. Among these misclassifications, one was a false negative and one a false positive. It is important to note that the current focus is on minimizing false negatives, as false positives would be detected by inspectors and subsequently examined in a laboratory setting, to confirm their status. Regarding the false positives, no definitive explanation could be identified, and they might be attributed to random modelling errors. With NIR spectra, using SNV as data pretreatment, a SIMCA model was obtained, showing 100% correct classification in both cross-validation and external validation. SNV was

effective here because it corrects scattering effects in NIR spectra, which were particularly pronounced in the cannabis-adulterated tobacco samples.

Table 3. Classification statistics for cross-validation and test set for SIMCA and PLS-DA models.

SIMCA												
Target Adulterant	Cannabis			Cannabis			Tobacco			Tobacco		
Technique	MIR			NIR			MIR			NIR		
	Precision	Specificity	Sensitivity	Precision	Specificity	Sensitivity	Precision	Specificity	Sensitivity	Precision	Specificity	Sensitivity
Cross validation	0.98	0.90	0.98	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.91
Test Set	0.95	0.66	0.95	1.00	1.00	1.00	0.91	0.66	1.00	1.00	1.00	1.00

PLS-DA												
Target Adulterant	Cannabis			Cannabis			Tobacco			Tobacco		
Technique	MIR			NIR			MIR			NIR		
	Precision	Specificity	Sensitivity	Precision	Specificity	Sensitivity	Precision	Specificity	Sensitivity	Precision	Specificity	Sensitivity
Cross validation	0.96	0.75	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
Test Set	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.66	1.00

Shifting to MIR spectroscopy and SIMCA for tobacco detection in herbal smoking products, the most effective model was obtained with autoscaling as data pretreatment. Autoscaling improved discrimination by normalizing spectral variability arising from heterogeneous sample matrices. The external test set for this model achieved an accuracy of 92%, with only one sample misclassified as a false positive. Cross-validation accuracy was similar at 98%, with one misclassified sample which was categorized as a false negative. This false negative sample exhibited a concentration of 10% tobacco, indicating potential contributions from factors such as sample matrix or instrumental variations. An explanation for the false positive sample in the external test set could not be found. The complexity of the herbal matrix and the similarities in the matrices of the different samples could be an explanation. Moving towards NIR spectroscopy and SIMCA, the best model was obtained with autoscaling. The external test set achieved an accuracy of 100%. Cross-validation accuracy was 98%, with one misclassified sample which was categorized as a false negative. This false negative sample exhibited a concentration of 5% tobacco, which is quite low, but this is probably due to random modelling errors, since lower concentrations are classified correctly. Thus, the choice of preprocessing method in each case was guided by both spectral characteristics and model performance, ensuring the chemometric models could reliably classify complex adulterated samples.

3.4. PLS-DA Model

In the next step, PLS-DA models were calculated using the same spectral data sets. The statistics for the best models are also summarized in Tables 2 and 3.

Employing MIR spectroscopy for cannabis detection in tobacco products, alongside autoscaling data pretreatment, PLS-DA achieved a 100% correct classification rate for the external test set. The process of cross-validation yielded a correct classification rate of 91%, with 81 of the 89 samples being accurately categorized. Delving deeper, the analysis unveiled that three samples were misclassified as positive, whereas five were inaccurately identified as negatives. This classification challenge seemed to revolve around specific high-dosage cannabis samples, with instances ranging from 10 to 30% cannabis samples erroneously labelled as negatives. Autoscaling normalizes variance across spectral variables, improving model performance, despite the complex heterogeneous matrices which are the subject of this study. Contrarily, the utilization of NIR spectroscopy, coupled with autoscaling data pretreatment, showed 100% correct classification for both the external test set and cross-validation.

MIR spectroscopy and PLS-DA for tobacco detection in herbal smoking products in tandem with autoscaling resulted in a 100% correct classification rate for both the test set and cross-validation. Conversely, the application of NIR spectroscopy revealed results of achieving a 92% external test set correct classification rate, with one sample being misclassified as positive and a 98% rate during cross-validation. Notably, one sample was misclassified during cross-validation, comprising a false negative, which is primarily associated with a low concentration of 5% tobacco. Autoscaling effectively normalized spectral variability arising from heterogeneous herbal matrices, allowing the PLS-DA models to discriminate accurately despite minor matrix interferences. Despite scrutiny, a concrete explanation for the false positive classifications remained elusive. Thus, the selection of autoscaling for both MIR and NIR data sets was guided by spectral characteristics and model performance, ensuring robust classification in complex adulterated samples.

4. Conclusions

The study systematically evaluated the effectiveness of MIR and NIR, combined with various data preprocessing and multivariate calibration methods (PCA, HCA, SIMCA, and PLS-DA) for the detection of cannabis and tobacco as adulterant in tobacco and herbal smoking products, respectively. Through data exploration with PCA and HCA, it was observed that both spectroscopic techniques are able to distinguish between adulterated and non-adulterated samples, and this was true in both cases studied here with the lowest threshold for adulteration (1% m/m). Based on these results, it was decided to continue with supervised modelling using SIMCA and PLS-DA. SIMCA modelling demonstrated high classification accuracy, with rates ranging from 90% to 100% across various preprocessing methods. The models for both cannabis and tobacco adulteration were particularly effective when using autoscaling as a preprocessing technique, achieving near-perfect classification rates, especially in NIR spectroscopy. Lastly, the PLS-DA model showed excellent performance, with perfect classification rates for the external test set using MIR and NIR for both targeted adulterants. Cross-validation results were slightly lower, but still demonstrated strong classification capabilities, especially with autoscaling as data pretreatment method. Comparing the different models, NIR spectroscopy combined with SIMCA yielded the overall best results for the detection of the targeted adulterants, though the performance of MIR is very close. This means that both spectroscopic techniques could be used by law enforcement, e.g., to detect cannabis in tobacco products in night life settings, or by inspectors, visiting shops or suspicious wholesalers selling tobacco and/or herbal smoking products, if they are combined with the correct data preprocessing and basic modelling techniques. Their applicability in this context would even increase, given the availability of highly performant portable devices for both techniques, to which the proposed approach could be transferred. This is especially true for new portable instruments which can be linked to laptops or tablets, enabling convenient on-site data collection and processing even with more advanced software. However, besides the availability of highly performant instruments, several practical challenges remain. Sample heterogeneity, variable moisture content, and the influence of different packaging materials can affect spectral quality and classification reliability. Furthermore, environmental factors (e.g., temperature, humidity) may introduce additional variability in real-world conditions. Therefore, future validation studies should focus on testing under operational settings with diverse product types, packaging formats, and storage conditions. Addressing these aspects will be crucial before these technologies can be routinely implemented by law enforcement agencies.

5. Limitations

It should be noted that the products analyzed in this study were limited in diversity, both in the number of samples or brands and geographically. This restricted sampling scope may limit the generalizability of the developed models, as product composition and manufacturing practices can vary considerably across brands and regions. Future research should therefore include a wider selection of products from different geographical areas and manufacturers to better capture inter-product variability and to validate the robustness and transferability of the proposed chemometric models.

In practical applications, several factors can influence the spectral quality and, consequently, the model performance. Sample heterogeneity, such as uneven particle size, inconsistent moisture levels, or the presence of natural additives, can cause spectral variability and reduce classification accuracy. Similarly, environmental factors, including temperature fluctuations and ambient humidity during data acquisition, can introduce baseline shifts or scattering effects in the spectra. These sources of variability highlight the need for robust preprocessing strategies and model validation under varied conditions.

The integration of the developed models into portable spectroscopic instruments offers significant translational potential. Next to the challenge of treating samples on site in the context of homogeneity of the samples, measuring conditions, etc., such deployment would require onboard chemometric software capable of rapid data preprocessing, multivariate analysis, and automated classification. Ensuring model interpretability and minimal operator training would be essential for reliable field use, particularly in regulatory inspections or product authentication scenarios. Future work will explore real-time model implementation in handheld spectrometers to evaluate performance under realistic environmental and operational settings.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/chemosensors13100370/s1>: Table S1: Concentration levels (% m/m) of cannabis samples in tobacco samples; Table S2: Tobacco–Cannabis Spiking Scheme; Table S3: Concentration levels (% m/m) of tobacco samples in herbal smoking samples.

Author Contributions: Conceptualization, E.D., E.A. and C.D.; methodology, Z.A. and E.D.; validation, Z.A. and E.D.; formal analysis, Z.A.; investigation, Z.A.; resources E.A. and E.D.; data curation, Z.A.; writing—original draft preparation, Z.A.; writing—review and editing, E.D., E.A., I.u.R. and C.D.; supervision, E.D. and E.A.; project administration E.D. and E.A.; funding acquisition, E.D., E.A. and Z.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by HEC (Higher Education Commission) Pakistan, funding no: PD/HEC/HRD/OSS-III/BIg-B2/2021/19331/19347 and SCIENSANO.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Ren, M.; Tang, Z.; Wu, X.; Spengler, R.; Jiang, H.; Yang, Y.; Boivin, N. The origins of cannabis smoking: Chemical residue evidence from the first millennium BCE in the Pamirs. *Sci. Adv.* **2019**, *5*, eaaw1391. [[CrossRef](#)] [[PubMed](#)]
2. Śliwińska-Mossoń, M.; Milnerowicz, H. The impact of smoking on the development of diabetes and its complications. *Diabetes Vasc. Dis. Res.* **2017**, *14*, 265–276. [[CrossRef](#)]
3. Kondo, T.; Nakano, Y.; Adachi, S.; Murohara, T. Effects of tobacco smoking on cardiovascular disease. *Circ. J.* **2019**, *83*, 1980–1985. [[CrossRef](#)] [[PubMed](#)]

4. Zhang, Y.; He, J.; He, B.; Huang, R.; Li, M. Effect of tobacco on periodontal disease and oral cancer. *Tob. Induc. Dis.* **2019**, *17*, 40. [CrossRef]
5. Ockene, I.S.; Miller, N.H. Cigarette smoking, cardiovascular disease, and stroke: A statement for healthcare professionals from the American Heart Association. *Circulation* **1997**, *96*, 3243–3247. [CrossRef]
6. Lee, E.S.; Seo, H.G. The factors associated with successful smoking cessation in Korea. *J. Korean Acad. Fam. Med.* **2007**, *28*, 39–44.
7. Fromme, H.; Schober, W. Waterpipes and e-cigarettes: Impact of alternative smoking techniques on indoor air quality and health. *Atmos. Environ.* **2015**, *106*, 429–441. [CrossRef]
8. Sutton, C.D.; Robinson, R.G. The marketing of menthol cigarettes in the United States: Population, messages, and channels. *Nicotine Tob. Res.* **2004**, *6* (Suppl. S1), S83–S91. [CrossRef]
9. European Union. Directive 2014/40/EU of the European Parliament and of the Council of 3 April 2014 on the approximation of the laws, regulations and administrative provisions of the Member States concerning the manufacture, presentation and sale of tobacco and related products and repealing Directive 2001/37/EC. *Off. J. Eur. Union* **2014**, *L127*, 1–38.
10. Gan, Q.; Yang, J.; Yang, G.; Goniewicz, M.; Benowitz, N.L.; Glantz, S.A. Chinese “herbal” cigarettes are as carcinogenic and addictive as regular cigarettes. *Cancer Epidemiol. Biomark. Prev.* **2009**, *18*, 3497–3501. [CrossRef]
11. Chen, A.; Glantz, S.; Tong, E. Asian herbal-tobacco cigarettes: “Not medicine but less harmful”? *Tob. Control* **2007**, *16*, e3. [CrossRef]
12. Schauer, G.L.; Rosenberry, Z.R.; Peters, E.N. Marijuana and tobacco coadministration in blunts, spliffs, and mulled cigarettes: A systematic literature review. *Addict. Behav.* **2017**, *64*, 200–211. [CrossRef]
13. Meier, E.; Hatsukami, D.K. A review of the additive health risk of cannabis and tobacco co-use. *Drug Alcohol Depend.* **2016**, *166*, 6–12. [CrossRef]
14. Akre, C.; Michaud, P.A.; Berchtold, A.; Suris, J.C. Cannabis and tobacco use: Where are the boundaries? A qualitative study on cannabis consumption modes among adolescents. *Health Educ. Res.* **2010**, *25*, 74–82. [CrossRef] [PubMed]
15. Bélanger, R.E.; Akre, C.; Kuntsche, E.; Gmel, G.; Suris, J.C. Adding tobacco to cannabis—its frequency and likely implications. *Nicotine Tob. Res.* **2011**, *13*, 746–750. [CrossRef] [PubMed]
16. Bélanger, R.E.; Marclay, F.; Berchtold, A.; Saugy, M.; Cornuz, J.; Suris, J.C. To what extent does adding tobacco to cannabis expose young users to nicotine? *Nicotine Tob. Res.* **2013**, *15*, 1832–1838. [CrossRef] [PubMed]
17. United Nations. Convention Against Illicit Traffic in Narcotic Drugs and Psychotropic Substances. 1998. Available online: https://www.unodc.org/pdf/convention_1988_en.pdf (accessed on 4 February 2025).
18. United Nations Office on Drugs and Crime. Méthodes Recommandées pour L’identification et L’analyse du Cannabis et des Produits du Cannabis. 2009. Available online: <https://www.unodc.org/documents/scientific/Cannabis-F.pdf> (accessed on 4 February 2025).
19. European Directorate for the Quality of Medicines. *Qualification of Equipment Annex 4 Qualification of IR Spectrophotometers*; (PA/PH/OMCL (07) 12 DEF CORR). European Directorate for the Quality of Medicines: Strasbourg, France, 2007. Available online: <https://www.edqm.eu/en/d/129347> (accessed on 4 February 2025).
20. Smith, B.C. *Infrared Spectral Interpretation: A Systematic Approach*; CRC Press: Boca Raton, FL, USA, 2011.
21. Griffiths, P.R.; de Haseth, J.A. *Infrared Spectroscopy: Fundamentals and Applications*; Wiley-Interscience: Hoboken, NJ, USA, 2007.
22. Chalmers, J.M.; Griffiths, P.R. *Handbook of Vibrational Spectroscopy*; Wiley: Chichester, UK, 2002.
23. Jonathan, B. Chapter 5: Diffuse reflectance spectroscopy. In *Modern Techniques and Applications in Molecular Spectroscopy*; John Wiley & Sons: Hoboken, NJ, USA, 1998; pp. 145–160.
24. Burns, D.A.; Ciurczak, E.W. (Eds.) *Handbook of Near-Infrared Analysis*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2007.
25. Workman, J.; Weyer, L. *Practical Guide to Interpretive Near-Infrared Spectroscopy*; CRC Press: Boca Raton, FL, USA, 2012.
26. Osborne, B.G.; Fearn, T.; Hindle, P.T. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*; Longman: Harlow, UK, 1993.
27. Ziegel, E.R. Handbook of Chemometrics and Qualimetrics, Part B. *Technometrics* **2000**, *42*, 218–219. [CrossRef]
28. Snee, R.D. Validation of regression models: Methods and examples. *Technometrics* **1977**, *19*, 415–428. [CrossRef]
29. Wold, S. Pattern recognition by means of disjoint principal components models. *Pattern Recognit.* **1976**, *8*, 127–139. [CrossRef]
30. Roggo, Y.; Chalou, P.; Maurer, L.; Lema-Martinez, C.; Edmond, A.; Jent, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J. Pharm. Biomed. Anal.* **2007**, *44*, 683–700. [CrossRef]
31. Gurbanov, R.; Gozen, A.G.; Severcan, F. Rapid classification of heavy metal-exposed freshwater bacteria by infrared spectroscopy coupled with chemometrics using supervised method. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2018**, *189*, 282–290. [CrossRef] [PubMed]
32. Saleh, A.A.; Hegazy, M.; Abbas, S.; Elkosasy, A. Development of distribution maps of spectrally similar degradation products by Raman chemical imaging microscope coupled with a new variable selection technique and SIMCA classifier. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *268*, 120654. [CrossRef] [PubMed]

33. Akhtar, Z.; Barhdadi, S.; De Braekeleer, K.; Delporte, C.; Adams, E.; Deconinck, E. Spectroscopy and chemometrics for conformity analysis of e-liquids: Illegal additive detection and nicotine characterization. *Chemosensors* **2024**, *12*, 9. [[CrossRef](#)]
34. Akhtar, Z.; Canfyn, M.; Vanhee, C.; Delporte, C.; Adams, E.; Deconinck, E. Evaluating MIR and NIR spectroscopy coupled with multivariate analysis for detection and quantification of additives in tobacco products. *Sensors* **2024**, *24*, 7018. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.