

Review

# Light and Shadow in Near-Infrared Spectroscopy: A Powerful Tool for *Cannabis sativa* L. Analysis

María del Carmen Díaz-Liñán , Verónica Sánchez de Medina, Carlos Ferreiro-Vera   
and María Teresa García-Valverde \*

Phytoplant Research S.L.U., The Science and Technology Park of Córdoba-Rabanales 21, Astrónoma Cecilia Payne Street, Centauro Building, B-1, 14014 Cordoba, Spain; mc.diaz@phytoplant.es (M.d.C.D.-L.); v.sanchez@phytoplant.es (V.S.d.M.); c.ferreiro@phytoplant.es (C.F.-V.)

\* Correspondence: mt.garcia@phytoplant.es

**Abstract:** *Cannabis sativa* L. is an ancient cultivar that has found applications in various fields, e.g., medicine, due to its beneficial effects. However, due to its psychotropic effects, the regulation of this cultivar has increased throughout the decades. In this context, the need for rapid and reliable analytical methods to ensure the quality control of Cannabis cultivars has become of extreme importance. NIRS has arisen as a powerful tool in this field due to its multiple advantages, e.g., non-destructive, rapid, and cost-effective. In this article, the chemometric techniques commonly employed in NIRS method development are described, along with their application for the analysis of Cannabis samples. Regarding qualitative methods, different mathematical treatments and classification models are explained. As for quantitative methods, the representative linear and non-linear modelling techniques applied for the development of prediction equations are described, alongside their application in the Cannabis field. To the best of our knowledge, this is the first time this type of review is written, since there are several articles which address cannabinoid determination, but the main purpose of this review is to enhance the potential of NIRS over the traditional techniques employed for the analysis of Cannabis samples.

**Keywords:** Cannabis; near-infrared spectroscopy; chemometrics; cannabinoids



**Citation:** Díaz-Liñán, M.d.C.; Sánchez de Medina, V.; Ferreiro-Vera, C.; García-Valverde, M.T. Light and Shadow in Near-Infrared Spectroscopy: A Powerful Tool for *Cannabis sativa* L. Analysis. *AppliedChem* **2023**, *3*, 526–545. <https://doi.org/10.3390/appliedchem3040033>

Academic Editor: Alessandra Biancolillo

Received: 17 October 2023

Revised: 17 November 2023

Accepted: 21 November 2023

Published: 29 November 2023



**Copyright:** © 2023 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/>).

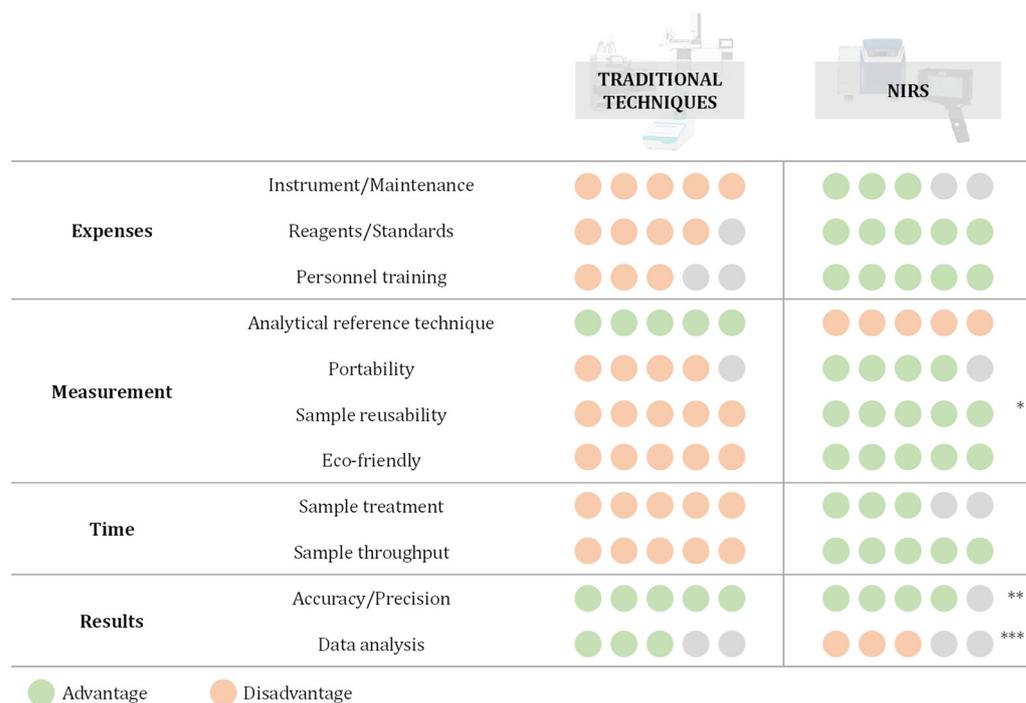
## 1. Introduction

*Cannabis sativa* L. is one of the oldest known cultivars and it has been exploited for its versatility and advantages in various fields, such as medicine and food or textile industry. Despite its negative connotation related to its recreational consumption, Cannabis has been used as a potential treatment or palliative measure for various illnesses and medical conditions such as chronic pain, epilepsy, multiple sclerosis, cancer, glaucoma, and neurodegenerative disorders, like Parkinson's and Huntington's disease, among others [1]. In this sense, the medical Cannabis industry focuses on the cultivation, processing, and distribution of Cannabis and Cannabis-derived products for medical purposes. It includes the production of pharmaceutical-grade Cannabis products, such as oils, tinctures, capsules, and more, which are used to treat various medical conditions. The medicinal Cannabis industry is subject to strict regulations and often requires specific licenses and certifications. Cannabis produces at least 144 distinct phytocannabinoids, the main constituents of the plant, which represent the potential cause of this medical benefit. The most prevalent cannabinoids are the non-psychoactive tetrahydrocannabinolic and cannabidiolic acid (THCA and CBDA, respectively); these transform into the biologically active neutral forms,  $\Delta^9$ -tetrahydrocannabinol ( $\Delta^9$ -THC) and cannabidiol (CBD), after undergoing the decarboxylation process. In recent years, there has been a growing interest in the scientific examination of various phytocannabinoids, including CBN (cannabinol), CBC (cannabichromene), and CBG (cannabigerol), with regard to their potential applications in the fields of medicine, cosmetics, and other purposes [2].

On the other hand, the non-psychoactive species of *Cannabis sativa* L., industrial hemp, is characterized by containing minimal concentrations of  $\Delta^9$ -THC. Hemp fibres can be used in the production of textiles, paper, and building materials, while hemp seeds can be processed into food products and dietary supplements. The common agricultural policy (CAP) of the European Union provides subsidies for the cultivation of these specific strains, as long as their  $\Delta^9$ -THC content does not exceed 0.2%. In this sense, according to Regulation (EU) No. 1307/2013 of the European Parliament, the verification of the  $\Delta^9$ -THC content must be carried out to prevent the cultivation of illicit drug-type Cannabis in hemp fields. However, this limitation may vary depending on the country, increasing up to 1% in Czech Republic [3].

For this reason, the importance of the quality control of the Cannabis industry has increased dramatically and numerous accredited testing laboratories have emerged in the last decade [4]. The complexity of the Cannabis sample challenges the quality control of this matrix, since many parameters may be determined. In this sense, analysis could be classified considering both favorable and unfavorable parameters. In the first group, different compounds such as cannabinoids, terpenes, flavonoids, alkaloids, or sterols, among others may be quantified [5–9]. All these parameters will give beneficial information of the vegetal material intended to be used as a medicinal product, since they provide certain characteristics to the Cannabis plant, such as flavour or smell. On the other hand, certain parameters such as pesticides, mycotoxins, or heavy metals are categorized as adverse factors, and their presence may need to be limited to specific quantities [10–13]. Therefore, they are also essential considerations for the quality control of Cannabis.

Traditionally, the determination of cannabinoids in Cannabis samples, has been performed via chromatographic techniques, such as gas (GC) [14–16] or liquid chromatography (LC) [17–19] coupled to different detectors, namely diode array (DAD), mass spectrometry (MS) and flame ionization detector (FID), among others [20]. These techniques present a series of advantages that make them appropriate for the analysis of cannabinoids and other minor analytes such as terpenes or flavonoids, which are depicted in Figure 1. LC and GC provide excellent separation capabilities, allowing the analysis of complex mixtures and permitting the determination of cannabinoids and other compounds with high selectivity and sensitivity. This is especially reinforced when coupled with detectors like mass spectrometry, which makes them suitable for the detection and quantification of compounds, even at low concentrations [21]. However, the achievement of superior analytical performance using these techniques involves the use of hazardous solvents such as acetonitrile, methanol, and/or hexane (Figure 1). On the other hand, it should be noted that the expenses associated with the acquisition, maintenance, and operation of the LC and GC equipment are significant, accompanied by the daily operation of these instruments that requires highly skilled personnel [22]. As for GC, this approach is not suitable for the direct determination of the acidic form of cannabinoids since high temperatures (up to 300 °C) are reached throughout the analysis and these compounds decarboxylate to their neutral form when heated due to their thermolabile nature. Additionally, some studies have demonstrated the potential degradation of cannabinoids in the chromatographic injector port, leading in inaccuracies [23]. For this reason, derivatization of the acidic cannabinoids is mandatory prior GC analysis, which can be time-consuming. Furthermore, temperature can also lead to conversion reactions [24–28], thus resulting in an inappropriate determination of the analyte. Further techniques, such as thin-layer chromatography (TLC) or nuclear magnetic resonance (NMR) have also been employed for the determination of cannabinoids [29,30]. However, these techniques are not suitable for quantification due to the limitations they present, i.e., elevated cost and maintenance in the case of NMR, or low sensitivity in the case of TLC.



\* After obtaining the predictive equations, NIRS can measure the sample directly without causing its destruction

\*\* Depending on the quality of the information provided by the reference techniques

\*\*\* Developing the predictive equations is a time-consuming process

**Figure 1.** Comparison of the traditional techniques employed for the analysis of Cannabis samples versus NIRS. Advantages are represented in greenish colour while disadvantages are in orangish colour, grey circles express unfilled. A scale from 1 to 5 has been employed to evaluate each parameter in terms of the degree of benefit/drawback. For instance, when comparing the 'Instrument/Maintenance expenses', a 5/5 disadvantage score is chosen for traditional techniques due to their characteristic high cost, while a 3/5 advantage score is selected for NIRS since they tend to have an intermediate cost.

Additional techniques have been utilized to determine further components of the plant. Malik et al. [31] describe the use of Kjeldahl method for the analysis of the nitrogen content in Cannabis as part of a comparative study of nutrition. Some other macro and microelements as well as trace elements were determined via flame atomic absorption spectroscopy (FAAS) and inductively coupled plasma optical emission spectroscopy (ICP-OES) [32,33]. In all the cases, sample treatment procedures require the use of hazardous chemicals, like nitric or sulfuric acid, at very high temperatures, which need to be handled with caution by trained personnel. It involves multiple steps, like sample digestion, and distillation and titration in the case of Kjeldahl, taking several hours to complete a single analysis (Figure 1). Moreover, it is imperative to follow the appropriate procedures and comply with environmental regulations when disposing of the chemical waste produced during the analysis. Despite all these disadvantages, the Kjeldahl approach is a well-established chemical method, used for many years to determine nitrogen in multiple samples of different fields, such as the food and beverage industry, agriculture and soil science, environmental analysis, and pharmaceutical industry [34]. On the other hand, measuring moisture content is particularly useful as it allows for results normalization, accounting for variations in sample humidity that may occur during storage, and can also provide information regarding the progress of cultivation [8,35]. This parameter could be determined employing several analytical techniques, such as a nuclear magnetic resonance (NMR) instrument [36], which has been used to measure the reference moisture content of ground and whole hemp materials. Additionally, the determination of sugars in Cannabis is a crucial aspect of plant analysis since it can provide valuable information regarding physiology or quality. Furthermore, sugars play a pivotal role in the biosynthesis of terpenes

and cannabinoids, and thus, understanding the sugar content can offer valuable insights into the capacity of the plant for producing these compounds, which are responsible for the plant's aroma and psychoactive effects. There are various analytical techniques that can be employed for determining sugar content, with HPLC being one of the most widely used methods [37,38].

As previously mentioned, there are adverse compounds of the Cannabis plant that also need to be determined in order to assess the quality control of Cannabis, i.e., pesticides, heavy metals or mycotoxins [11–13]. These compounds can be introduced in the Cannabis plant via various sources, e.g., soil, water, or air [39,40] and can be harmful to human health if consumed, leading to potential neurological problems, organ damage, or cancer, among others. For this reason, many countries have established regulations and limits, and thus, quality assessment by a specialized laboratory is mandatory for licensed Cannabis producers [40]. Numerous analytical techniques are typically employed for the determination of heavy metals, i.e., ICP-MS, AAS, or X-ray fluorescence (XRF) [41]; pesticides, i.e., GC-MS or LC-MS; and mycotoxins, i.e., LC-MS [11].

One of the main common disadvantages of all the aforementioned techniques, is the employment of expensive equipment that require expertise personnel as well as reactants and solvents. In this context, near-infrared spectroscopy (NIRS) has been gaining much attention in the last decade due to its multiple advantages (Figure 1). On the one hand, its broad spectral range facilitates the detection and quantification of numerous compounds or properties in complex mixtures without the need for extensive separation methods. Additionally, it is a low-cost and rapid technique that can provide results within a few minutes, which is especially useful for high-throughput analysis or applications where quick results are required [42]. Since it is a non-destructive technique, samples can be used after NIRS measurements for other applications [36,43], which is particularly advantageous when working with valuable or limited sample quantities [44]. On the other hand, its versatility allows broad application across various sample types, including solids, liquids, and gases, being particularly adept at simultaneous analysis of multiple components within a sample [45]. Nowadays, with the development of portable NIRS instruments, measurements can be performed in situ, and consequently, sample treatment is either entirely or almost entirely avoided [42]. Therefore, NIRS is truly in line with the green analytical chemistry principles, i.e., diminishing the utilization of hazardous chemicals and reagents, employing equipment that is energy-efficient, and producing minimal waste [46]. However, the main drawback of NIRS technology is the interpretation of the spectra, which is difficult, and the use of mathematical and statistical methods, i.e., chemometrics, is mandatory to extract the important information related to chemical mechanisms [47,48].

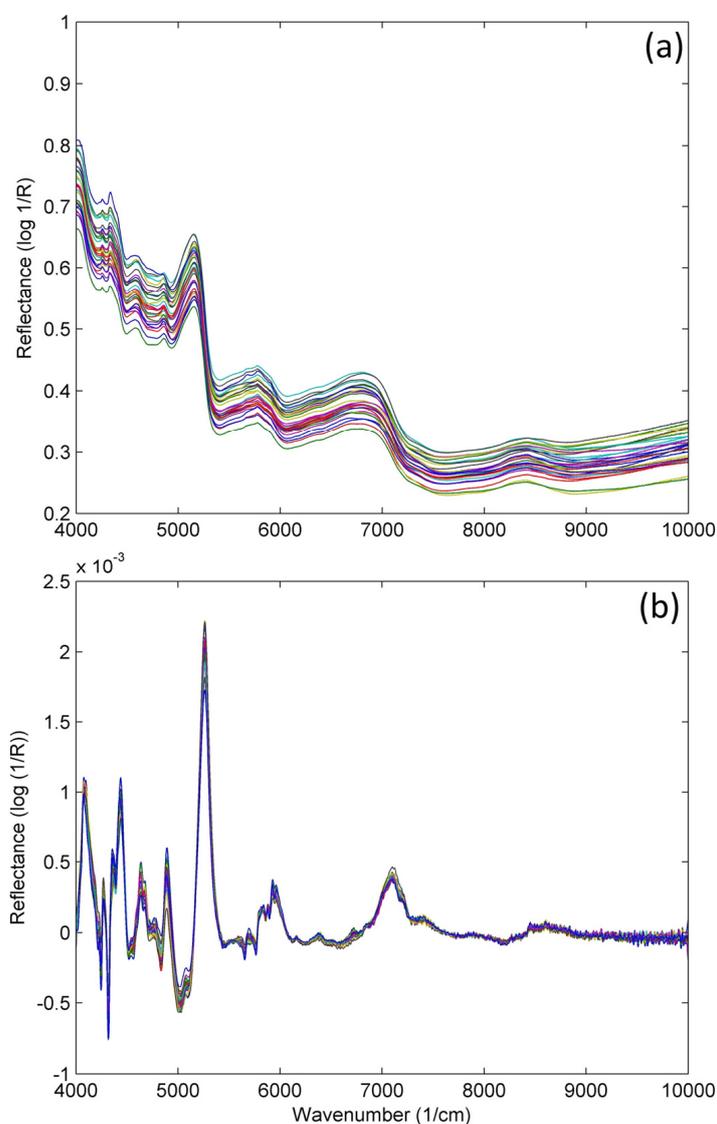
Once the potential of NIRS against traditional techniques has been settled, in this article, the different chemometric techniques commonly employed in NIRS method development will be described, along with their application for the analysis of *Cannabis sativa* L. samples. For this purpose, the article is divided in Sections 2 and 3, i.e., qualitative and quantitative methods, in which a thorough explanation of the different multivariate statistical techniques will be provided. Although there are several articles that review the different analytical techniques employed for the determination of cannabinoids [49], to the best of our knowledge, this is the first time that all this information along with the theoretical basis of NIRS method development is gathered in a revision article.

## 2. Qualitative Methods

The term “qualitative methods” refers to the identification and classification of samples based on their NIR spectra by means of spectral patterns. These methods are particularly useful when aiming for a differentiation between types or classes of samples without quantifying specific compounds. In this sense, *Cannabis sativa* L. is a type of cultivar with multiple characteristics than can be employed for its classification, e.g., chemotype, genotype, and growth stage, among others [50–52]. Additionally, NIRS methodology has been used for the classification of samples regarding the crop year or location [53].

Furthermore, to develop a robust NIR methodology, a large number of samples is required in order to cover a wide range of variability among the samples. As a consequence, a big matrix of data and variables is usually obtained, thus hindering the ability to easily extract the important information within the data [54]. There are different chemometric techniques that can facilitate the process of data management (Table 1), which will be described in this section, along with their application to the analysis of *Cannabis sativa* L. samples.

Prior method development, preprocessing of the spectra is a very important step in order to avoid interferences related to factors that may affect the information of interest, e.g., sample particle size variations, outliers, scatter effects or baseline shifts, among others [52,55]. For this purpose, different mathematical treatments can be used, e.g., standard normal variate (SNV), detrending, Savitzky–Golay polynomial derivative filters, multiplicative scatter correction (MSC), etc. [56]. These techniques are useful to prevent the effects previously mentioned while preserving the shape and integrity of the original spectral features, which is particularly interesting in the case of complex spectra [57]. As can be seen in Figure 2, the quality of the spectra is considerably improved after preprocessing, with sharper and well-differentiated bands.



**Figure 2.** NIR spectra of Cannabis samples (a) raw spectra and (b) spectra after preprocessing (smoothing, MSC, first derivative, and mean-centered). Reproduced from reference [51] with permission from Elsevier, copyright 2016.

Once the data are preprocessed, the classification can be performed following two approaches, i.e., supervised and unsupervised. Regarding the first approach, the supervised methods, there is previous knowledge about some characteristics of the samples, such as the categories in which the samples are classified. However, when developing unsupervised methods, no previous information about the samples is provided and, therefore, self-interpretation of the differences encountered among the spectra is needed [58].

## 2.1. Unsupervised Methods

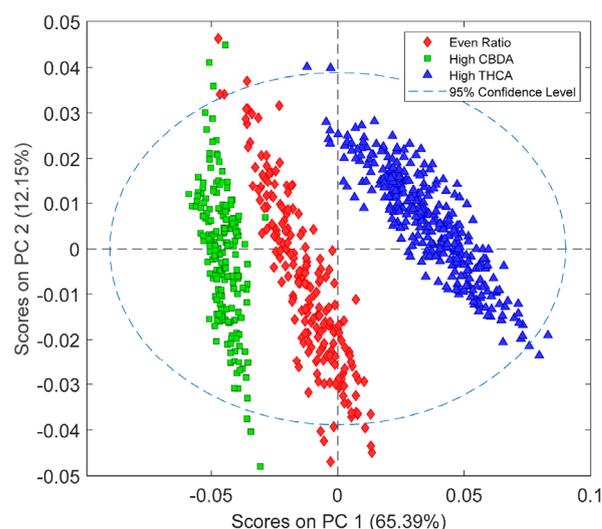
### 2.1.1. Principal Component Analysis

The aim of principal component analysis (PCA) is to reduce the dimensionality of the spectral data and identify the most significant variations, thus facilitating the visualization and interpretation of sample groupings [58]. In order to achieve this goal, PCA finds a new coordinate system that represents the data with a reduced number of uncorrelated variables called principal components (PCs) [59]. In this sense, PCA is used to obtain a graphical representation with no previous knowledge about the samples.

Considering that PCA diminishes the complexity of the spectral data, other advantages arise, namely data processing, since redundant or highly correlated variables are eliminated. Additionally, PCA facilitates the detection of outliers, i.e., samples that deviate significantly from the majority of data points, by means of their abnormal coordinates in the transformed space. By combining all these features, PCA effectively identifies important characteristics within a dataset, thus being a highly valuable technique [60]. In the field of Cannabis, PCA has been particularly interesting, since it allows researchers to differentiate not only between chemotypes, but also to distinguish among illegal or legal Cannabis plantations depending on the country regulations, thus becoming a great help for the authorities considering that further analysis and sample treatment can be avoided [51,52,61]. To give an example, Duchateau et al. [52] employed different chemometric techniques to compare the performance of two NIR devices, i.e., a benchtop and a handheld device, for the classification of 189 Cannabis samples as legal or illegal according to the European and Swiss legislation. In this article, PCA played a crucial role in visualizing the spectra in two- or three-dimensional plots. Its purpose was to perform an exploratory analysis, ensuring that the data could be effectively modelled afterwards using supervised techniques, and that good predictive results are not based on coincidence or due to modelling of the noise.

Borille et al. [51] also employed the NIRS technology combined with chemometric methods for the growth stage classification of 29 sample triplicates obtained from Cannabis cultivated in a greenhouse from seized seeds. In this article, PCA by intervals (iPCA) was employed in the raw spectra as a preprocess step in order to eliminate irrelevant information and select the best spectral range for the classification. This technique is an extension of the traditional PCA that evaluates intervals rather than point estimates, thus providing a more comprehensive and robust analysis of the data. Subsequent PCA was also used for the purpose of visualizing the clusters of samples according to similarity.

Relatedly, Tran et al. [42] developed prediction models using two different NIRS instruments, i.e., benchtop and handheld, for the quantification of cannabinoids in Cannabis samples. Before quantitative method development, they applied different preprocessing techniques and PCA to take an overview of the classification of the data. As can be seen in Figure 3, PCA revealed three well-defined classes depending on the cannabinoid content, i.e., high CBDA, high THCA, and even ratio, for the benchtop instrument, thus showing the potential of this technique to preliminary classification of samples.



**Figure 3.** Principal components analysis of the Cannabis samples using the benchtop NIR instrument. Adapted from reference [42].

### 2.1.2. Hierarchical Clustering Analysis

Hierarchical clustering analysis (HCA) is an unsupervised multivariate analysis technique commonly used in NIRS technology. HCA consists of clustering the samples into groups by means of the similarities between the spectra, so that the within-group similarities are larger compared to the between-group similarities. Thus, by performing successive divisions, a tree-like structure, i.e., dendrogramme, is obtained [62]. In order to obtain this dendrogramme, two main approaches are employed: agglomerative and divisive. The agglomerative approach is a bottom-up strategy in which the samples are considered as individual clusters that are progressively grouped according to their similarity until a single cluster containing all the samples is formed. On the other hand, the divisive approach is a top-down method that separates a single cluster containing all the samples into smaller subclusters based on the differences between samples or groups.

In the article previously mentioned by Duchateau et al. [52], HCA was employed to do a preliminary evaluation of the different clusters of samples. For the benchtop NIR, three major clusters were obtained, separating the samples according to the content of  $\Delta^9$ -THC in higher than 1% (*w/w*), between 0.2 and 1% (*w/w*) and less than 0.2% (*w/w*). As for the handheld device, HCA showed to major clusters: one with samples with a  $\Delta^9$ -THC content higher than 1% (*w/w*) and a second cluster with samples containing less than 1% (*w/w*). Thus, by employing HCA, a global vision of the similarities between samples is possible, and therefore, the suitability of a subsequent supervised method is ensured.

Furthermore, in the article mentioned in the previous section by Borille et al. [51], HCA revealed three main groups according to the three growth stages employed. However, the dendrogram did not entirely reveal the specific clustering of the growth stage, which confirms that these chemometric techniques are used only for a preliminary exploratory analysis and not for a precise classification, for which more advanced chemometric tools are required.

### 2.1.3. Non-Hierarchical Clustering Analysis

The objective of non-hierarchical clustering methods is to obtain one final partition of the data. In this case, the number of clusters is fixed by the user, which is an important drawback to these techniques since there is no previous knowledge about the samples. Some examples of non-hierarchical clustering methods are the k-means, which assess the similarities between samples by means of distance measurements; the Density Based Spatial Clustering of Applications with Noise (DBSCAN), based on sample density; or the Self-Organising Map (SOM), which studies the relationship between variables and

samples, among others [63]. However, no applications of these methods to Cannabis samples combined with NIRS spectroscopy have been found.

## 2.2. Supervised Methods

### 2.2.1. Soft Independent Modelling of Class Analogy

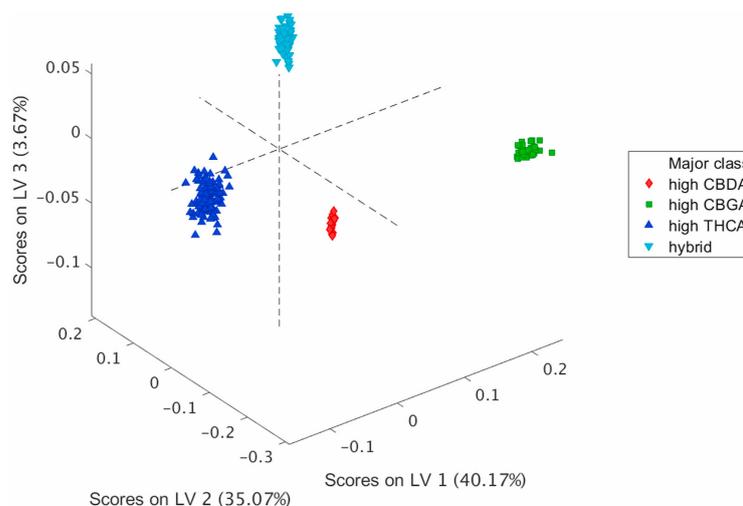
Soft Independent Modelling of Class Analogy (SIMCA) is a class modelling technique commonly used in chemometrics for pattern recognition and classification, thus being a supervised method [64,65]. Samples are placed in a PCs space that describes a certain class created by SIMCA after applying PCA, in order to evaluate whether they belong to it or not. The number of PCs determined by SIMCA is key, since important features from the model may be excluded when lacking PCs, thus hindering the selectivity. On the other hand, employing an excessive number of PCs can lead to increase in the noise and overfitting of the model [66]. As implied on its name, SIMCA is focused on evaluating the similarities among the samples within a class rather than on discrimination between classes [67].

Pereira et al. [50] developed a method using NIRS hyperspectral imaging and machine learning methods for the detection and identification of illegal plantations of *Cannabis sativa* L. PCA was firstly employed to explore the characteristics of the images and identify the potential clusters related to *Cannabis sativa* L. Different areas of the images were selected in order to collect the pixels from important parts of the plant, e.g., margin, veins, etc. Prior to SIMCA method development, sparse PCA was performed to eliminate the variables that were not informative. Then, the training set of 401 pixels/spectrum selected via sparse PCA were used to build the SIMCA model. By selecting another region of interest from the same image, a validation set was obtained and the developed SIMCA model was applied, obtaining good results in terms of sensitivity and specificity, thus ensuring the suitability of the model for the identification of *Cannabis sativa* L. with a low false-positive rate.

### 2.2.2. Partial Least Squares Discriminant Analysis

Another supervised method is the so-called partial least squares discriminant analysis (PLS-DA), which combines dimensionality reduction and prediction model construction [68]. Although this technique has been gaining attention in recent decades, it is prone to overfitting in some cases in which the number of variables significantly exceeds the number of samples. Therefore, cross-validation becomes an essential step, not only for characteristic selection and classification, but also for mere data visualization. On the other hand, it is a multi-step procedure that involves different mathematical operations and parameters [68]. However, if all the potential obstacles are carefully taken into account and a thorough evaluation of every step of the procedure is carried out, PLS-DA can be a powerful technique for managing highly multidimensional data.

There are several articles in which this technique is employed so as to take an overlook of the data that are going to be subsequently fitted into a quantitative method or to develop very accurate qualitative methods for the classification of Cannabis samples [20,42,69,70]. For instance, Birenboim et al. [20] developed a method for the classification of Cannabis cultivars and the quantification of major cannabinoids and terpenes via Fourier transform near-infrared spectroscopy (FT-NIR) combined with chemometrics. Prior development of the quantitative method, multivariate classification and regression models were used, e.g., PLS-DA. This technique was particularly useful for major class separation of the samples in four categories, namely high-THC, high-CBD, hybrid, and high-CBG, for which the FT-NIR spectra of the ground inflorescence samples were employed, as observed in Figure 4. The results showed no misclassified samples, with sensitivity and specificity values of 1, which ensures that the classification model is highly accurate. Furthermore, the root mean standard error of cross validation (RMSECV) and calibration (RMSEC) ratio, along with the root mean standard error of prediction (RMSEP) and the RMSECV ratio, were both below 1.5, thus indicating a low probability of model overfitting to the data.



**Figure 4.** Partial least squares discriminant analysis classification of the FT-NIR spectra of Cannabis inflorescence samples. Reproduced from reference [20] with permission from Elsevier, copyright 2022.

**Table 1.** Highlighted applications of chemometric methods in the analysis of Cannabis samples.

Instrument	Spectral Range (nm)	Samples	Parameter	Spectra Pretreatment	Chemometric Method	Ref.
FOSS NIRSystem 6500 (benchtop)	1100–2500	Inflorescences and leaves	$\Delta^9$ -THC	-	PCA	[61]
PerkinElmer™ Frontier MIR/NIR (benchtop) Thermo Scientific MicroPHAZIR RX (handheld)	1000–2500 1600–2500	Inflorescences	$\Delta^9$ -THC	SNV and Savitzky–Golay filters First derivative and SNV	PCA, HCA, SIMCA, k-NN, and PLS-DA	[52]
PerkinElmer 400 IR	1000–2500	Aerial parts	Growth stage	Savitzky–Golay filters, MSC, and mean centering	HCA, PCA, PLS-DA, and SVM-DA	[51]
Bruker MPA II FT-NIR (benchtop) Viavi MicroNIR Onsite-W (handheld)	870–2500 950–1650	Dried inflorescences	14 cannabinoids	Detrend, SNV, and normalization	PCA and PLS-DA	[42]
Specim, SisuChema (handheld)	-	Leaves	Detection and classification of Cannabis	SNV, MSC, Savitzky–Golay filters	PCA and SIMCA	[50]
ThermoFisher, Antaris II FT-NIR (benchtop)	1000–2500	Inflorescences	CBDA, CBGA and THCA	Savitzky–Golay filters, SNV, MSC, and mean centering	PLS-DA	[20]
VIAMI, microNIR (portable)	900–1700	Hemp flours	CBD, $\Delta^9$ -THC and CBG	SNV, MSC, mean centering, normalization, Savitzky–Golay filters	PCA	[69]
VIAMI, microNIR (portable)	900–1700	Oral fluids	$\Delta^9$ -THC	SNV, MSC, normalization, and Savitzky–Golay filters	PLS-DA	[70]
Perten DA7250	950–1650	Ground and whole hemp	$\Delta^9$ -THC	Mean centering	LDA	[36]

$\Delta^9$ -THC: tetrahydrocannabinol; PCA: principal component analysis; SNV: standard normal variation; HCA: hierarchical clustering analysis; SIMCA: soft independent modelling of class analogy; k-NN: k-nearest neighbour; MSC: multiplicative scatter correction; PLS-DA: partial least squares discriminant analysis; SVM-DA: support vector machine discriminant analysis; CBDA: cannabidiolic acid; CBGA: cannabigerolic acid; THCA: tetrahydrocannabinolic acid; CBD: cannabidiol; CBG: cannabigerol; LDA: linear discriminant analysis.

Similarly, Tran et al. [42] developed different prediction models using two NIRS instruments, i.e., a benchtop FT-NIR and a handheld microNIR, for the classification and quantification of cannabinoids in *Cannabis sativa* L. samples. After performing PCA to check data quality and have an overview of the trends within the population, which showed three different clusters based on the chemotype, PLS-DA models were carried out to ensure how accurately the chemotypes could be predicted. The results showed three well-defined categories, namely high-CBDA, high-THCA, and even-ratio for the benchtop device, with a classification accuracy of 100% for both high-CBDA and high-THCA models and 99.4% for even-ratio model. The models obtained by using the handheld device also showed the same three categories as the benchtop instrument; however, only the high-THCA model has a classification accuracy of 100%, and the other models showed lower classification accuracy values compared to the benchtop instrument.

San Nicolas et al. [71] developed a non-invasive method for the classification of Cannabis cultivars based on hyperspectral imaging along with PCA and PLS-DA. In this case, two different approaches employing PLS-DA were proposed, i.e., direct calibration with the complete dataset, which comprise 502 flower spectra; and a two-layer hierarchical model, which classifies the stem separated from the rest of the plant. Both classification methods showed promising results, with an 89.47% correct classification of the samples for the direct calibration, and a 91.23% for the hierarchical approach.

### 2.2.3. Parametric and Non-Parametric Methods

Artificial neural networks (ANN) are non-linear, non-parametric supervised methods composed of several layers of neurons, namely inputs output and hidden [72]. A neuron is a processing unit that transforms input data into output by means of back-propagation of the hidden layers, which represent the modelling process. Several parameters have to be optimized when developing ANN methods, e.g., the number of hidden layers, which can be an intricate and time-consuming process if the proper training algorithm is not selected [72,73]. Despite the fact that it is challenging to ascertain the structure of an ANN, these methods are suitable for classification purposes due to their ability to calculate and approximate functions of any form [74,75].

There are several articles in which ANN are applied to create a model for the determination of different parameters from samples of hemp extracts [76] or for optimizing in vitro germination and growth of hemp seeds [77], among others. However, no specific applications of this technique to Cannabis samples in combination with NIRS technology have been found.

On the other hand, k-nearest neighbours (KNN) is a non-parametric supervised technique used for classification via evaluation of distances between samples [58]. Typically, the Euclidean distance between the samples from the validation set and the samples from the training set is calculated and the different classes in which samples can be divided are established [78]. Then, the unknown samples are placed in the class to which the majority of its KNN from the training set belong.

Linear discriminant analysis (LDA) is a linear, parametric, and, as its name indicates, discriminant supervised technique originally described by R. Fisher [79]. This method aims to reduce the dimensionality of the data set, similarly to what PCA provides; however, LDA finds the hyperplane that maximizes the distance between the different classes [80]. Furthermore, when employing LDA, the location of the data set is not changed, as opposed to PCA. Su et al. [36] developed a NIRS method for rapid measurement of moisture and cannabinoid contents in samples of *Cannabis sativa* L., along with a qualitative method for the classification of the samples in legal or illegal via LDA. The application of this technique yielded up to 94% correct classifications, thus showing the potential of LDA for discriminant analysis. A more flexible technique is Quadratic Discriminant Analysis (QDA), since it can learn quadratic boundaries, as opposed to LDA, which is based on linear boundaries [81]. Borregaard et al. [82] carried out a study about the discrimination of crop and weed based on high-dimensional spectral data. They stated that QDA and LDA

classification present similarities, with a lower performance than PCA/SIMCA and PLS methods in illegal or legal Cannabis plantations.

Table 1 summarizes the applications mentioned throughout the present article, which comprise different chemometric techniques employed for the classification of cannabinoids and other parameters in Cannabis samples, alongside the mathematical techniques employed for spectra pretreatment.

### 3. Quantitative Methods

Quantitative methods in near-infrared spectroscopy (NIRS) refer to analytical approaches used to determine the concentration or amount of specific compounds or properties in a sample. They are valuable for quality control, potency assessment, and determining the chemical composition of Cannabis samples in research, forensic, and regulatory settings [22,43]. These methods involve establishing a relationship between the spectral data obtained from NIRS measurements and the concentration of the target analyte or property of interest [83]. Despite all the advantages of NIRS, which were previously described in Figure 1, it is important to note that calibration models for Cannabis analysis should be developed and validated using appropriate reference values and considering the specific Cannabis matrices and regulatory requirements of the intended application [36]. Therefore, high-quality reference data are essential to ensure the accuracy, precision, and reliability of the calibration model and subsequent predictions. In order to achieve this goal, a set of representative samples with known reference values is initially introduced in the NIRS equipment to obtain their corresponding spectra. These samples should cover the range of concentrations expected in the samples to be analyzed [20]. The reference values are typically obtained using a reference method, such as traditional wet chemistry techniques. Although this step may appear inconsequential, it truly holds significant importance within the process as the accurate prediction of the model relies heavily on these data.

Once all the data are obtained, as was the case with qualitative methods, the NIRS spectra collected from the calibration sample set may be preprocessed to avoid interferences related to factors that may affect the information of interest, e.g., sample particle size variations, outliers, scatter effects, or baseline shifts, among others. This may involve steps like baseline correction, smoothing, normalization, or noise reduction techniques [56]. The preprocessed spectral data are then correlated with the corresponding reference values to establish a calibration model. Various multivariate statistical techniques, such as partial least squares (PLS) regression or principal component regression (PCR), are commonly used. These methods aim to find the best mathematical relationship between the spectral data and the reference values. Finally, the developed calibration model needs to be validated using an independent set of samples not used in the model development. These validation samples are analyzed using NIRS, and the predicted values are compared to their reference values to assess the accuracy and reliability of the model. Statistical parameters like root mean square error of prediction (RMSEP) or correlation coefficients are typically used to evaluate the performance of the model. According to the ICH Guideline Q2(R2) on the validation of analytical procedures, if RMSEP is comparable to root mean-squared error of calibration (RMSEC), then the accuracy of the method is confirmed [84]. Another parameter that is usually employed is standard error of prediction (SEP), which is similar to RMSEP but independent of bias, unlike RMSEP [85]. Calibration in NIRS is an ongoing process that should be continuously improved over time by introducing new samples that can enrich the performance of the predictive equation and increase its robustness.

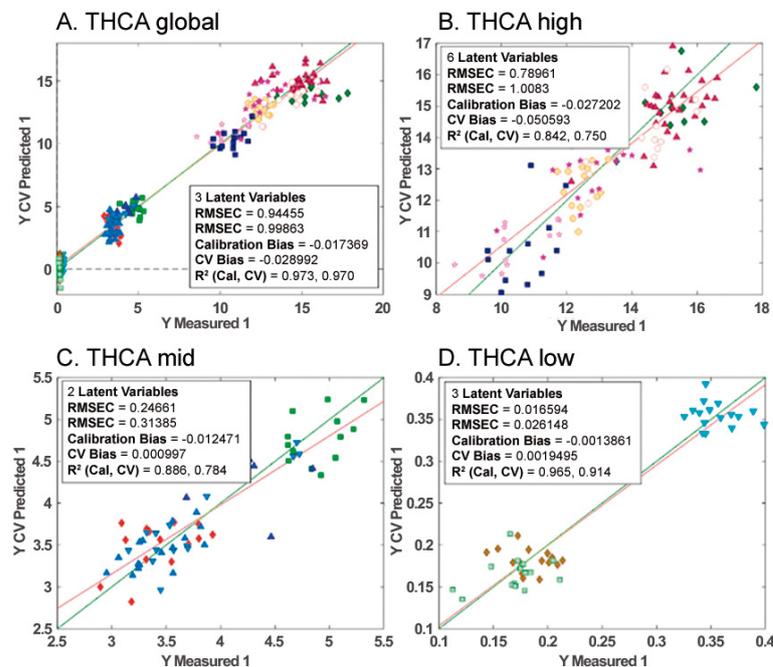
Quantitative NIRS methods find applications in various fields such as pharmaceutical analysis, food quality control, agricultural monitoring, and environmental analysis. In the field of Cannabis, this technique has garnered increasing attention in recent years, gradually establishing itself as an integral part of the quality control process for this medicinal product [20,71]. For the calibration model development, as aforementioned, there are multiple techniques commonly employed for the obtention of predictive equations, which will be described in the next section.

### 3.1. Linear Regression Multivariate Statistical Techniques

#### 3.1.1. Partial Least Squares Regression

Partial Least Squares Regression (PLS-R) is a predictive technique that originated from Herman Wold's concept of PLS and has gained considerable importance in various fields over time. This method is particularly useful when dealing with datasets exhibiting a high degree of multicollinearity [86]. In order to solve the problem mentioned, PLS-R performs iterative least squares fitting on latent variables, which are found as linear combinations of the initial variables. By doing so, PLS-R sets the spectral data variables that best describe the reference values [58]. In contrast to other multivariate statistical techniques, such as PCR, PLS-R provides more accurate predictions by placing stronger emphasis on the relationship between the independent and dependent variables during dimensionality reduction.

Owing to PLS-R being one of the most employed multivariate techniques, numerous articles describe the use of this approach for developing calibration models for various matrices in NIRS. In the particular case of Cannabis, PLS-R is usually employed for the construction of regression models for cannabinoids [22,36,42–44,55,69,70,87] and/or terpenes [20], among other parameters [36,88]. Birenboim et al. [20] describe the use of Fourier transform near-infrared spectroscopy (FT-NIR) to determine cannabinoid and terpene content in Cannabis inflorescence samples. Spectral regions of 1450–1880 and 2130–2350 nm were fundamental for predicting all cannabinoids and terpenes by means of PLS-R. As can be observed in Figure 5A, when representing the complete dataset, three well-defined clusters are formed, i.e., THCA high (Figure 5B), THCA mid (Figure 5C), and THCA low (Figure 5D). This is also observed for the rest of cannabinoids. The authors indicate the high predictive capabilities of the developed models; however, according to Williams [89], RPD values are low for that classification (Table 2), and they should be used instead for screening purposes.



**Figure 5.** Correlation between the predicted values of THCA (*y*-axis) via PLS-R and the references values (*x*-axis) provided through HPLC-DAD at global (A), high (B), mid (C) and low (D) concentration. Colours and shapes correspond to the different strains employed for calibration. Adapted from reference [20] with permission from Elsevier, copyright 2022.

Sánchez-Carnerero et al. [48] made a comparative study using two equipment, a dispersive NIR and a FT-NIR spectrometers. In this case, they employed PLS-R for the determination of various cannabinoids, affording RPD values better than 3 for  $\Delta^9$ -THC,

CBC and CBD, being close to 6 for the latter. They also build prediction equations for CBDV,  $\Delta^9$ -THCV,  $\Delta^8$ -THC, CBG, and CBN, obtaining RPD values close to 2 in these cases (Table 2), and therefore being useful for screening purposes. The large number of samples for calibration, as well as the dispersion of data (SD) in  $\Delta^9$ -THC, CBC, and CBD, play an important role in obtaining good predictive results. Both instruments provide similar results for all cannabinoids, although the chemometric treatment of data was varied.

### 3.1.2. Principal Component Regression

Principal component regression (PCR) is a multivariate analysis technique based on the combination of PCA and least-squares regression. PCR mainly focuses on dimensionality reduction, unlike PLS-R, in which the prediction of the dependent variables is emphasized. This is one of the reasons to explain that PCR components are always orthogonal, i.e., they are not affected by the dependent variable, as opposed to PLS-R, in which components are not necessarily orthogonal and are chosen to retrieve as much predictive information as possible [90]. However, although PCR interpretability can be challenging owing to the components being linear combinations of the original predictors, it is a very useful technique in those cases in which dimensionality reduction is the priority. PCR models have been extensively employed in different matrices; nevertheless, only one NIRS application in Cannabis has been found. Townsend et al. [91] developed an application note explaining the determination of  $\Delta^9$ -THC and CBD content in Cannabis flowers via FT-NIR. For this purpose, PCR was employed as chemometric model, affording values of standard error of prediction (SEP) of 0.73 and 0.92% (Table 2), for total CBD and total THC, respectively, which demonstrates the potential of this PCR-based FT-NIR model.

Further lineal regression multivariate statistical techniques, namely multiple linear regression (MLR), are also helpful in those cases in which the number of independent variables is small, and the main objective is the interpretation of the results and not to reduce dimensionality or to predict [92]. However, although MLR has been widely used in various matrices, no NIRS approaches employing MLR in Cannabis have been found.

## 3.2. Non-Linear Regression Multivariate Statistical Techniques

### 3.2.1. Artificial Neural Networks

Although all the aforementioned techniques have been widely used throughout the years, there are some situations in which these methods are not suitable for data treatment, e.g., when the mathematical model describing the data set is unknown. In this context, artificial neural networks (ANN) emerges as a powerful modelling technique since it can retrieve useful information when going through all the data and is capable of modelling complex non-linear relationships [93]. For this reason, it has great potential when dealing with multicollinearity and a large number of independent variables. However, due to the complicated relationships provided by ANN, the interpretation of the model can be a hurdle. Furthermore, this technique usually requires a higher amount of data and computational resources.

ANN models have been used in a wide variety of fields and matrices, from biological to commercial or food and drink approaches. As for Cannabis applications, there are few in which ANN models are applied. For example, Gloerfelt-Tarp et al. [47] developed a NIR-based chemometric application for the quantification of 12 cannabinoids in plant material, with emphasis on the discrimination between neutral and carboxylic forms of each cannabinoid. For this purpose, different machine learning algorithms were employed, including deep neural network and random forest, affording values of root mean standard error of validation ( $RMSE_v$ ) in the range of 0.001 and 0.560 (%) (Table 2). Valinger et al. [76] also employed ANN modelling to predict different physical and chemical properties in hemp extracts, namely total dissolved solids, extraction yield, total polyphenolic content, and antioxidant activity. As observed in Table 2, the values of  $RMSE_v$  were in the range of 0.0140–305.5601% for solid–liquid extraction (SLE) and 0.0320–21.8810% for microwave-assisted extraction (MAE).

**Table 2.** Highlighted applications of multivariate statistical techniques in the analysis of Cannabis samples.

Instrument	Spectral Range (nm)		Samples	Parameter	Regression Model	n		$r_v^2$		RMSE <sub>v</sub> (%)		RMSEP (%)	SEP (%)	RPD	Ref.			
FT-NIR (handheld)	1350–2560		Dry hemp	CBD, total CBD, $\Delta^9$ -THC, and total THC	PLS-R	67–72		0.9100–0.9500		-	-	0.02–0.61	-	-	[22]			
VIAMI, microNIR (portable)	900–1700		Hemp flours	CBD, $\Delta^9$ -THC, and CBG	PLS-R	10		0.9741–0.9980		0.005	0.005–0.007		-	-	[69]			
Acousto-Optic Tunable Filter NIR	1200–2200		Dry hemp	$\Delta^9$ -THC and CBD	PLS-R	91–103		0.77		0.0140–0.4310		-	-	2.04–2.07	[55]			
Bruker MPA II FT-NIR (benchtop)	Viavi MicroNIR Onsite-W (handheld)	870–2500	950–1650	Dried inflorescences	14 cannabinoids	PLS-R	734	730	0.2500–0.9800	0.2100–0.9800	0.0800–7.000	0.0800–6.530	0.0600–5.5100	0.0800–6.2300	-	-	[42]	
VIAMI, microNIR (portable)	900–1700		Oral fluids	$\Delta^9$ -THC	PLS-R	50		0.989		1.1	1.3		-	-	[70]			
Resonon Pika XC2 hyperspectral camera	400–1000		Fresh flowers	Fresh leaves	CBD, $\Delta^9$ -THC, CBG, CBDA, THCA, and CBGA	PLS-R	100		0.5100–0.8500	0.4200–0.71	0.9000–20.6700	0.1600–3.7600	-	-	1.43–2.62	1.32–1.88	[44]	
Pertea DA7250	950–1650		Ground hemp	MC	5 cannabinoids	PLS-R	115		0.91	0.0300–0.8500	-	-	1.28	0.02–0.92	-	[36]		
			Whole hemp	MC	5 cannabinoids		194		0.94	0.03–0.89	-	-	1.24	0.01–0.60				
TellSpec NIR-S-G1 (handheld)	900–1700		Resins	$\Delta^9$ -THC	PLS-R	-		0.02		5.19	3.87		-	-	1.51	[43]		
Viavi Solutions MicroNIR (handheld)	950–1650					0.67		2.5	1.46		2.26							
ThermoFisher Antaris II FT-NIR	1000–2500		Dried inflorescences	10 cannabinoids	9 terpenes	PLS-R	47–237	84–218	0.6250–0.9900	0.7000–0.8870	0.0100–1.0080	0.0032–0.0400	0.0110–1.2750	0.0037–0.0416	-	1.87–10.87	1.78–3.00	[20]
FOSS NIR Systems 6500 (benchtop)	Bruker FT-NIR (portable)	400–2498	800–2500	Dried leaves and inflorescences	8 cannabinoids	PLS-R	189		0.5400–0.9800	0.7800–0.9900	-	-	0.03–1.72	0.04–1.79	1.25–6.03	1.52–6.00	[48]	
PerkinElmer Spectrum Two FT-NIR	1000–2500		Dried flowers	CBD and $\Delta^9$ -THC	PCR	302		0.9700–0.9800		-	-	0.73–0.92	-	-	[91]			
Hone HL-EVT9-Neo NIR (portable)	1250–2500		Dried flowers	12 cannabinoids	ANN	249		0.0300–1.0000		0.0010–0.5600		-	-	-	[47]			
Control development NIR spectrophotometer	904–1699		Hemp extracts	TDS, EY, TPC, and AC	ANN (SLE)	ANN (MAE)	-		0.5925–0.9547	0.6459–0.9434	0.0140–305.5601	0.0320–21.8810	-	-	-	[76]		
Bruker Matrix-F FT-NIR	800–2500		Hemp oil	CBD and total CBD	SOSVEN	sPLS-R	20		0.9828–0.9864	0.9810–0.9844	6.4000–6.6000	6.8700–7.0000	-	-	-	[57]		

CBD: cannabidiol;  $\Delta^9$ -THC: tetrahydrocannabinol; PLS-R: partial least squares regression; CBG: cannabigerol; CBDA: cannabidiolic acid; THCA: tetrahydrocannabinolic acid; CBGA: cannabigerolic acid; MC: moisture content; PCR: principal component regression; ANN: artificial neural network; SLE: solid–liquid extraction; MAE: microwave-assisted extraction; SOSVEN: self-optimizing support vector elastic net; sPLS-R: super partial least squares regression.

### 3.2.2. Support Vector Machine

Support vector machine (SVM) is a multivariate statistical technique originally proposed to address classification issues, but it can be applied to various situations and fields, e.g., bioinformatics or handwriting recognition, among others [94]. The main objective of SVM is to find a hyperplane that best separates the different classes of a data set based on particular patterns in those classes or observations. Unlike ANN, SVM models are transparent; i.e., it is easier to interpret which data points contribute to the model. However, SVM applications are mainly designed for binary classification, while ANN can handle multiclass classification without the need of combination with other techniques. Nonetheless, SVM models are preferred when overfitting is a concern, since they tend to be less prone to cause this problem [95]. Chen et al. [57] proposed a NIR methodology for the quantification of CBD in hemp oil via comparison of PLS-R and self-optimizing support vector elastic net (SOSVEN) models. This technique is an advanced variant of SVM that combines the typical elements of SVM with characteristics from elastic nets (EN), which are usually employed in linear regression. SOSVEN is particularly useful for feature and model selection, along with hyperparameter optimization. In this case, SOSVEN had lower validation errors when compared to PLS-R to predict the concentration of CBD and total CBD (Table 2), thus demonstrating the potential of this multivariate statistical technique.

### 3.3. Near-Infrared Hyperspectral Imaging

Although NIRS can predict physical and biochemical characteristics in diverse sample types with high efficiency, simplicity, and accuracy, its measurement scope is restricted to a relatively small section of the specimen for determining average composition values. This shortcoming is especially problematic when dealing with heterogeneous samples, such as Cannabis samples, as NIRS spectroscopy fails to provide important information on the spatial distribution of quality parameters [96]. However, the integration of NIRS-hyperspectral imaging systems facilitates the simultaneous acquisition of spatial and spectral information [73].

Near-infrared hyperspectral imaging (NIR-HSI) is a highly advanced methodology, which can capture up to several hundred images of different wavelength, offering a detailed spectral response of target features [97]. This technique is particularly adept at discerning even the most subtle variations in ground covers, as well as tracking changes over time. Previous research has also demonstrated that HSI surpasses multispectral images in terms of effectively monitoring vegetation properties, such as the leaf area index (LAI), differentiating between crop types, retrieving crop biomass, and assessing leaf nitrogen content [98]. HSI has emerged as a highly promising method for the non-invasive assessment of diverse constituents of the Cannabis plant. This technique involves the measurement of pixel reflectance and subsequent correlation with cannabinoid content, enabling effective and reliable evaluation of the properties of the plant. Holmes et al. [99] utilized NIRS-HSI to estimate the content of cannabidiolic acid (CBDA) in flowers and leaves of *Cannabis sativa* L. The Gaussian Process Regression (GPR) model was chosen as it possessed the capacity to effectively predict the outcomes obtained through LC-MS. The proposed model displayed significant potential as a screening technique for implementation in the cultivation of this crop. Additionally, Lu et al. [44] describe a HSI technology for non-destructive quantification of major cannabinoids, including CBD,  $\Delta^9$ -THC (tetrahydrocannabinol), CBG (cannabigerol) and their acid forms in fresh floral and leaf materials of industrial hemp on a dry weight basis. Parsimonious PLS models were utilized, obtaining the best RPD values of 2.6 for CBD and  $\Delta^9$ -THC in flowers. This value indicates that the prediction of these cannabinoids is fair, being appropriate for screening purposes [100]. The lack of accuracy in the prediction method could be related to the measurement wavelength range used in this article, which was in the short-wave near-infrared (SW-NIR) region, from 400 to 1000 nm. However, according to Sánchez-Carnerero et al. [48], cannabinoids mainly absorb in the 1064–2357 nm range. Therefore, extending the measurement region to a wider scale could be a solution, as additional information may have been lost in the

shorter region. On the other hand, Abeysekera et al. [101] developed a method based on HSI to determine THCA in Cannabis plant samples. In this case, the utilized camera worked from 645 nm to 2070 nm, obtaining increased spectral information to be related to cannabinoids concentration. Partial least squares feature selection (PLSFS) was selected as the regression model, as it provided the best accuracy in estimating of the THCA content. Despite the multiple benefits of NIR-HSI, the technique presents several disadvantages to take into consideration, such as limited accessibility to some researchers or industries with budget constraints, due to the high cost of acquisition and maintenance of HSI systems. Additionally, HSI generates large amounts of data due to the high number of spectral bands. Analyzing and processing such voluminous data can be computationally intensive and time-consuming [102]. On the other hand, the spectral interference and the environmental sensitivity makes it difficult to obtain good predictions to be used in quality control or process control evaluation [97].

Table 2 summarizes the applications mentioned throughout the present article, which comprise various quantitative multivariate statistical techniques for the quantification of cannabinoids, terpenes, and/or other parameters in Cannabis samples via NIRS.

#### 4. Conclusions

Traditionally, Cannabis has been analysed using different analytical techniques that, despite their inherent benefits, come with significant disadvantages, hindering their application for routine analysis. In the era of heightened environmental consciousness, the demand for more eco-friendly methodologies has intensified. Near-infrared spectroscopy (NIRS) satisfies this necessity as moderately-priced equipment that efficiently operates within seconds without the requirement of toxic reagents. However, its primary drawbacks include the necessity of analytical reference techniques for providing reference values and the expertise in chemometrics to correlate spectral and numerical data, necessary for developing predictive equations. This article offers a comprehensive review of the most employed chemometric techniques, classified into qualitative and quantitative methods, and their application to obtain NIRS equations for predicting cannabinoids, terpenes, and other parameters in Cannabis samples.

Future applications of NIRS in Cannabis analysis should be headed towards developing new predictive equations, expanding beyond cannabinoids, terpenes, and moisture content. Further interesting compounds of Cannabis, such as flavonoids, phenols, or alkaloids, shows considerable potential in influencing the organoleptic properties of the future medicine. Furthermore, compounds like heavy metals or pesticides, which are related to possible contamination of the plant, could also benefit from these applications. The portability of this equipment open doors to new possibilities, enabling direct plant monitoring without the necessity of harvesting.

**Author Contributions:** Conceptualization, M.d.C.D.-L. and M.T.G.-V.; writing—original draft preparation, M.d.C.D.-L. and M.T.G.-V.; writing—review and editing, M.d.C.D.-L., M.T.G.-V., V.S.d.M. and C.F.-V.; visualization, M.d.C.D.-L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** Authors M.d.C.D.-L., M.T.G.-V., V.S.d.M. and C.F.-V. are employed by the company Phytoplant Research SLU. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Zeine, R.R.; Teasdale, B.W. *Medical Cannabis and the Effects of Cannabinoids on Fighting Cancer, Multiple Sclerosis, Epilepsy, Parkinson's, and Other Neurodegenerative Diseases*; IGI Global: Hershey, PA, USA, 2023; ISBN 1668456532.
2. EMCDDA. *EMCDDA Cannabis Laws in Europe. Questions and Answers for Policymaking*; EMCDDA: Lisbon, Portugal, 2023.
3. Sgrò, S.; Lavezzi, B.; Caprari, C.; Polito, M.; D'Elia, M.; Lago, G.; Furlan, G.; Girotti, S.; Ferri, E.N. Delta9-THC Determination by the EU Official Method: Evaluation of Measurement Uncertainty and Compliance Assessment of Hemp Samples. *Anal. Bioanal. Chem.* **2021**, *413*, 3399–3410. [[CrossRef](#)]
4. Hädener, M.; König, S.; Weinmann, W. Quantitative Determination of CBD and THC and Their Acid Precursors in Confiscated Cannabis Samples by HPLC-DAD. *Forensic Sci. Int.* **2019**, *299*, 142–150. [[CrossRef](#)] [[PubMed](#)]
5. Micalizzi, G.; Vento, F.; Alibrando, F.; Donnarumma, D.; Dugo, P.; Mondello, L. *Cannabis sativa* L.: A Comprehensive Review on the Analytical Methodologies for Cannabinoids and Terpenes Characterization. *J. Chromatogr. A* **2021**, *1637*, 461864. [[CrossRef](#)] [[PubMed](#)]
6. Izzo, L.; Castaldo, L.; Narváez, A.; Graziani, G.; Gaspari, A.; Rodríguez-Carrasco, Y.; Ritieni, A. Analysis of Phenolic Compounds in Commercial *Cannabis sativa* L. Inflorescences Using UHPLC-Q-Orbitrap HRMS. *Molecules* **2020**, *25*, 631. [[CrossRef](#)] [[PubMed](#)]
7. Radwan, M.M.; Chandra, S.; Gul, S.; ElSohly, M.A. Cannabinoids, Phenolics, Terpenes and Alkaloids of Cannabis. *Molecules* **2021**, *26*, 2774. [[CrossRef](#)] [[PubMed](#)]
8. Macherone, A. Chapter One-Comprehensive Analytical Testing of Cannabis and Hemp. In *Analysis of Cannabis*; Ferrer, I., Thurman, E.M., Eds.; Comprehensive Analytical Chemistry; Elsevier: Amsterdam, The Netherlands, 2020; Volume 90, pp. 3–29.
9. AL Ubeed, H.M.S.; Bhuyan, D.J.; Alsherbiny, M.A.; Basu, A.; Vuong, Q.V. A Comprehensive Review on the Techniques for Extraction of Bioactive Compounds from Medicinal Cannabis. *Molecules* **2022**, *27*, 604. [[CrossRef](#)]
10. Atapattu, S.N.; Johnson, K.R.D. Pesticide Analysis in Cannabis Products. *J. Chromatogr. A* **2020**, *1612*, 460656. [[CrossRef](#)] [[PubMed](#)]
11. Brown, A.K.; Xia, Z.; Bulloch, P.; Idowu, I.; Francisco, O.; Stetefeld, J.; Stout, J.; Zimmer, J.; Marvin, C.; Letcher, R.J.; et al. Validated Quantitative Cannabis Profiling for Canadian Regulatory Compliance-Cannabinoids, Aflatoxins, and Terpenes. *Anal. Chim. Acta* **2019**, *1088*, 79–88. [[CrossRef](#)]
12. Buchicchio, L.; Asselborn, L.; Schneider, S.; van Nieuwenhuyse, A.; Moris, G.; Schummer, C. Investigation of Aflatoxin and Ochratoxin A Contamination of Seized Cannabis and Cannabis Resin Samples. *Mycotoxin Res.* **2022**, *38*, 71–78. [[CrossRef](#)]
13. Wilcox, J.; Pazdanska, M.; Milligan, C.; Chan, D.; MacDonald, S.J.; Donnelly, C. Analysis of Aflatoxins and Ochratoxin A in Cannabis and Cannabis Products by LC-Fluorescence Detection Using Cleanup with Either Multiantibody Immunoaffinity Columns or an Automated System with In-Line Reusable Immunoaffinity Cartridges. *J. AOAC Int.* **2020**, *103*, 494–503. [[CrossRef](#)]
14. Cardenia, V.; Gallina Toschi, T.; Scappini, S.; Rubino, R.C.; Rodriguez-Estrada, M.T. Development and Validation of a Fast Gas Chromatography/Mass Spectrometry Method for the Determination of Cannabinoids in *Cannabis sativa* L. *J. Food Drug Anal.* **2018**, *26*, 1283–1292. [[CrossRef](#)]
15. Pellegrini, M.; Marchei, E.; Pacifici, R.; Pichini, S. A Rapid and Simple Procedure for the Determination of Cannabinoids in Hemp Food Products by Gas Chromatography-Mass Spectrometry. *J. Pharm. Biomed. Anal.* **2005**, *36*, 939–946. [[CrossRef](#)] [[PubMed](#)]
16. Lachenmeier, D.W.; Kroener, L.; Musshoff, F.; Madea, B. Determination of Cannabinoids in Hemp Food Products by Use of Headspace Solid-Phase Microextraction and Gas Chromatography? *Mass Spectrometry. Anal. Bioanal. Chem.* **2004**, *378*, 183–189. [[CrossRef](#)]
17. Gul, W.; Gul, S.W.; Radwan, M.M.; Wanas, A.S.; Mehmedic, Z.; Khan, I.I.; Sharaf, M.H.M.; ElSohly, M.A. Determination of 11 Cannabinoids in Biomass and Extracts of Different Varieties of Cannabis Using High-Performance Liquid Chromatography. *J. AOAC Int.* **2015**, *98*, 1523–1528. [[CrossRef](#)] [[PubMed](#)]
18. McRae, G.; Melanson, J.E. Quantitative Determination and Validation of 17 Cannabinoids in Cannabis and Hemp Using Liquid Chromatography-Tandem Mass Spectrometry. *Anal. Bioanal. Chem.* **2020**, *412*, 7381–7393. [[CrossRef](#)] [[PubMed](#)]
19. Stolker, A.A.M.; van Schoonhoven, J.; de Vries, A.J.; Bobeldijk-Pastorova, I.; Vaes, W.H.J.; van den Berg, R. Determination of Cannabinoids in Cannabis Products Using Liquid Chromatography-Ion Trap Mass Spectrometry. *J. Chromatogr. A* **2004**, *1058*, 143–151. [[CrossRef](#)]
20. Birenboim, M.; Kengisbuch, D.; Chalupowicz, D.; Maurer, D.; Barel, S.; Chen, Y.; Fallik, E.; Paz-Kagan, T.; Shimshoni, J.A. Use of Near-Infrared Spectroscopy for the Classification of Medicinal Cannabis Cultivars and the Prediction of Their Cannabinoid and Terpene Contents. *Phytochemistry* **2022**, *204*, 113445. [[CrossRef](#)]
21. Pourseyed Lazarjani, M.; Torres, S.; Hooker, T.; Fowlie, C.; Young, O.; Seyfoddin, A. Methods for Quantification of Cannabinoids: A Narrative Review. *J. Cannabis Res.* **2020**, *2*, 35. [[CrossRef](#)] [[PubMed](#)]
22. Yao, S.; Ball, C.; Miyagusuku-Cruzado, G.; Giusti, M.M.; Aykas, D.P.; Rodriguez-Saona, L.E. A Novel Handheld FT-NIR Spectroscopic Approach for Real-Time Screening of Major Cannabinoids Content in Hemp. *Talanta* **2022**, *247*, 123559. [[CrossRef](#)] [[PubMed](#)]
23. García-Valverde, M.T.; Sánchez-Carnerero Callado, C.; Díaz-Liñán, M.C.; Sánchez de Medina, V.; Hidalgo-García, J.; Nadal, X.; Hanuš, L.; Ferreiro-Vera, C. Effect of Temperature in the Degradation of Cannabinoids: From a Brief Residence in the Gas Chromatography Inlet Port to a Longer Period in Thermal Treatments. *Front. Chem.* **2022**, *10*, 1038729. [[CrossRef](#)]
24. Vacek, J.; Vostalova, J.; Papouskova, B.; Skarupova, D.; Kos, M.; Kabelac, M.; Storch, J. Antioxidant Function of Phytocannabinoids: Molecular Basis of Their Stability and Cytoprotective Properties under UV-Irradiation. *Free Radic. Biol. Med.* **2021**, *164*, 258–270. [[CrossRef](#)]

25. Yangsud, J.; Santanasasuwan, S.; Ahkkarachinoreh, P.; Maha, A.; Madaka, F.; Suksaeree, J.; Songsak, T.; Vutthipong, A.; Monton, C. Stability of Cannabidiol,  $\Delta^9$ -Tetrahydrocannabinol, and Cannabinol under Stress Conditions. *Adv. Tradit. Med.* **2021**, *21*, 475–484. [[CrossRef](#)]
26. Golombek, P.; Müller, M.; Barthlott, I.; Sproll, C.; Lachenmeier, D.W. Conversion of Cannabidiol (CBD) into Psychotropic Cannabinoids Including Tetrahydrocannabinol (THC): A Controversy in the Scientific Literature. *Toxics* **2020**, *8*, 41. [[CrossRef](#)]
27. Czégény, Z.; Nagy, G.; Babinszki, B.; Bajtel, Á.; Sebestyén, Z.; Kiss, T.; Csupor-Löffler, B.; Tóth, B.; Csupor, D. CBD, a Precursor of THC in e-Cigarettes. *Sci. Rep.* **2021**, *11*, 8951. [[CrossRef](#)] [[PubMed](#)]
28. Hazekamp, A.; Peltenburg, A.; Verpoorte, R.; Giroud, C. Chromatographic and Spectroscopic Data of Cannabinoids from *Cannabis sativa* L. *J. Liq. Chromatogr. Relat. Technol.* **2005**, *28*, 2361–2382. [[CrossRef](#)]
29. Mano-Sousa, B.J.; Maia, G.A.S.; Lima, P.L.; Campos, V.A.; Negri, G.; Chequer, F.M.D.; Duarte-Almeida, J.M. Color Determination Method and Evaluation of Methods for the Detection of Cannabinoids by Thin-layer Chromatography (TLC). *J. Forensic Sci.* **2021**, *66*, 854–865. [[CrossRef](#)]
30. Leite, J.D.A.; de Oliveira, M.V.L.; Conti, R.; Borges, W.d.S.; Rosa, T.R.; Filgueiras, P.R.; Lacerda, V.; Romão, W.; Neto, Á.C. Extraction and Isolation of Cannabinoids from Marijuana Seizures and Characterization by <sup>1</sup>H NMR Allied to Chemometric Tools. *Sci. Justice* **2018**, *58*, 355–365. [[CrossRef](#)] [[PubMed](#)]
31. Malík, M.; Praus, L.; Tlustoš, P. Comparison of Recirculation and Drain-to-Waste Hydroponic Systems in Relation to Medical Cannabis (*Cannabis sativa* L.). *Plants. Ind. Crops Prod.* **2023**, *202*, 117059. [[CrossRef](#)]
32. Michaud, K. Using Flame Atomic Absorption Spectroscopy to Analyze for Heavy Metal Content in Cannabidiol Products. Ph.D. Thesis, University of Massachusetts Lowell, Lowell, MA, USA, 2021.
33. Menezes, I.M.N.R.; Nascimento, P.D.A.; Yamamoto, C.I.; Oliveira, A. Evaluation of Trace Elements in Cannabis Products. *J. Food Compos. Anal.* **2022**, *113*, 104721. [[CrossRef](#)]
34. Potin, F.; Lubbers, S.; Husson, F.; Saurel, R. Hemp (*Cannabis sativa* L.) Protein Extraction Conditions Affect Extraction Yield and Protein Quality. *J. Food Sci.* **2019**, *84*, 3682–3690. [[CrossRef](#)]
35. Craven, C.B.; Wawryk, N.; Jiang, P.; Liu, Z.; Li, X.-F. Pesticides and Trace Elements in Cannabis: Analytical and Environmental Challenges and Opportunities. *J. Environ. Sci.* **2019**, *85*, 82–93. [[CrossRef](#)]
36. Su, K.; Maghirang, E.; Tan, J.W.; Yoon, J.Y.; Armstrong, P.; Kachroo, P.; Hildebrand, D. NIR Spectroscopy for Rapid Measurement of Moisture and Cannabinoid Contents of Industrial Hemp (*Cannabis sativa*). *Ind. Crops Prod.* **2022**, *184*, 115007. [[CrossRef](#)]
37. Alonso-Esteban, J.I.; Pinela, J.; Ćirić, A.; Calhella, R.C.; Soković, M.; Ferreira, I.C.F.R.; Barros, L.; Torija-Isasa, E.; de Sánchez-Mata, M.C. Chemical Composition and Biological Activities of Whole and Dehulled Hemp (*Cannabis sativa* L.). *Seeds. Food Chem.* **2022**, *374*, 131754. [[CrossRef](#)] [[PubMed](#)]
38. Pattnaik, F.; Nanda, S.; Kumar, V.; Naik, S.; Dalai, A.K.; Mohanty, M.K. Extraction of Sugars and Cellulose Fibers from Cannabis Stems by Hydrolysis, Pulping, and Bleaching. *Chem. Eng. Technol.* **2022**, *45*, 962–970. [[CrossRef](#)]
39. Nie, B.; Henion, J.; Ryona, I. The Role of Mass Spectrometry in the Cannabis Industry. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 719–730. [[CrossRef](#)] [[PubMed](#)]
40. Nelson, J.; Jones, C.; Heckle, S.; Anderson, L. Determination of Heavy Metals in a Variety of Cannabis and Cannabis-Derived Products, First Action 2021 03. *J. AOAC Int.* **2022**, *105*, 1640–1651. [[CrossRef](#)]
41. Laza, A.; Orozco, E.; Baldo, M.F.; Raba, J.; Aranda, P.R. Determination of Arsenic (V) in Cannabis Oil by Adsorption on Multiwall Carbon Nanotubes Thin Film Using XRF Technique. *Microchem. J.* **2020**, *158*, 105265. [[CrossRef](#)]
42. Tran, J.; Vassiliadis, S.; Elkins, A.C.; Cogan, N.O.I.; Rochfort, S.J. Developing Prediction Models Using Near-Infrared Spectroscopy to Quantify Cannabinoid Content in *Cannabis sativa*. *Sensors* **2023**, *23*, 2607. [[CrossRef](#)] [[PubMed](#)]
43. Deidda, R.; Coppey, F.; Damergi, D.; Schelling, C.; Coïc, L.; Veuthey, J.-L.; Sacré, P.-Y.; De Bleye, C.; Hubert, P.; Esseiva, P.; et al. New Perspective for the In-Field Analysis of Cannabis Samples Using Handheld near-Infrared Spectroscopy: A Case Study Focusing on the Determination of  $\Delta^9$ -Tetrahydrocannabinol. *J. Pharm. Biomed. Anal.* **2021**, *202*, 114150. [[CrossRef](#)]
44. Lu, Y.; Li, X.; Young, S.; Li, X.; Linder, E.; Suchoff, D. Hyperspectral Imaging with Chemometrics for Non-Destructive Determination of Cannabinoids in Floral and Leaf Materials of Industrial Hemp (*Cannabis sativa* L.). *Comput. Electron. Agric.* **2022**, *202*, 107387. [[CrossRef](#)]
45. Rossi, G.B.; Lozano, V.A. Simultaneous Determination of Quality Parameters in Yerba Mate (*Ilex Paraguariensis*) Samples by Application of near-Infrared (NIR) Spectroscopy and Partial Least Squares (PLS). *LWT* **2020**, *126*, 109290. [[CrossRef](#)]
46. Sajid, M.; Plotka-Wasyłka, J. Green Analytical Chemistry Metrics: A Review. *Talanta* **2022**, *238*, 123046. [[CrossRef](#)] [[PubMed](#)]
47. Gloerfelt-Tarp, F.; Hewavitharana, A.K.; Mieog, J.; Palmer, W.M.; Fraser, F.; Ansari, O.; Kretschmar, T. Using a Global Diversity Panel of *Cannabis sativa* L. to Develop a near InfraRed-Based Chemometric Application for Cannabinoid Quantification. *Sci. Rep.* **2023**, *13*, 2253. [[CrossRef](#)] [[PubMed](#)]
48. Sánchez-Carnerero Callado, C.; Núñez-Sánchez, N.; Casano, S.; Ferreira-Vera, C. The Potential of near Infrared Spectroscopy to Estimate the Content of Cannabinoids in *Cannabis sativa* L.: A Comparative Study. *Talanta* **2018**, *190*, 147–157. [[CrossRef](#)]
49. Deidda, R.; Dispas, A.; De Bleye, C.; Hubert, P.; Ziemons, É. Critical Review on Recent Trends in Cannabinoid Determination on Cannabis Herbal Samples: From Chromatographic to Vibrational Spectroscopic Techniques. *Anal. Chim. Acta* **2022**, *1209*, 339184. [[CrossRef](#)] [[PubMed](#)]

50. Pereira, J.F.Q.; Pimentel, M.F.; Amigo, J.M.; Honorato, R.S. Detection and Identification of *Cannabis sativa* L. Using near Infrared Hyperspectral Imaging and Machine Learning Methods. A Feasibility Study. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2020**, *237*, 118385. [[CrossRef](#)]
51. Borille, B.T.; Marcelo, M.C.A.; Ortiz, R.S.; Mariotti, K.d.C.; Ferrão, M.F.; Limberger, R.P. Near Infrared Spectroscopy Combined with Chemometrics for Growth Stage Classification of Cannabis Cultivated in a Greenhouse from Seized Seeds. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2017**, *173*, 318–323. [[CrossRef](#)]
52. Duchateau, C.; Kauffmann, J.; Canfyn, M.; Stévigny, C.; De Braekeleer, K.; Deconinck, E. Discrimination of Legal and Illegal Cannabis Spp. According to European Legislation Using near Infrared Spectroscopy and Chemometrics. *Drug Test. Anal.* **2020**, *12*, 1309–1319. [[CrossRef](#)]
53. Singh, M.; Bowman, M.J.; Berhow, M.A.; Price, N.P.J.; Liu, S.X. Application of near Infrared Spectroscopy for Determination of Relationship between Crop Year, Maturity Group, Location, and Carbohydrate Composition in Soybeans. *Crop Sci.* **2021**, *61*, 2409–2422. [[CrossRef](#)]
54. Beć, K.B.; Grabska, J.; Huck, C.W. NIR Spectroscopy of Natural Medicines Supported by Novel Instrumentation and Methods for Data Analysis and Interpretation. *J. Pharm. Biomed. Anal.* **2021**, *193*, 113686. [[CrossRef](#)]
55. Jarén, C.; Zambrana, P.C.; Pérez-Roncal, C.; López-Maestresalas, A.; Ábrego, A.; Arazuri, S. Potential of NIRS Technology for the Determination of Cannabinoid Content in Industrial Hemp (*Cannabis sativa* L.). *Agronomy* **2022**, *12*, 938. [[CrossRef](#)]
56. Jianhua Yi Yifei Sun, Z.Z.N.L.; Lu, J. Near-Infrared Reflectance Spectroscopy for the Prediction of Chemical Composition in Walnut Kernel. *Int. J. Food Prop.* **2017**, *20*, 1633–1642. [[CrossRef](#)]
57. Chen, Z.; de Boves Harrington, P.; Griffin, V.; Griffin, T. In Situ Determination of Cannabidiol in Hemp Oil by Near-Infrared Spectroscopy. *J. Nat. Prod.* **2021**, *84*, 2851–2857. [[CrossRef](#)] [[PubMed](#)]
58. Roggo, Y.; Chalus, 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](#)]
59. Jackson, J.E. Getting Started. In *A User's Guide to Principal Components*; John Wiley & Sons: Hoboken, NJ, USA, 1991; pp. 4–25.
60. Bro, R.; Smilde, A.K. Principal Component Analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [[CrossRef](#)]
61. Wilson, N.; Heinrich, M. The Use of Near Infrared Spectroscopy to Discriminate between THC-Rich and Hemp Forms of Cannabis. *Planta Med.* **2006**, *72*, P\_260. [[CrossRef](#)]
62. Brereton, R.G. Pattern Recognition. In *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons, Ltd: Chichester, UK, 2003; pp. 183–269.
63. Gülağız, F.K.; Şahin, S. Comparison of Hierarchical and Non-Hierarchical Clustering Algorithms. *Int. J. Comput. Eng. Inf. Technol.* **2017**, *9*, 6–14.
64. Wold, S. Pattern Recognition by Means of Disjoint Principal Components Models. *Pattern Recognit.* **1976**, *8*, 127–139. [[CrossRef](#)]
65. Wold, S.; Sjöström, M. SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. In *Chemometrics: Theory and Application*; ACS Publications: Washington, DC, USA, 1977; pp. 243–282.
66. Chen, Z.; de Boves Harrington, P. Automatic Soft Independent Modeling for Class Analogies. *Anal. Chim. Acta* **2019**, *1090*, 47–56. [[CrossRef](#)]
67. Vitale, R.; Cocchi, M.; Biancolillo, A.; Ruckebusch, C.; Marini, F. Class Modelling by Soft Independent Modelling of Class Analogy: Why, When, How? A Tutorial. *Anal. Chim. Acta* **2023**, *1270*, 341304. [[CrossRef](#)]
68. Lee, L.C.; Liong, C.-Y.; Jemain, A.A. Partial Least Squares-Discriminant Analysis (PLS-DA) for Classification of High-Dimensional (HD) Data: A Review of Contemporary Practice Strategies and Knowledge Gaps. *Analyst* **2018**, *143*, 3526–3539. [[CrossRef](#)]
69. Risoluti, R.; Gullifa, G.; Battistini, A.; Materazzi, S. Monitoring of Cannabinoids in Hemp Flours by MicroNIR/Chemometrics. *Talanta* **2020**, *211*, 120672. [[CrossRef](#)]
70. Risoluti, R.; Gullifa, G.; Battistini, A.; Materazzi, S. MicroNIR/Chemometrics: A New Analytical Platform for Fast and Accurate Detection of  $\Delta^9$ -Tetrahydrocannabinol (THC) in Oral Fluids. *Drug Alcohol Depend.* **2019**, *205*, 107578. [[CrossRef](#)]
71. San Nicolas, M.; Villate, A.; Alvarez-Mora, I.; Olivares, M.; Aizpurua-Olaizola, O.; Amigo, J. *NIR-Hyperspectral Imaging and Machine Learning for Non-Invasive Chemotype Classification in Cannabis sativa L.*; Elsevier: Amsterdam, The Netherlands, 2023.
72. Shaffer, R.E.; Rose-Pehrsson, S.L.; McGill, R.A. A Comparison Study of Chemical Sensor Array Pattern Recognition Algorithms. *Anal. Chim. Acta* **1999**, *384*, 305–317. [[CrossRef](#)]
73. Wang, H.; Chen, D.; Chen, Y. The Integrated Strategy of Pattern Classification and Its Application in Chemistry. *Chemom. Intell. Lab. Syst.* **2004**, *70*, 23–31. [[CrossRef](#)]
74. Specht, D.F. Probabilistic Neural Networks. *Neural Netw.* **1990**, *3*, 109–118. [[CrossRef](#)]
75. Park, J.; Sandberg, I.W. Universal Approximation Using Radial-Basis-Function Networks. *Neural Comput.* **1991**, *3*, 246–257. [[CrossRef](#)] [[PubMed](#)]
76. Valinger, D.; Jurina, T.; Šain, A.; Matešić, N.; Panić, M.; Benković, M.; Gajdoš Kljusurić, J.; Jurinjak Tušek, A. Development of ANN Models Based on Combined UV-vis-NIR Spectra for Rapid Quantification of Physical and Chemical Properties of Industrial Hemp Extracts. *Phytochem. Anal.* **2021**, *32*, 326–338. [[CrossRef](#)] [[PubMed](#)]
77. Aasim, M.; Kadir, R.; Akgur, O.; Yildirim, B.; Mustafa, Z.; Nadeem, M.A.; Baloch, F.S.; Karakoy, T.; Yilmaz, G. Machine Learning (ML) Algorithms and Artificial Neural Network for Optimizing In Vitro Germination and Growth Indices of Industrial Hemp (*Cannabis sativa* L.). *Ind. Crops Prod.* **2022**, *181*, 114801. [[CrossRef](#)]

78. Coomans, D.; Massart, D.L. Alternative K-Nearest Neighbour Rules in Supervised Pattern Recognition. *Anal. Chim. Acta* **1982**, *138*, 153–165. [CrossRef]
79. Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugen.* **1936**, *7*, 179–188. [CrossRef]
80. Xanthopoulos, P.; Pardalos, P.M.; Trafalis, T.B. Linear Discriminant Analysis. In *Robust Data Mining*; Pardalos, P.M., Pintér, J.D., Robinson, S.M., Terlaky, T., Thai, M.T., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 27–33.
81. Massart, D.L.; Vandeginste, B.G.M.; Deming, S.N.; Michotte, Y.; Kaufman, L. Supervised Pattern Recognition. In *Chemometrics: A Textbook*; Vandeginste, B.G.M., Kaufman, L., Eds.; Elsevier: Amsterdam, The Netherlands, 2003; Volume 2, pp. 385–413.
82. Borregaard, T.; Nielsen, H.; Nørgaard, L.; Have, H. Crop–Weed Discrimination by Line Imaging Spectroscopy. *J. Agric. Eng. Res.* **2000**, *75*, 389–400. [CrossRef]
83. Geskovski, N.; Stefov, G.; Gigopulu, O.; Stefov, S.; Huck, C.W.; Makreski, P. Mid-Infrared Spectroscopy as Process Analytical Technology Tool for Estimation of THC and CBD Content in Cannabis Flowers and Extracts. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2021**, *251*, 119422. [CrossRef] [PubMed]
84. Food and Drug Administration. *International Council for Harmonisation Validation of Analytical Procedures Q2(R2)*; Food and Drug Administration: Rockville, MD, USA, 2022.
85. Van Wyngaard, E.; Blancquaert, E.; Nieuwoudt, H.; Aleixandre-Tudo, J.L. Infrared Spectroscopy and Chemometric Applications for the Qualitative and Quantitative Investigation of Grapevine Organs. *Front. Plant Sci.* **2021**, *12*, 723247. [CrossRef] [PubMed]
86. Helland, I. Partial Least Squares Regression. In *Wiley StatsRef: Statistics Reference Online*; Wiley: New York, NY, USA, 2014.
87. Espel Grekopoulos, J. 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. [CrossRef] [PubMed]
88. Toonen, M.A.J.; Maliepaard, C.; Reijmers, T.H.; van der Voet, H.; Mastebroek, H.D.; van den Broeck, H.C.; Ebskamp, M.J.M.; Kessler, W.; Kessler, R.W. Predicting the Chemical Composition of Fibre and Core Fraction of Hemp (*Cannabis sativa* L.). *Euphytica* **2004**, *140*, 39–45. [CrossRef]
89. Williams, P. The RPD Statistic: A Tutorial Note. *NIR News* **2014**, *25*, 22–26. [CrossRef]
90. Jamwal, R.; Amit, Kumari, S.; Balan, B.; Kelly, S.; Cannavan, A.; Singh, D.K. Rapid and Non-Destructive Approach for the Detection of Fried Mustard Oil Adulteration in Pure Mustard Oil via ATR-FTIR Spectroscopy–Chemometrics. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2021**, *244*, 118822. [CrossRef]
91. Townsend, D.; Eustis, I.; Lewis, M.; Rodgers, S.; Smith, K.; Bohman, A. *The Determination of Total THC and CBD Content in Cannabis Flower by Fourier Transform Near Infrared Spectroscopy*; PerkinElmer: Waltham, MA, USA, 2018; pp. 1–5.
92. Marill, K.A. Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression. *Acad. Emerg. Med.* **2004**, *11*, 94–102. [CrossRef]
93. Zou, J.; Han, Y.; So, S.-S. Overview of Artificial Neural Networks. In *Artificial Neural Networks—Methods and Applications*; Humana Press: Totowa, NJ, USA, 2008; pp. 14–22.
94. Mammone, A.; Turchi, M.; Cristianini, N. Support Vector Machines. *WIREs Comput. Stat.* **2009**, *1*, 283–289. [CrossRef]
95. Pisner, D.A.; Schnyer, D.M. Support Vector Machine. In *Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 101–121.
96. Deidda, R.; Damergi, D.; Coppey, F.; De Bleye, C.; Coic, L.; Sacre, P.-Y.; Hubert, P.; Ziemons, E.; Esseiva, P.; Veuthey, J.-L. Handheld Near Infrared Spectroscopy for Cannabis Analysis: From the Analytical Problem to the Chemometric Solution. *Chimométrie* **2020**, *1*, 2–3. Available online: <https://orbi.uliege.be/handle/2268/244774> (accessed on 10 August 2023).
97. Adesokan, M.; Alamu, E.O.; Otegbayo, B.; Maziya-Dixon, B. A Review of the Use of Near-Infrared Hyperspectral Imaging (NIR-HSI) Techniques for the Non-Destructive Quality Assessment of Root and Tuber Crops. *Appl. Sci.* **2023**, *13*, 5226. [CrossRef]
98. Lu, B.; Dao, P.D.; Liu, J.; He, Y.; Shang, J. Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sens.* **2020**, *12*, 2659. [CrossRef]
99. Holmes, W.S.; Ooi, M.P.-L.; Abeysekera, S.; Kuang, Y.C.; Simpkin, R.; Caddie, M.; Nowak, J.; Demidenko, S. On Machine Learning Methods to Estimate Cannabidiolic Acid Content of *Cannabis sativa* L. from near-Infrared Hyperspectral Imaging. In Proceedings of the 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Kuala Lumpur, Malaysia, 22–25 May 2023; IEEE: Piscataway, NJ, USA; pp. 1–6.
100. Heil, K.; Schmidhalter, U. An Evaluation of Different NIR-Spectral Pre-Treatments to Derive the Soil Parameters C and N of a Humus–Clay-Rich Soil. *Sensors* **2021**, *21*, 1423. [CrossRef]
101. Abeysekera, S.K.; Robinson, A.; Ooi, M.P.-L.; Kuang, Y.C.; Manley-Harris, M.; Holmes, W.; Hirst, E.; Nowak, J.; Caddie, M.; Steinhorn, G.; et al. Sparse Reproducible Machine Learning for near Infrared Hyperspectral Imaging: Estimating the Tetrahydrocannabinolic Acid Concentration in *Cannabis sativa* L. *Ind. Crops Prod.* **2023**, *192*, 116137. [CrossRef]
102. Wang, C.; Liu, B.; Liu, L.; Zhu, Y.; Hou, J.; Liu, P.; Li, X. A Review of Deep Learning Used in the Hyperspectral Image Analysis for Agriculture. *Artif. Intell. Rev.* **2021**, *54*, 5205–5253. [CrossRef]

**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.