Spectroscopy and Chemometrics for Conformity Analysis of e-Liquids: Illegal Additive Detection and Nicotine Characterization

Zeb Akhtar [1,2], Sophia Barhdadi [1], Kris De Braekeleer [3], Cedric Delporte [3,4], Erwin Adams [2]
and Eric Deconinck [1,*]

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

Abstract: Vaping electronic cigarettes (e-cigarettes) has become a popular alternative to smoking tobacco. When an e-cigarette is activated, a liquid is vaporized by heating, producing an aerosol that users inhale. While e-cigarettes are marketed as less harmful than traditional cigarettes, there are ongoing concerns about their long-term health effects, including potential lung damage. Therefore, it is essential to closely monitor and study the composition of e-liquids. E-liquids typically consist of propylene glycol, glycerin, flavorings and nicotine, though there have been reports of non-compliant nicotine concentrations and the presence of illegal additives. This study explored spectroscopic techniques to examine the conformity of nicotine labeling and detect the presence of the not-allowed additives: the caffeine, taurine, vitamin E and cannabidiol (CBD) in e-liquids. A total of 236 e-liquid samples were carefully selected for analysis. Chemometric analysis was applied to the collected data, which included mid-infrared (MIR) and near-infrared (NIR) spectra. Supervised modeling approaches such as partial least squares-discriminant analysis (PLS-DA) and soft independent modeling of class analogy (SIMCA) were employed to classify the samples, based on the presence of nicotine and the targeted additives. This study demonstrates the efficacy of MIR and NIR spectroscopic techniques in conjunction with chemometric methods (SIMCA and PLS-DA) for detecting specific molecules in e-liquids. MIR with autoscaling data preprocessing and PLS-DA achieved 100% classification rates for CBD and vitamin E, while NIR with the same approach achieved 100% for CBD and taurine. Overall, MIR combined with PLS-DA yielded the best classification across all targeted molecules, suggesting its preference as a singular technique.

Keywords: e-liquids; spectroscopic techniques; chemometric techniques; data analysis

1. Introduction

In the early 2000s, Hon Lik, a Chinese chemist, developed and obtained patents for the technology behind e-cigarettes. He developed it as an alternative to traditional tobacco cigarettes, with a reduced risk of cancer. As worldwide cigarette consumption declined due to anti-smoking regulations, large multinational tobacco companies recognized the potential of e-cigarettes to generate substantial profits in the future. They initiated global marketing campaigns promoting e-cigarettes as healthier alternatives to smoking, effective tools for smoking cessation, or as a means to reduce overall cigarette consumption. As a result, the use of e-cigarettes has significantly increased on a global scale over the past

decade, with current estimates suggesting that there are more than 55 million individuals who use e-cigarettes, commonly referred to as "vaping products" [1,2].

The primary mechanism of an e-cigarette involves the vaporization of an e-liquid through an atomizer, resulting in the formation of an aerosol. This aerosol comprises microscopic particles derived from the e-liquid and the atomizer. The composition of the e-liquid primarily consists of propylene glycol, (vegetable) glycerin, and nicotine, with the addition of food-grade flavorings as supplementary additives.

E-cigarettes have gained popularity as a potential substitute for smoking cessation, primarily due to their resemblance to traditional cigarettes and the perception of delivering nicotine in a more realistic manner. Some research studies, such as the one conducted by Hajek et al. (2019) [3], suggest that e-cigarettes could be helpful in aiding individuals to quit smoking. However, there is an ongoing debate surrounding the safety of e-cigarettes, primarily due to the limited availability of long-term safety data. Several studies have raised concerns about the general safety of e-cigarettes. One significant risk associated with using e-cigarettes as a smoking cessation aid is the possibility that individuals may replace traditional cigarettes with e-cigarettes and continue using them for an extended period, believing them to be harmless. This is particularly of concern among young individuals who have never smoked before, as it may make e-cigarettes more appealing, create a false sense of security, and potentially serve as a gateway to other substances, as suggested by Wadgave and Nagesh (2016) [4].

Legislation regarding these e-liquids varies from country to country. In some countries they are considered as consumer products, while at the European level, the Tobacco Products Directive 2014/40/EU (TPD) was revised by the European Parliament in 2014 to ensure the quality and safety of products. It has since added a set of rules governing e-cigarettes, considering nicotine-containing e-liquids as tobacco products in the European Union (EU). A number of stricter regulations regarding e-cigarette product promotion, packaging, and warning labels, as well as capped maximum amounts for cartridges and e-liquids, are included in the amended TPD. Additionally, in some countries like Belgium, legislation does not allow the sale and purchase of e-liquids through the internet. Indeed, in the new Royal Decree, nicotine-free products are also not allowed to be sold online. A problem with e-liquids is the label conformity for the nicotine content and the additives, like caffeine, taurine, vitamins and CBD (when combined with nicotine), which are banned by the authorities [5–8]. Control of manufacturer claims regarding nicotine concentration and delivery is particularly glaring in this regard. Before putting their items on the EU market, manufacturers must also notify the relevant authorities. Moreover, more stringent criteria are proposed for the substances found in e-liquids, making chemical characterization of these compounds necessary to monitor compliance with the TPD and to regulate the quality of the goods on the market (European Parliament and the Council of the European Union, Directive 2014/40/EU) [5].

Nowadays, nicotine and additives in e-liquids are determined using techniques such as polarographic analysis [9], atomic absorption spectrometry, liquid chromatography UV [10], UV spectrophotometry [11], and gas chromatography [12]. These techniques have some disadvantages, like cumbersome processing of the samples, high cost and the need to be performed in a laboratory. However, for inspectors to know which samples to seize, a convenient on-site method for characterizing e-liquids, i.e., nicotine content and additive content, is necessary. Due to its distinct benefits, (near-)infrared spectroscopy technology has recently been widely used in various industries [13]. It has demonstrated already its efficacy in the identification of active pharmaceutical ingredients in medications [14–16], as well as illicit ingredients in cosmetics and nutritional supplements [16]. Up till now, there are not many reports on its usage for e-cigarettes.

FTIR (Fourier-Transform Infrared) spectroscopy is a powerful technique that operates in the infrared region of the electromagnetic spectrum. It utilizes an interferometer to measure the spectrum simultaneously, providing high-resolution data. This makes FTIR spectroscopy well suited for structural characterization and the identification of com-

pounds. Its ability to provide detailed information about chemical bonds and functional groups makes it valuable in various fields such as organic chemistry, materials science, and pharmaceutical analysis. Additionally, FTIR spectroscopy allows for quantitative analysis, making it useful for determining the concentration of components in a sample. It requires minimal sample preparation and can provide accurate results, even with small sample sizes [17–20]. Mid-Infrared (MIR) spectroscopy operates in the mid-infrared region. It focuses on the fundamental vibrations of chemical bonds, which occur in this region [21,22]. Near-Infrared (NIR) spectroscopy on the other hand, covers the near-infrared region of the electromagnetic spectrum. It involves measuring the absorption and reflection of NIR light by samples. The NIR region contains overtone and combination bands, which provide information about the chemical composition of a sample. NIR spectroscopy is commonly used in the pharmaceutical industry, agriculture, and food analysis [23]. It is valuable for assessing chemical properties like moisture content, protein, fat, and sugar levels in samples. Additionally, NIR spectroscopy can be used for qualitative and quantitative analysis, making it a versatile tool for composition analysis [24].

In this research, the use of MIR and NIR spectroscopy was explored, in combination with chemometric modeling, which allowed for the identification and detection of nicotine and different additives in the e-liquids.

2. Materials and Methods

2.1. E-Liquid Sample Set

A total of two hundred and thirty-six (236) samples were collected for analysis. These samples were obtained from inspections conducted by federal government authorities at various vaping shops in Belgium, as well as from the interception of post packages ordered online. To maintain sample integrity, all collected samples were stored under controlled conditions. Specifically, they were stored at room temperature (15 to 25 °C) and protected from light.

The samples were all screened for the presence of nicotine and illegal additives, using accredited gas chromatography—mass spectrometry and liquid chromatography—high-resolution mass spectrometry approaches [25]. These results can be found in Supplementary Table S1.

2.2. Reagents and Chemicals

USP Grade Nicotine 99.9% was purchased from Vaping Zone (Columbia, SC, USA). α-Tocopherol ($\geq$96%) and taurine ($\geq$99%) were obtained from Sigma Aldrich (St. Louis, MO, USA). Caffeine and CBD (100.0% purity) were purchased from Fagron (Nazareth, Belgium) and ethanol (99.9%) from Biosolve (Valkenswaard, The Netherlands).

2.3. Sample Preparation

For this study, a series of e-liquids were spiked with pure taurine, CBD, vitamin E and caffeine, following the subsequent steps. Firstly, stock solutions of each compound were prepared individually, as shown in Table 1. To spike the e-liquid, the desired amount of the e-liquid sample was brought, in a clean vial. Spiked samples were prepared by spiking zero-liquid samples with the required volumes of additive stock solutions to achieve the desired spiked concentrations in the e-liquid, as shown in Table 1. The spiked e-liquid was thoroughly mixed using a vortex mixer or gently shaken to ensure uniform distribution of the compounds. Control samples were prepared by adding the corresponding volume of solution, but free of additive.

Table 1. Concentrations of stock solutions and spiking concentrations from each stock solution for the different additives.

Additive Compounds	Solvent	Concentrations of Stock Solutions	Spiking Concentrations
CBD	Ethanol (100%)	5 mg/mL	3.5–19.5% m/m [26]
Vitamin E	Ethanol (100%)	15 mg/mL	500–2000 mg/kg or ppm [27]
Taurine	Milli Q water	20 mg/mL	100 mg/L–10,000 mg/L [28]
Caffeine	Milli Q water	0.25 mg/mL	20–100 ppm [29]
Nicotine	e-liquid matrix	100 mg/mL	0–25 mg/mL

2.4. Data Acquisition

2.4.1. MIR

Utilizing a Nicolet iS10 FTIR spectrometer (ThermoFisher Scientific, Waltham, MA, USA) with a Smart iTR accessory and a deuterated triglycine sulphate (DTGS) detector, MIR spectra were acquired. The Smart iTR attachment, employing a singular-bounce diamond crystal, underwent a weekly calibration procedure involving a polystyrene film standard. Subsequent to the instrumentation setup, infrared (IR) spectra spanning the wavenumber range from 4000 to 400 cm^{-1} were systematically recorded. Each spectrum entailed 32 accumulative scans at a spectral resolution of 4 cm^{-1}. The acquired spectral datasets underwent processing using OMNIC software version 8.3, developed by ThermoFisher Scientific, Madison, WI, USA. Post data acquisition, the employed diamond crystal underwent meticulous cleaning through a sequential procedure involving methanol-soaked soft-tissue cleaning and subsequent air drying. Prior to evaluating each analyte, a blank measurement was executed to assess the crystal's susceptibility to contamination and carry-over effects. Criteria for contamination, as defined by the European Directorate for the Quality of Medicines and HealthCare (EDQM) (2007) [30] were rigorously adhered to. Furthermore, in the pursuit of sustained instrument precision, hourly background spectra against ambient air were consistently recorded and accounted for.

2.4.2. NIR

All samples underwent scanning, utilizing a Frontier MIR/NIR Spectrometer™ (PerkinElmer, Waltham, MA, USA). Reflectance mode was employed to measure spectra using the NIR Reflectance Accessory, employing an 8 cm^{-1} resolution within the 10,000–4000 cm^{-1} range. These spectra were generated by averaging 16 scans. For each replicate, background spectra were gathered with the assistance of a diffuse reflector from PerkinElmer. The background was captured between individual samples. To capture the spectra, background subtraction was conducted, and arithmetic functions were employed to correct for background influences.

3. Chemometric Methods

The general steps of the procedure followed during the data or chemometric analysis of the collected (N)IR data are represented in Figure 1.

3.1. Data Preprocessing

Data preprocessing in chemometrics for IR spectroscopy involves a series of steps, undertaken to enhance the quality and reliability of the spectral data. Figure 1 illustrates the main steps involved in the data analyses. Baseline correction was performed in the MIR and NIR software to address baseline drift or curvature caused by instrumental or environmental factors, ensuring accurate representation of the spectral features. Wavelength range selection was carried out in NIR spectroscopy using its software (Frontier MIR/NIR Perkin Elmer, Waltham, MA, USA), which focused on relevant regions of the spectrum that contained the most meaningful chemical information for the specific analysis. Then the other steps were carried out in Matlab. Normalization methods were used to eliminate intensity variations arising from sample concentration or instrumental factors, allowing for

fair comparisons between spectra. In this study, autoscaling and standard normal variate (SNV) were explored. Derivative transformations were employed to enhance spectral features and resolve overlapping peaks, enabling better characterization and analysis [31,32].
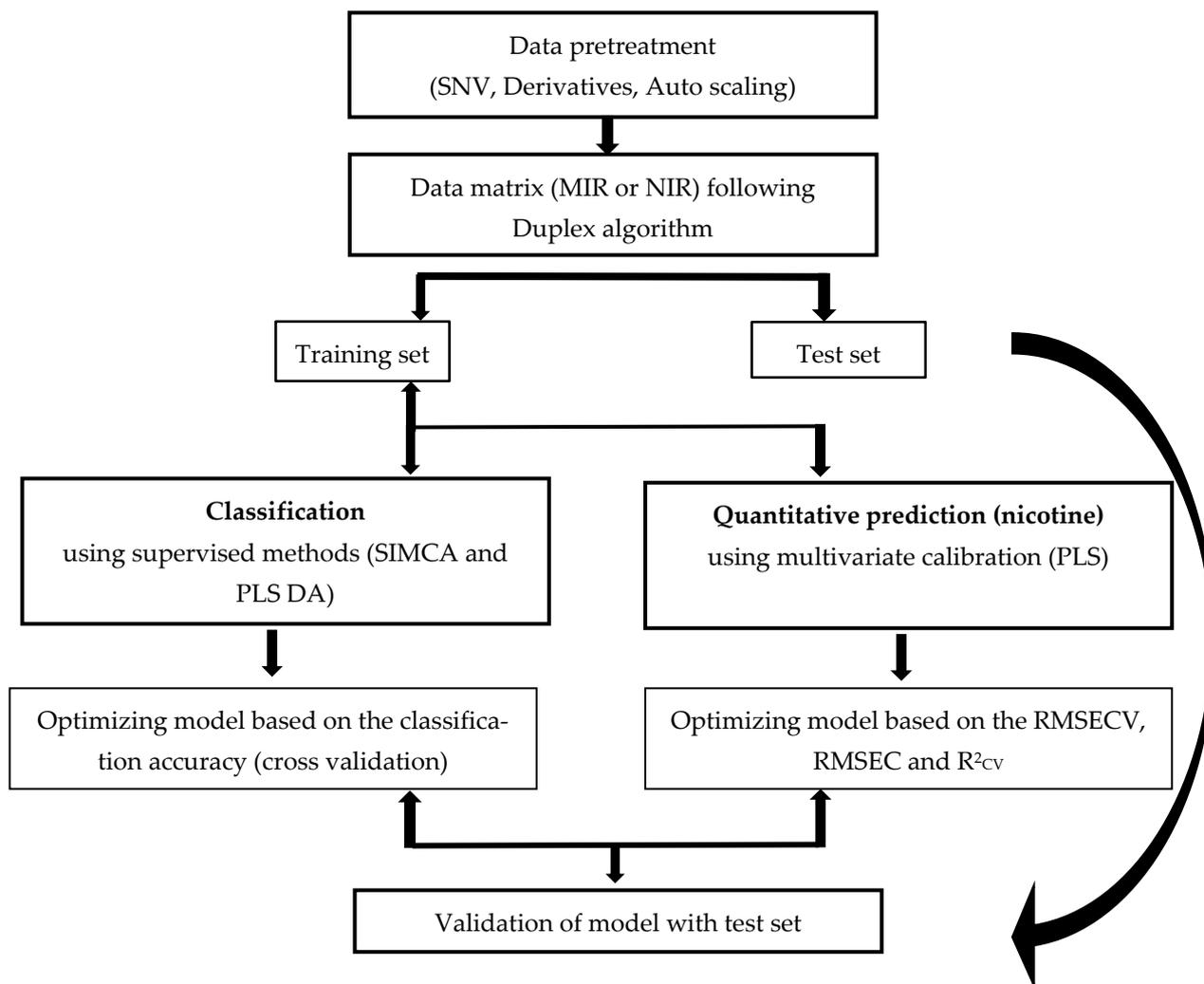
Figure 1. General steps involved in the application of chemometrics methods to treat near- and mid-infrared spectral data.

3.2. Selection of Training and Test Set

For model validation, an external test set needs to be chosen to assess the performance of the models developed. In this study, the Duplex algorithm was utilized (Figure 1). This method is designed to ensure that the test set represents the entire original dataset and is uniformly distributed across the data space [33]. The Duplex algorithm relies on Euclidean distances to select samples in pairs. It starts by identifying the two samples with the highest Euclidean distance between them, and assigns them to the first set. Subsequently, another pair of samples with the highest Euclidean distance is chosen and added to the test set. This iterative process continues until the desired number of samples is obtained for the test set. The first set and the remaining samples that were not selected form the training set [33]. Approximately 20% of the total samples were allocated to the test set for external model validation, while the remaining 80% (the training set) were used to construct the models.

3.3. Soft Independent Modeling of Class Analogy (SIMCA)

SIMCA is a supervised classification technique that focuses on emphasizing the similarity within classes, rather than the discrimination between classes. This approach is known as disjoint-class modeling. In SIMCA, each class is modeled separately, using principal component analysis (PCA) [34]. The modeling process involves constructing a space around the training samples of each class. This space is defined using two distance measures: Euclidean distance towards the SIMCA model and the Mahalanobis distance determined in the space of scores. The Euclidean distance measures how close a new sample's projection is to the SIMCA model of a particular class, while the Mahalanobis distance takes into account the correlations between variables and calculates the distance relative to the covariance structure of the class [35]. When a new sample is analyzed, its projection is compared to the defined spaces around the training samples of each class. If the projection falls within the space defined around the training samples of a specific class, the new sample is assigned to that class. The advantage of SIMCA is its ability to handle complex datasets with multiple classes, by building separate models for each class. This approach allows for the capture of the inherent variability within each class, and for making reliable predictions or classifications based on the proximity of new samples to the existing class models [36,37].

3.4. Partial Least Squares (PLS)

PLS is a supervised projection technique that shares similarities with PCA. In PLS, latent variables are derived as linear combinations of the observed variables, defined to maximize the covariance with a specific response variable. PLS is commonly used for regression tasks, where the response variable is continuous, such as dosage or concentration. PLS-discriminant analysis (PLS-DA) is a variation of PLS, specifically designed for classification tasks. It allows working with categorical response variables, enabling the use of PLS as a classification technique. PLS-DA is widely employed for pattern recognition and classification problems [38,39].

3.5. Software

In this study, all data treatments were performed using Matlab version R2019b, a software package for scientific and numerical computing. The ChemoAC toolbox version 4.1, developed by the ChemoAC Consortium in Brussels, Belgium, was used for the SIMCA and PLS algorithms, and PCA was conducted utilizing the PLS toolbox from Eigenvector Research, Inc., Manson, WA, USA.

4. Results and Discussion

4.1. Qualitative Models

Qualitative models were constructed using the spectra of all samples in the sample set. Initially, PCA was applied to explore the potential discriminatory capability between samples containing the targeted compound and other samples within the dataset (results not shown). Subsequently, binary supervised classification models were developed, using both SIMCA and PLS-DA. The sample set was categorized into five distinct groups of samples, i.e., nicotine-containing samples, caffeine-containing samples, taurine-containing samples, vitamin E-containing samples and finally CBD-containing samples. The objective of these models was to effectively classify and distinguish samples, based on the presence of the targeted molecules. Tables 2 and 3 and Figures 2 and 3 show for each combination of spectroscopic technique, chemometric technique and target molecule, the performance of the best model obtained and their sensitivity, precision and specificity. For the ideal model, values for sensitivity, precision and specificity would be 1.00.

Table 2. Comprehensive overview of the performance of MIR and NIR, data pretreatment methods, and SIMCA in the classification of various target molecules.

Spectroscopic Technique	Data Pretreatment	Chemometric Technique	Target Molecule	Samples in External Test Set	Correctly Classified Samples (External Test Set)	Samples in Training Set	Correct Classification Rate (Cross-Validation)
MIR	2nd derivative	SIMCA	Nicotine	48	83% (40/48)	188	81% (153/188)
	Autoscaling		CBD	48	98% (47/48)	188	95% (178/188)
	1st derivative		Vit E	48	92% (44/48)	188	97% (182/188)
	Autoscaling		Taurine	48	96% (46/48)	188	88% (166/188)
	1st derivative		Caffeine	48	92% (44/48)	188	93% (175/188)
NIR	SNV	SIMCA	Nicotine	48	92% (44/48)	188	95% (178/188)
	Autoscaling		CBD	48	100%	183	93% (170/183)
	Autoscaling		Vit E	48	96% (46/48)	183	93% (170/183)
	Autoscaling		Taurine	48	100%	188	93% (175/188)
	SNV		Caffeine	48	83% (40/48)	183	95% (173/183)

Table 3. Comprehensive overview of the performance of MIR and NIR, data pretreatment methods, and PLS-DA in the classification of various target molecules.

Spectroscopic Technique	Data Pretreatment	Chemometric Technique	Target Molecule	Samples in External Test Set	Correct Classification Rate (External Test Set)	Samples in Training Set	Correct Classification Rate (Cross Validation)
MIR	1st derivative	PLS-DA	Nicotine	48	92% (44/48)	188	82% (155/188)
	Autoscaling		CBD	48	100%	188	94% (177/188)
	2nd derivative		Vit E	48	100%	188	94% (177/188)
	Autoscaling		Taurine	48	96% (46/48)	188	100%
	2nd derivative		Caffeine	48	94% (45/48)	188	93% (175/188)
NIR	Autoscaling	PLS-DA	Nicotine	48	83% (40/48)	183	95% (173/183)
	Autoscaling		CBD	48	100%	183	93% (170/183)
	Autoscaling		Vit E	48	96% (46/48)	183	95% (173/183)
	1st derivative		Taurine	48	100%	188	93% (175/188)
	Autoscaling		Caffeine	48	83% (40/48)	183	91% (167/183)
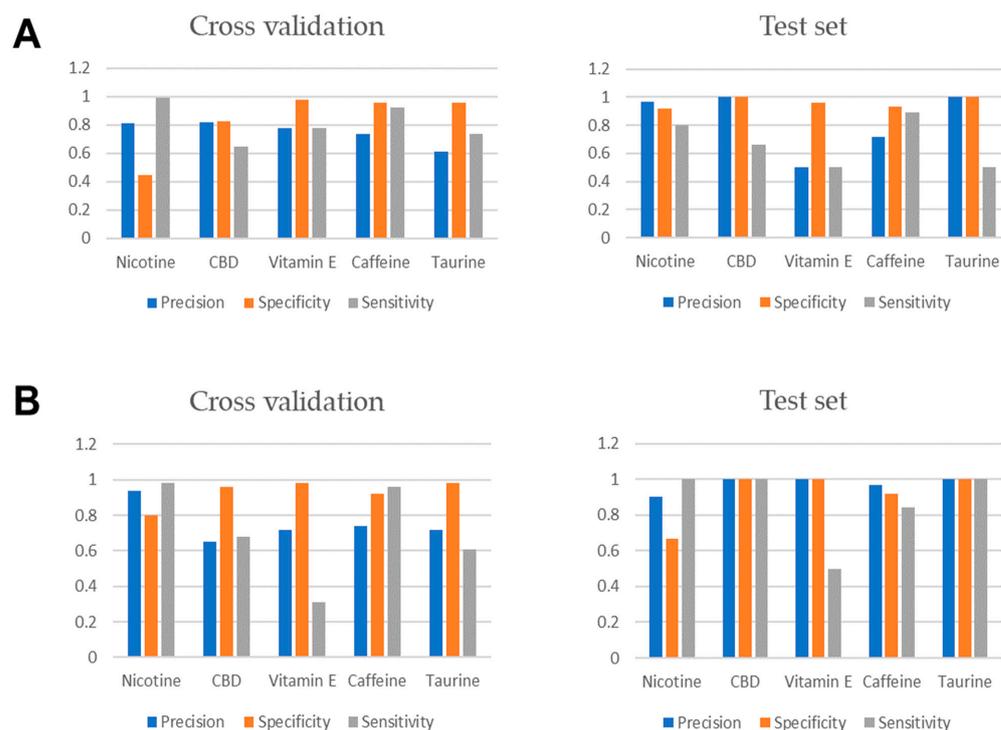
Figure 2. Classification statistics for cross-validation and test set for SIMCA model with (**A**) MIR and (**B**) NIR spectroscopy.
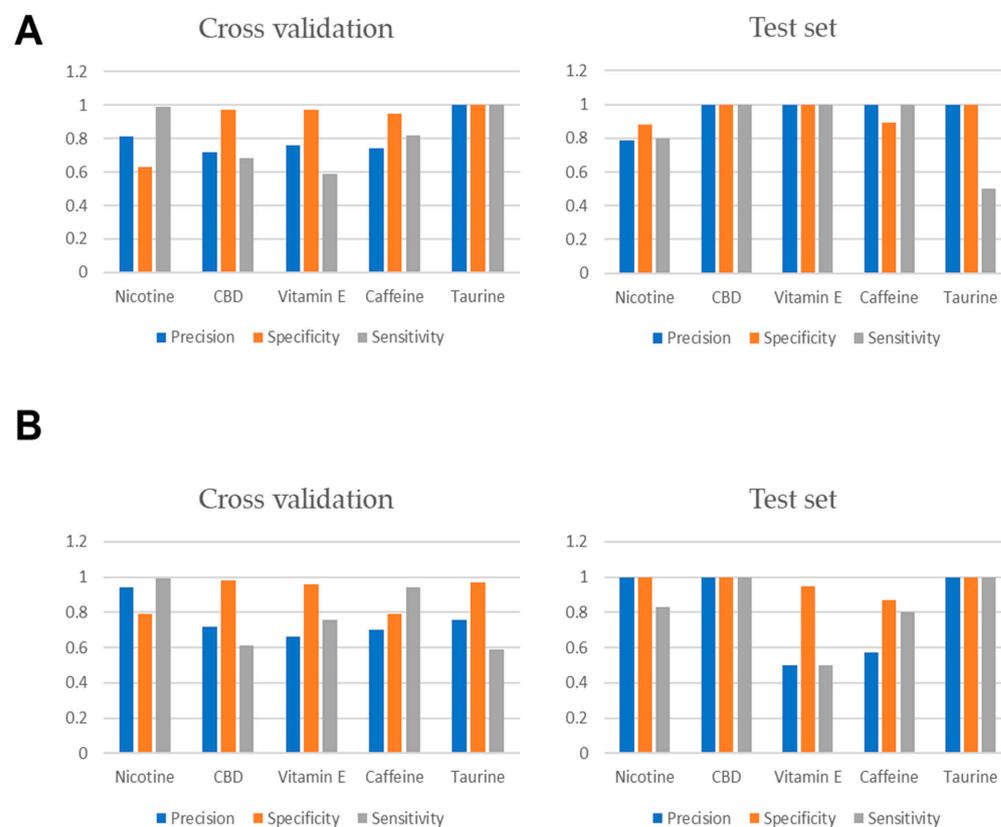
Figure 3. Classification statistics for cross-validation and test set for PLS-DA model with (**A**) MIR and (**B**) NIR spectroscopy.

4.1.1. SIMCA

MIR and NIR spectra were obtained for all 236 samples. Prior to modeling, various data preprocessing methods such as auto scaling, SNV, and 1st and 2nd derivative were explored. These preprocessing techniques aimed to optimize the data and to improve their interpretation. The target molecules analyzed in the study were nicotine, CBD, vitamin E, taurine, and caffeine. These molecules were chosen based on their relevance from a public health and legal point of view [5]. The total number of samples was always 236, but the number of samples positive for the target compound changed, since for some compounds more positive samples were present in the data set than for others. After the data preprocessing steps were completed, SIMCA was applied to both the MIR and NIR data and to each compound separately (binary modeling). Subsequently, the dataset was divided into two distinct sets: the training set and the test set. The training set played a dual role, serving to construct and build the models and the subject for internal validation, employing 10-fold cross-validation. After model construction and internal validation, the selected test set, which remained completely separate from the training data, was dedicated to external validation. This external validation allowed for an unbiased evaluation of the models' performance on new, unseen spectral data, ensuring the reliability and robustness of the developed models when predicting spectra from unknown samples. The models provided a means to identify the similarities and dissimilarities between the spectra of the known classes, allowing for the classification of unknown samples, based on their spectral characteristics.

Globally, SIMCA, with various data pretreatment methods, achieved correct classification rates ranging from 83% to 100% for the external test set and from 81% to 97% for cross-validation, as shown in Table 1.

For nicotine, the most effective model based on MIR was achieved through utilizing the 2nd derivative as a data pretreatment technique. This model yielded a 83% correct classification rate on the external test set, while six samples were classified as false positive and two as false negative. There was a cross-validation correct classification rate (CCR) of 81%, with 35 samples misclassified. Among these misclassifications, one was a false negative, while the rest were false positives. Although the specificity during cross-validation was observed to be relatively low (0.45), the precision values (as shown in Figure 2) remained acceptable. Upon a closer examination of misclassifications for the test set and training set, it was determined that the false negative could be attributed to the low nicotine dosage in this sample (2 mg/mL). Regarding the other false positives, no definitive explanation could be identified, and they might be attributed to random modeling errors. 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. Shifting to NIR spectroscopy with SNV data pretreatment and the SIMCA technique for nicotine classification, the external test set achieved an accuracy of 92%, with only four samples misclassified as false positive. Cross-validation accuracy was notably higher at 95%, with 10 misclassified samples. Among these misclassifications, eight samples were categorized as false positives and two as false negatives. Notably, these false negative samples exhibited concentrations of 6 mg/mL and 25 mg/mL, indicating potential contributions from factors such as sample matrix or instrumental variations. Explanations are currently unavailable for the instances in either the test set or the cross-validation set that were incorrectly identified as positives, but are actually negative.

Moving on to CBD, the use of autoscaling of MIR spectra led to the most successful SIMCA model, achieving a cross-validation CCR of 95%, with only 10 misclassified samples. The test set performance reached an impressive 98%, with one sample being misclassified, which was classified as false negative, having a lower concentration of 0.05 mg/mL. Especially for the external validation, very good features for precision, specificity and sensitivity (Figure 2) could be obtained. The slightly lower performance during cross-validation could potentially be attributed to the relatively limited number of positive samples within the dataset. Delving into misclassifications, it became evident that seven samples were inaccu-

rately categorized as false negatives, and three as false positives. Intriguingly, there were six instances where both low- and high-concentration samples were misclassified, spanning from 0.05 mg/mL to 0.2 mg/mL. This highlights the presence of errors within the generated model. False classifications in the model are often due to the fact that samples show a difference towards the samples used for training. This is called deviation from the model. Otherwise, the model looks for the best fit of the training samples and allows margins of mistakes called random modeling errors. Or, the sample can also be an outlier, in which case it is not the model causing the error, but the sample or the measurement (e.g., spectral error due to changes in settings, temperature variation etc...). For CBD classification using NIR spectroscopy and autoscaling data pretreatment, a perfect classification rate of 100% was achieved on the external test set, accurately categorizing all 48 samples. Cross-validation yielded a CCR of 93%, with 170 out of 183 samples correctly classified. An analysis of misclassifications showed that seven samples were misclassified as positive, while the rest was falsely categorized as negative. The concentration of misclassified samples ranged from 0.05 mg/mL to 0.2 mg/mL.

Regarding vitamin E, external testing revealed misclassification in merely four samples, in which two samples were misclassified as positive and two as negative; the highest CCR of 97% during cross-validation was attained using the 1st derivative of Mid-IR spectra. This indicated that the model accurately identified vitamin E in 182 out of 188 training set samples. An examination of misclassification patterns for each class unveiled an estimated three samples erroneously categorized as false negatives and three as false positives. No evident explanation could be discerned for these instances for both the test set and cross-validation, as the concentration of misclassified samples ranged from 0.15 mg/mL to 0.6 mg/mL. Utilizing NIR spectroscopy with autoscaling data pretreatment and the SIMCA technique, the classification of vitamin E demonstrated an impressive external test set classification rate of 96%, accurately identifying 46 out of 48 samples, while 2 samples were misclassified as negative, having a concentration of 0.15 mg/mL. Cross-validation maintained model reliability, with a correct classification rate of 93%, yet 13 samples were misclassified as false negatives and 5 as false positives. The concentration of misclassified samples ranged from 0.1 mg/mL to 0.25 mg/mL, suggesting misclassification of low-dose samples as negatives. No definitive explanations were found for the false-positive classifications, indicating potential modeling errors.

In the classification scenario for taurine, the application of MIR spectroscopy with autoscaling data pretreatment and SIMCA resulted in an external test set accuracy of 96%, with only two misclassified samples; these were misclassified as negative, both having a lower concentration of 0.187 mg/mL. The cross-validation accuracy, slightly lower at 88%, accurately classified 166 out of 188 samples. A closer examination of misclassifications revealed that eight samples were misclassified as negatives, with concentrations ranging from 0.187 mg/mL to 0.4 mg/mL. Additionally, around five samples were erroneously classified as positive. These misclassifications might stem from modeling errors. Taurine classification involving NIR spectroscopy with autoscaling data pretreatment and the SIMCA technique resulted in a perfect classification rate of 100% for the external test set, but thirteen samples were misclassified during cross-validation. Misclassifications included eight samples categorized as false negatives and five as false positives. No conclusive explanations were identified, suggesting they could be random modeling errors.

The study aimed to classify caffeine, using MIR spectroscopy with 1st-derivative data pretreatment and SIMCA. The external test set exhibited a classification accuracy of 92%, accurately identifying 44 out of 48 samples; 3 samples were identified as false positive, for which no explanation was found, and 1 as false negative, having a concentration of 0.07 mg/mL. Cross-validation demonstrated a similar accuracy rate of 93%, with 175 out of 188 samples correctly classified. While scrutinizing the misclassifications, it was observed that four samples were misclassified as false negatives and nine as false positives. The false-negative instances were associated with low concentrations,

ranging from 0.06 mg/mL to 0.1 mg/mL. Lastly, for caffeine classification involving NIR spectroscopy, the external test set achieved a correct classification rate of 83.3%, accurately identifying 40 out of 48 samples, with only 4 samples misclassified as negative. Cross-validation maintained a similar level of accuracy, at 95%, with two samples misclassified as negative and eight as positive. No clear explanations emerged for these misclassifications for both test set and cross-validation.

Analyzing the overall results, the combination of NIR spectroscopy and SIMCA produced robust models for identifying the majority of additives, showcasing superior performance across the targeted molecules. However, when examining caffeine, MIR proved to be the preferred method. Consequently, comprehensive screening for the five targeted molecules mandates the incorporation of both MIR and NIR techniques, to ensure the best approach.

4.1.2. PLS-DA Model

PLS-DA was employed. PLS-DA is a specialized variant of PLS, designed for classification tasks, enabling the use of multivariate analysis techniques to classify samples based on their spectral characteristics.

The study evaluated the classification performance of different target molecules using two different spectroscopic techniques: MIR and NIR, as shown in Table 3 and Figure 3.

Employing MIR spectroscopy alongside 1st-derivative data pretreatment, the PLS-DA chemometric technique achieved a 92% correct classification rate on the external test set for nicotine. This achievement involved the precise classification of 44 out of 48 samples, while 4 samples were misclassified as positive. However, a clarification could not be located within the available information. The process of cross-validation yielded a correct classification rate of 82%, with 155 out of 188 samples being accurately categorized. Delving deeper, the analysis unveiled that 31 samples were misclassified as positive, whereas 2 were inaccurately identified as false negatives. This classification challenge seemed to revolve around specific high-dosage nicotine samples, with instances such as a 10 mg/mL and a 25 mg/mL sample erroneously labeled as false negatives. Contrarily, the utilization of NIR spectroscopy, coupled with autoscaling data pretreatment, showed an external test set classification rate of 83%, accurately categorizing 40 out of 48 samples, while 8 samples were misclassified as positive. The available data did not yield an explanation for these misclassifications. However, cross-validation demonstrated good proficiency, having a rate of 95%, ensuring the correct classification of 173 out of 183 samples. Among these, nine samples were unfortunately classified as false positives, and a sole sample was falsely labeled as negative. The variance in success rates underscores the role of spectroscopic techniques and data preprocessing methods in nicotine classification.

Shifting the focus to CBD classification, the implementation of MIR spectroscopy in tandem with autoscaling data pretreatment exhibited remarkable robustness, boasting a perfect 100% classification rate on the external test set, and appropriately categorizing all 48 samples. Further, cross-validation attained a commendable 94% correct classification rate, successfully classifying 177 out of 188 samples. However, six samples were categorized as false negatives, with concentrations spanning from low to high, ranging from 0.05 mg/mL to 0.25 mg/mL. Moreover, an additional five samples were falsely identified as positive. Conversely, the application of NIR spectroscopy revealed equally impressive results, achieving a 100% external test set classification rate and a 93% rate during cross-validation. Notably, 13 samples were misclassified during cross-validation, comprising 8 false negatives, primarily associated with low concentrations (0.05 mg/mL–0.1 mg/mL), and 5 false positives. Despite scrutiny, a concrete explanation for the false-positive classifications remained elusive.

For the assessment of vitamin E, MIR spectroscopy employing 2nd-derivative data pretreatment yielded compelling outcomes. Remarkably, the external test set showcased a flawless classification rate of 100%, expertly categorizing all 48 samples. For cross-validation, maintaining a good 93.6% correct classification rate, 177 out of

188 samples could be correctly identified. The misclassification analysis divulged that nine samples were falsely labeled as negatives, their concentrations spanning from 0.1 mg/mL to 0.6 mg/mL, while four samples were inaccurately identified as false positives. In contrast, when NIR spectroscopy with autoscaling data pretreatment was employed, the external test set achieved a respectable 96% classification rate, resulting in only two misclassified samples, with one identified as false positive and one as false negative. The false-negative sample was found to have a low concentration of 0.1 mg/mL. Yet, during cross-validation, the performance remained noteworthy, at 95%, although 10 samples fell into the category of misclassification. These instances comprised seven false negatives, ranging from 0.1 mg/mL to 0.4 mg/mL, and three false positives. Notably, plausible explanations for these classifications remained elusive.

For taurine with MIR spectroscopy and autoscaling data pretreatment, PLS-DA attained a 96% correct classification rate during the external test set evaluation, precisely identifying 46 out of 48 samples, with 2 samples being misclassified as false negative. An absence of clarification was evident from the information regarding the misclassification. Cross-validation achieved a remarkable 100% accuracy, underscoring the model's dependability in classifying taurine samples. Alternatively, utilizing NIR spectroscopy in conjunction with 1st-derivative data pretreatment, the model achieved a 100% classification rate on the external test set. During cross-validation, the model's prowess remained evident, with a 93% correct classification rate, yet 13 samples were misclassified. Among these, nine samples were inaccurately labeled as false negatives, featuring concentrations ranging from 0.187 mg/mL to 0.6 mg/mL, while four samples were wrongly categorized as false positives. Despite exploration, definitive explanations for these misclassifications remained elusive, pointing towards potential modeling discrepancies.

For the classification of caffeine, MIR spectroscopy, in conjunction with 2nd-derivative data pretreatment of the external test set, achieved 94% correct classification rate, categorizing 45 out of 48 samples, which were identified as false positives; no explanation could be discovered within the information provided for this misclassification. Cross-validation demonstrated a slightly reduced accuracy of 92.5%, expertly classifying 175 out of 188 samples. Closer inspection of misclassification patterns unveiled five samples misclassified as false negatives and eight as false positives. Turning to NIR spectroscopy, alongside autoscaling data pretreatment, the model garnered an 83% correct classification rate on the external test set, while six samples were classified as false positives and two as false negatives, with concentrations of 0.06 mg/mL and 0.08 mg/mL. Remarkably, cross-validation yielded a 91% accuracy rate, accurately classifying 167 out of 183 samples. Among these, 10 samples were wrongly classified as false positives, and 6 as false negatives. Evidently, the misclassification of false negatives in both spectroscopic techniques could be attributed to the low dosage of caffeine in these instances, with concentrations ranging from 0.06 mg/mL to 0.1 mg/mL. Due to the intricate matrix composition, the model encountered challenges in accurately discerning caffeine presence in these particular samples. Conversely, the false-positive classifications suggested a possible connection with modeling inaccuracies.

Based on the results in Table 3, it can be said that, in general, very good performing models were obtained with both MIR and NIR, as correct classification rates were almost all above 90%. However, NIR combined with PLS-DA resulted in a correct classification rate of only 83% for both nicotine and caffeine, which is not bad, but clearly less performant than the nicotine and caffeine models based on MIR. From these results, it is clear that MIR is the preferred technique for characterizing the five targeted additives in e-liquids when PLS-DA is used for modeling.

4.2. Quantitative Models for Nicotine

Since problems were reported with sample conformity concerning the nicotine content of nicotine containing e-liquids, it was explored as to whether a quantitative PLS model could be constructed for the nicotine content.

Only the samples containing nicotine (real + spiked ones) were retained to create quantitative models. In a first step, an attempt was made to build PLS regression models using only the spiked samples and using the real samples as test set (results not shown). Although these models showed promising results based on cross-validation, the prediction of the real samples was not satisfactory, pointing at a lack of variability in the training set. Therefore, all nicotine-positive samples (real + spiked ones) were mixed, resulting in a sample set of, in total, 176 samples. This sample set was then split into training and test sets, using the duplex algorithm, resulting in a training set of 147 samples and a test set of 29 samples. Both sets contained spiked, as well as real, samples, and the whole range of nicotine content was covered in the sample set (1–25 mg/mL).

Based on the NIR spectra, no meaningful regression models could be obtained. This is probably due to the high variety of the matrices, and especially the differences in color between the samples. It is generally known that NIR suffers from interference due to coloring agents. Therefore, it was decided to focus on MIR.

PLS models were constructed using the MIR spectral data and exploring different data pretreatment techniques. The most optimal was obtained using 10 PLS factors and a combination of SNV and the 1st derivative as data pretreatment. However, when predicting the 29 samples in the test set, a root-mean-square error of prediction (RMSEP) of 5.23 and an R^2 between real and predicted values of only 0.46 was obtained. These results are insufficient to allow MIR to be used as a quantitative tool for nicotine-content determination. This is probably due to the complexity and the high variability in matrices represented by the e-liquids. Indeed, even though the basis of most samples is a mixture of propylene glycol and glycerol, the ratio in which they are present can differ. Also, the different e-liquids contain a huge range of different additives, especially colorants and aromatic molecules. This hypothesis is confirmed by the fact that good predictive models (results not shown) could be obtained using only the spiked samples, so with a sample set with limited variability in the matrix. Further, more complex chemometric techniques like orthogonal-PLS, artificial neural networks and support vector machines were explored, but no significant improvement towards the PLS model could be observed.

This research points to the fact that quantification using classical FTIR spectroscopy might not be feasible, due to the high variety in the samples in a real-life setting, like for example, in the case of inspections.

5. Conclusions

Spectroscopic techniques (MIR and NIR) with chemometric techniques (SIMCA and PLS-DA) can be highly effective in detecting different target molecules in e-liquids. Previous research suggests that both MIR and NIR spectroscopy can distinguish between nicotine and non-nicotine e-liquids without sample preparation. These methods have potential applications in custom screening and label-compliance checks, enhancing consumer information on nicotine content [40]. In this research, MIR and NIR spectroscopy was explored in combination with chemometric modeling which allowed for the identification and detection of nicotine and different additives in the e-liquids. This study shows the potential of spectroscopic analysis in the preliminary screening of e-liquids. In the case of MIR spectroscopy, using autoscaling data pretreatment with PLS-DA yielded exceptional classification rates of 100% for CBD and vitamin E, indicating that this combination is well suited for accurately identifying these particular target molecules in the samples. Similarly, for NIR spectroscopy, autoscaling data pretreatment with PLS-DA achieved perfect classification rates of 100% for CBD and taurine. This underscores the robustness and reliability of the autoscaling data pretreatment combined with PLS-DA in NIR spectroscopy. For the other targeted molecules in this study, optimized classification rates could be obtained.

Overall when looking to all targeted molecules together, MIR combined with PLS-DA resulted in the best classification. Therefore, if only one technique should be chosen, MIR is to be preferred. It should also be highlighted that at this stage of initial screening, false-negative samples are the major issue, since false-positive samples will be seized and

send for further analysis to a laboratory where they will be declared compliant with the legislation, while false-negative samples would be released to be sold to consumers.

However, quantification using classical FTIR spectroscopy might not be feasible, due to the high variety in the samples in a real life setting, like for example in the case of inspections.

In practice, the presented approach could be used by inspectors and custom services to quickly check if nicotine is present (comparison with the label) or non-allowed additives are present in the e-liquids sold at the different selling points, especially when the approach is transferred to portable devices.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/chemosensors12010009/s1, Table S1: Concentration and Attributes of Substances in Sample Analysis.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The information presented in this research is not accessible to the general public because of certain limitations, such as concerns related to privacy and ethics.

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

References

1. Felberbaum, M. So Far, Anything Goes in Ads for E-Cigarettes. 2013. Available online: http://www.utsandiego.com/news/2013/Aug/12/tp-so-far-anything-goes-in-ads-for-e-cigarettes (accessed on 18 October 2023).
2. Rachel, G.; Benowitz, N.; Glantz, S.A. E-cigarettes: A scientific review. *Circulation* **2014**, *129*, 1972–1986.
3. Hajek, P.; Phillips-Waller, A.; Przulj, D.; Pesola, F.; Myers Smith, K.; Bisal, N.; Li, J.; Parrott, S.; Sasieni, P.; Dawkins, L.; et al. A randomized trial of e-cigarettes versus nicotine-replacement therapy. *NEJM* **2019**, *380*, 629–637. [CrossRef] [PubMed]
4. Wadgave, U.; Nagesh, L. Nicotine replacement therapy: An overview. *Int. J. Health Sci.* **2016**, *10*, 425. [CrossRef]
5. Buettner-Schmidt, K.; Miller, D.R.; Balasubramanian, N. Electronic cigarette refill liquids: Child-resistant packaging, nicotine content, and sales to minors. *J. Pediatr. Nurs.* **2016**, *31*, 373–379. [CrossRef] [PubMed]
6. Goniewicz, M.L.; Kuma, T.; Gawron, M.; Knysak, J.; Kosmider, L. Nicotine levels in electronic cigarettes. *NTR* **2012**, *15*, 158–166. [CrossRef] [PubMed]
7. Hadwiger, M.E.; Trehy, M.L.; Ye, W.; Moore, T.; Allgire, J.; Westenberger, B. Identification of amino-tadalafil and rimonabant in electronic cigarette products using high pressure liquid chromatography with diode array and tandem mass spectrometric detection. *J. Chromatogr. A* **2010**, *1217*, 7547–7555. [CrossRef] [PubMed]
8. European Parliament and the Council of the European Union. Directive 2014/40/EU. *Off. J. Eur. Union* **2014**, *2014*, 127/1–127/38.
9. Hannisdal, A.; Mikkelsen, Ø.; Schrøder, K.H. Analysis of nicotine in antismoking pharmaceutical products by differential pulse polarography and voltammetry. Collect. Czechoslov. *Chem. Commun.* **2007**, *72*, 1207–1213.
10. Rai, M.; Ramachandran, K.N.; Gupta, V.K. Spectrophotometric method for the determination of total tobacco alkaloids and nicotine. *Analyst* **1994**, *119*, 1883–1885. [CrossRef]
11. Clayton, P.M.; Vas, C.A.; Bui, T.T.; Drake, A.F.; McAdam, K. Spectroscopic studies on nicotine and nornicotine in the UV region. *Chirality* **2013**, *25*, 288–293. [CrossRef]
12. Aszyk, J.; Woźniak, M.K.; Kubica, P.; Kot-Wasik, A.; Namieśnik, J.; Wasik, A. Comprehensive determination of flavouring additives and nicotine in e-cigarette refill solutions. Part II: Gas-chromatography–mass spectrometry analysis. *J. Chromatogr. A.* **2017**, *1517*, 156–164. [CrossRef] [PubMed]
13. Beć, K.B.; Grabska, J.; Huck, C.W. Principles and applications of miniaturized near-infrared (NIR) spectrometers. *Chem. Eur. J.* **2021**, *27*, 1514–1532. [CrossRef] [PubMed]
14. Deconinck, E.; Sacre, P.Y.; Courselle, P.; De Beer, J.O. Chemometrics and chromatographic fingerprints to discriminate and classify counterfeit medicines containing PDE-5 inhibitors. *Talanta* **2012**, *100*, 123–133. [CrossRef] [PubMed]

15. Deconinck, E.; Bothy, J.L.; Desmedt, B.; Courselle, P.; De Beer, J.O. Detection of whitening agents in illegal cosmetics using attenuated total reflectance-infrared spectroscopy. *JPBA* **2014**, *98*, 178–185. [CrossRef] [PubMed]

16. Deconinck, E.; Van Campenhout, R.; Aouadi, C.; Canfyn, M.; Bothy, J.L.; Gremeaux, L.; Blanckaert, P.; Courselle, P. Combining attenuated total reflectance-infrared spectroscopy and chemometrics for the identification and the dosage estimation of MDMA tablets. *Talanta* **2019**, *195*, 142–151. [CrossRef] [PubMed]

17. Kharbach, M.; Yu, H.; Kamal, R.; Barra, I.; Marmouzi, I.; Cherrah, Y.; Alaoui, K.; Bouklouze, A.; Vander Heyden, Y. New insights into the Argan oil categories characterization: Chemical descriptors, FTIR fingerprints, and chemometric approaches. *Talanta* **2021**, *225*, 122073. [CrossRef] [PubMed]

18. Liu, H.; Chen, Y.; Shi, C.; Yang, X.; Han, D. FT-IR and Raman spectroscopy data fusion with chemometrics for simultaneous determination of chemical quality indices of edible oils during thermal oxidation. *LWT* **2020**, *119*, 108906. [CrossRef]

19. Bunaciu, A.A.; Aboul-Enein, H.Y.; Fleschin, S. Recent applications of fourier transform infrared spectrophotometry in herbal medicine analysis. *Appl. Spectrosc. Rev.* **2011**, *46*, 251–260. [CrossRef]

20. Kim, S.W.; Ban, S.H.; Chung, H.; Cho, S.; Chung, H.J.; Choi, P.S.; Yoo, O.J.; Liu, J.R. Taxonomic discrimination of flowering plants by multivariate analysis of Fourier transform infrared spectroscopy data. *Plant Cell Rep.* **2004**, *23*, 246–250. [CrossRef]

21. Zin, N.B.M.; Azemin, A.; Rodi, M.M.M.; Rashid, Z.M.; Mohd, K.S. Application of FTIR fingerprints coupled with chemometric for comparison of stingless bee propolis from different extraction methods. *MJFAS* **2019**, *15*, 350–355.

22. Blanco, M.; Villarroya, I.N.I.R. NIR spectroscopy: A rapid-response analytical tool. TrAC. *Trends Anal. Chem.* **2002**, *21*, 240–250. [CrossRef]

23. Tsuchikawa, S.; Ma, T.; Inagaki, T. Application of near-infrared spectroscopy to agriculture and forestry. *Anal. Sci.* **2022**, *38*, 635–642. [CrossRef] [PubMed]

24. Dumont, E.; De Bleye, C.; Sacré, P.Y.; Netchacovitch, L.; Hubert, P.; Ziemons, E. From near-infrared and Raman to surface-enhanced Raman spectroscopy: Progress, limitations and perspectives in bioanalysis. *Bioanalysis* **2016**, *8*, 1077–1103. [CrossRef] [PubMed]

25. Barhdadi, S.; Moens, G.; Canfyn, M.; Vanhee, C.; Desmedt, B.; Courselle, P.; Rogiers, V.; Vanhaecke, T.; Deconinck, E. Impact of the revised European Tobacco Product Directive on the quality of e-cigarette refill liquids in Belgium. *NTR* **2021**, *23*, 227–234.

26. Joan, E.G. Construction and validation of quantification methods for determining the cannabidiol content in liquid pharma-grade formulations by means of near-infrared spectroscopy and partial least squares regression. *Med. Cannabis Cannabinoids* **2019**, *2*, 43–55.

27. Man, Y.B.C.; Ammawath, W.; Mirghani, M.E.S. Determining α-tocopherol in refined bleached and deodorized palm olein by Fourier transform infrared spectroscopy. *Food Chem.* **2005**, *90*, 323–327. [CrossRef]

28. Triebel, S.; Sproll, C.; Reusch, H.; Godelmann, R.; Lachenmeier, D.W. Rapid analysis of taurine in energy drinks using amino acid analyzer and Fourier transform infrared (FTIR) spectroscopy as basis for toxicological evaluation. *Amino Acids* **2007**, *33*, 451–457. [CrossRef] [PubMed]

29. Naveen, P.; Lingaraju, H.B.; Deepak, M.; Medhini, B.; Prasad, K.S. Method development and validation for the determination of caffeine: An alkaloid from coffea arabica by high-performance liquid chromatography method. *Pharmacogn. Res.* **2018**, *10*, 88.

30. 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.

31. McClure, W.F. Near-infrared spectroscopy the giant is running strong. *Anal. Chem.* **1994**, *66*, 42A–53A. [CrossRef]

32. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777. [CrossRef]

33. Snee, R.D. Validation of regression models: Methods and examples. *Technometrics* **1977**, *19*, 415–428. [CrossRef]

34. Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J.K.; Holmes, E.; Trygg, J. OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **2006**, *20*, 341–351. [CrossRef]

35. De Maesschalck, R.; Candolfi, A.; Massart, D.L.; Heuerding, S. Decision criteria for soft independent modelling of class analogy applied to near infrared data. *Chemom. Intell. Lab. Syst.* **1999**, *47*, 65–77. [CrossRef]

36. 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 A Mol. Biomol. Spectrosc.* **2018**, *189*, 282–290. [CrossRef] [PubMed]

37. 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 A Mol. Biomol. Spectrosc.* **2022**, *268*, 120654. [CrossRef] [PubMed]

38. Kemsley, E.K. Discriminant analysis of high-dimensional data: A comparison of principal components analysis and partial least squares data reduction methods. *Chemom. Intell. Lab. Syst.* **1996**, *33*, 47–61. [CrossRef]

39. Chevallier, S.; Bertrand, D.; Kohler, A.; Courcoux, P. Application of PLS-DA in multivariate image analysis. *J. Chemom.* **2006**, *20*, 221–229. [CrossRef]

40. Deconinck, E.; Bothy, J.L.; Barhdadi, S.; Courselle, P. Discriminating nicotine and non-nicotine containing e-liquids using infrared spectroscopy. *JPBA* **2016**, *120*, 333–341. [CrossRef]