

Review

Challenges in the Use of AI-Driven Non-Destructive Spectroscopic Tools for Rapid Food Analysis

Wenyang Jia ¹, Konstantia Georgouli ^{1,2}, Jesus Martinez-Del Rincon ³ and Anastasios Koidis ^{1,*}

¹ Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, Belfast BT9 5DL, UK; wjia02@qub.ac.uk (W.J.); kgeorgouli01@qub.ac.uk (K.G.)

² Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

³ Institute of Electronics, Communications and Information Technology, Queen's University Belfast, Belfast BT3 9DT, UK; j.martinez-del-rincon@qub.ac.uk

* Correspondence: t.koidis@qub.ac.uk

Abstract: Routine, remote, and process analysis for foodstuffs is gaining attention and can provide more confidence for the food supply chain. A new generation of rapid methods is emerging both in the literature and in industry based on spectroscopy coupled with AI-driven modelling methods. Current published studies using these advanced methods are plagued by weaknesses, including sample size, abuse of advanced modelling techniques, and the process of validation for both the acquisition method and modelling. This paper aims to give a comprehensive overview of the analytical challenges faced in research and industrial settings where screening analysis is performed while providing practical solutions in the form of guidelines for a range of scenarios. After extended literature analysis, we conclude that there is no easy way to enhance the accuracy of the methods by using state-of-the-art modelling methods and the key remains that capturing good quality raw data from authentic samples in sufficient volume is very important along with robust validation. A comprehensive methodology involving suitable analytical techniques and interpretive modelling methods needs to be considered under a tailored experimental design whenever conducting rapid food analysis.

Keywords: chemometrics; spectroscopy; challenges; research; validation; methodology; modelling; guidelines



Citation: Jia, W.; Georgouli, K.; Martinez-Del Rincon, J.; Koidis, A. Challenges in the Use of AI-Driven Non-Destructive Spectroscopic Tools for Rapid Food Analysis. *Foods* **2024**, *13*, 846. <https://doi.org/10.3390/foods13060846>

Academic Editor: Zhiming Guo

Received: 2 February 2024

Revised: 5 March 2024

Accepted: 8 March 2024

Published: 10 March 2024



Copyright: © 2024 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

Official methods adopted by regulatory authorities to protect a food product, define its authenticity, and tackle food adulteration are often chemical, ‘targeted’ methods that are designed to identify and quantify specific known compounds or markers (i.e., targets) as major or minor ingredients within the food. One of the targeted methods is chromatography, which produces specific signal peaks after sample extraction following in-column separation. Similar output is produced using direct mass spectrometry analysis, depending on the complexity of the sample and the analytical needs [1]. These peaks are then translated to specific analytes and quantified. The concern with these methods is that they are generally slow, non-environmentally friendly and require substantial initial capital investment and experienced staff to perform them. A new generation of rapid methods is emerging based on vibrational spectroscopy (mid-infrared, near-infrared, and Raman spectroscopy) and other sensors such as nuclear magnetic resonance spectroscopy and vision technology, fuelled by the rise in AI-based analytics. For instance, the coupling of multispectral/hyperspectral imaging and e-nose can rapidly produce a highly complex signal based on the properties and molecular structure of the foodstuffs that are analysed [2]. In contrast with targeted methods, in “untargeted” analysis, the outcome of the determination is not derived from the precise isolation and quantification of specific, known markers but from the chemical or molecular fingerprint of analytes extracted by

the raw total spectrum, which is theoretically the sum of the signal from the sample's constituents [3]. A common characteristic of all untargeted methods is that they provide a large number of signal peaks (in this case, absorbance or reflectance) coded as chemical information, presented in the form of numerous variables (wavelength or wavenumbers). The chemical information in the spectral data is often included within a lot of instrumental and environmental 'noise' [4]. The extraction and interpretation of this useful information are usually difficult and laborious. To process these multivariate data, a combination of mathematical and statistical techniques is used, commonly known as 'chemometrics'. Nowadays, chemometrics, also referred to as "machine learning", methods are used for developing prediction/decision models based on the knowledge acquired from calibration data using the principles of pattern recognition theory [5,6]. Recently, machine learning (ML) has been associated with the generic term 'Artificial Intelligence' (AI), although this is inaccurate as AI is a much broader term [7]. Thanks to these methods, information in multidimensional data can be automatically processed, elucidated, interpreted, and discriminated, resulting in advanced models that can authenticate, detect adulteration, and determine intrinsic quality parameters in foodstuffs [8]. Although not designed to completely replace the official regulatory 'targeted' methods, untargeted methods still have their place in the routine analysis and screening of foodstuffs. Especially after the pandemic, routine, remote, and process analysis for foodstuffs has been gaining attention and can provide more confidence for all stakeholders involved [9].

Currently, research on chemometrics and its application in the research field of food authenticity is thriving. The growth in the number of papers in the area is 700% in less than 20 years (Figure 1).

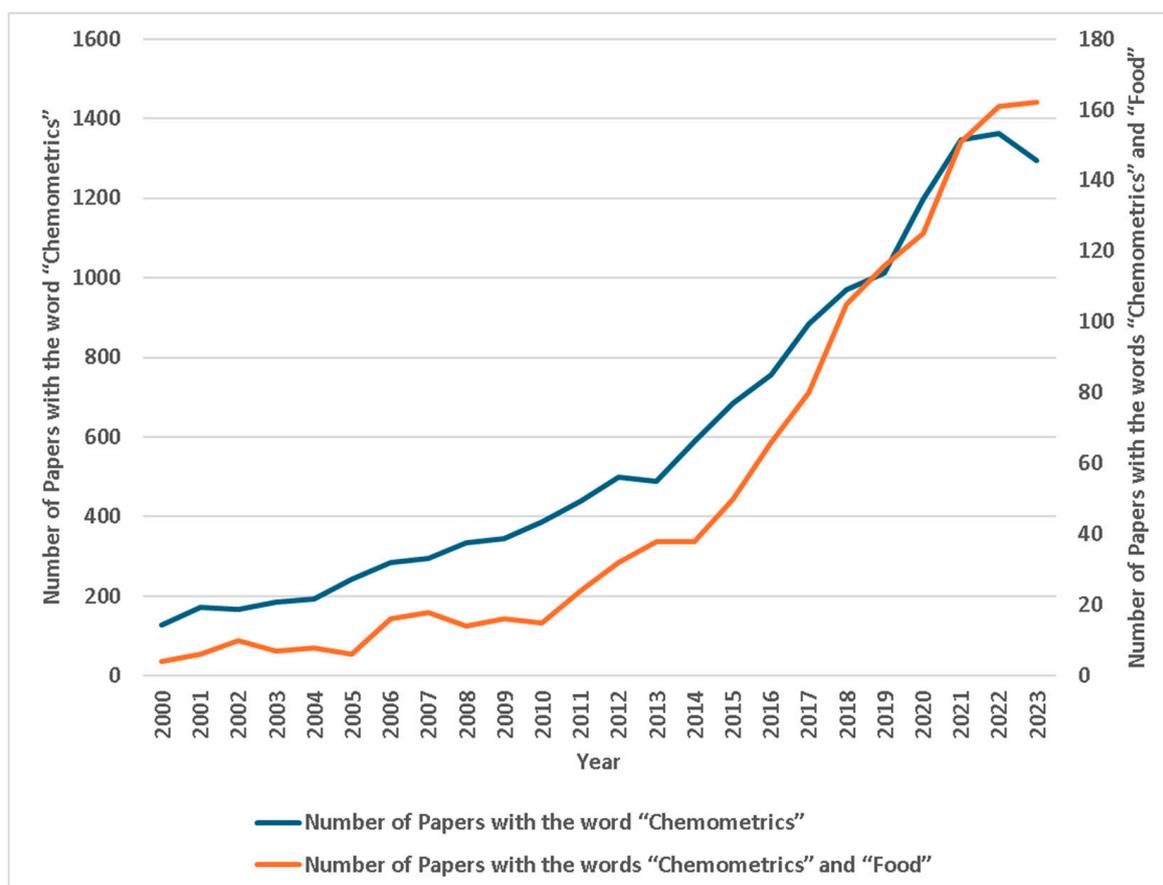


Figure 1. Rise of research papers per year with the word "chemometrics" (main axis) and a combination of "chemometrics" and "food" (secondary axis).

This figure depicts the rise in the number of research papers per year containing the keywords “chemometrics” and a combination of “chemometrics” and “Food”. The data show a significant increase over the years, illustrating the growing interest and research activity in this field. As the term ‘chemometrics’ is now replaced by ‘ML’ or ‘AI’, the actual studies related to the field of rapid analysis using the techniques described from 2022 and onwards could be even higher than demonstrated here (Figure 1).

Nevertheless, new chemometrics techniques are being developed, and existing techniques that are sufficiently generic and based on a solid mathematical background are being improved to make them applicable to a number of problems. In the literature, chemometrics has been used as an essential tool for non-destructive tools to assess various scenarios such as beer classification, rice discrimination, and sensory evaluation [10]. In addition, the use of chemometrics has enabled the unified analysis of data derived from more than one analytical technique in a new field of “data fusion” and, consequently, more information about the analytes, although improved accuracy is not always guaranteed [11].

Current chemometric implementations in the literature are plagued by some common problems. One of them is the number of unique samples that the research studies are based on for developing or validating chemometric models, which evidently appears rather small. This constrains the use of more advanced AI techniques such as deep neural networks and transformers, which are state-of-the-art in machine learning but require hundreds if not thousands of samples to be effectively trained [12]. Another issue is the variable selection techniques that are based on visual inspection of the data and not algorithmic or knowledge-driven methodologies. In limited capacity in some research publications, advanced chemometrics methods are being applied in excess in an effort to add depth to the existing results and discussion. However, there are cases where they arguably do not add to the narrative; they end up vaguely supporting the original arguments made by the rest of the techniques and usually do not serve the objective of the study in which they appear.

Based on the above, this review paper aims to critically evaluate the methodology used by the authors of chemometric-based studies, highlight the analytical methods and current challenges, and identify some of the major reasons for inconsistencies in the practices used. Finally, this review will also provide practical suggestions for academic and industrial research to improve the technical approach taken in future scientific studies in order to provide innovative solutions, tackle overrated activities, and ensure food authenticity.

2. Challenges and Sources of Error in Current Research Studies

The research challenges in food analysis with regard to instrumental analytical techniques, experimental design, the application of ML/AI/chemometrics, and model validation, as well as analysts’ perspectives, are critically discussed below.

2.1. Analytical Methods

Among the various analytical techniques developed, molecular (including genomics and proteomics) and chromatographic methodologies have been used mainly for food authentication studies over the last few years [13]. As implied above, these targeted methods, such as high-performance liquid chromatography (HPLC), offer accurate, precise, and reliable results. However, despite their benefits, these methods are typically time-consuming and destructive and require expensive equipment and highly skilled personnel, which limits their application to the real-time screening of processes required by industry, where timely decisions are an essential requirement [14]. This general trend toward the use of molecular and chromatographic techniques poses fundamental questions regarding their adoption in an industrial setting. Specifically, the heterogeneity and physical nature of food samples usually require sample preparation, which is time-consuming, laborious, and destructive for classical targeted analytical techniques [15]. Besides sample preparation, the sample size used is small (<1 g) and therefore unlikely to be representative of the high-throughput production in industry, adding another angle that is important: the frequency

of sampling, which by design is low when using targeted analytical methods. Most notably, these methods may allow other types of adulterants to remain undetected due to the targeted analysis of specific compounds or analytes of interest [16].

On the contrary, spectroscopic techniques such as vibrational spectroscopy enable rapid, easy, non-destructive, high-throughput, and relatively low-cost screening techniques with minimal sample preparation, with them being capable of online food analysis [17]. Here, the spectral information extracted is related to the molecular structure and, specifically, the vibrational behaviour of molecular bonds within a sample after the interaction of the sample with light [18]. Moreover, vibrational spectroscopy is characterised as a green analytical tool since it is an environmentally friendly option, minimising sample pretreatment with hazardous reagents and solvents [19]. Other techniques providing multivariate data such as hyper- and multispectral imaging have been developed as well to ensure food authenticity. As an advanced spectroscopy technique, hyperspectral imaging captures the spectrum of every pixel from the foodstuffs of interest, with it being developed regarding the authenticity of different types of foodstuffs, such as flour, meat, fruits, and oil.

Although existing spectroscopic techniques coupled with chemometrics are capable of meeting food screening demands, a set of specific and challenging tasks should be overcome for their inclusion in real-world industrial applications and official methods (Figure 2). Therefore, research is needed to develop both reliable, state-of-the-art spectroscopic sensors and robust ML/AI chemometric models, as discussed in Section 2.3.

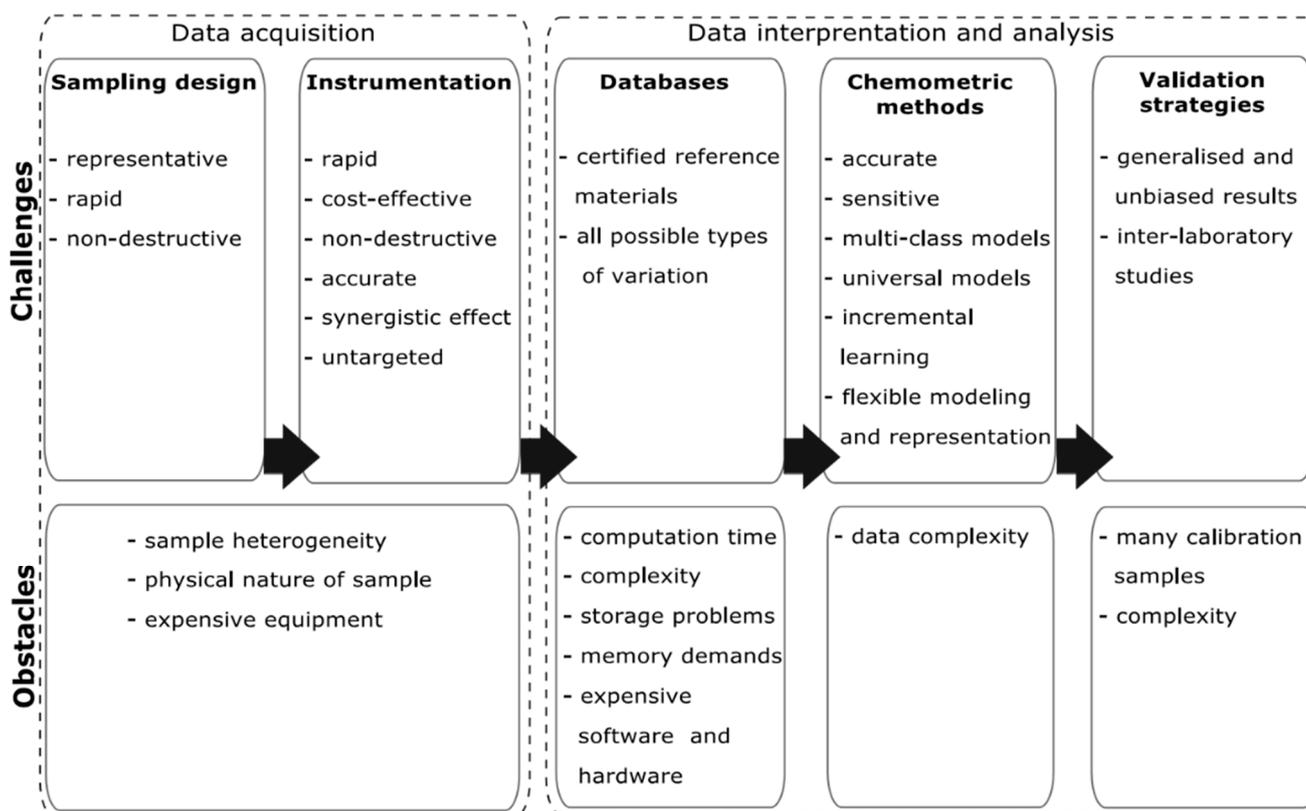


Figure 2. Modern research challenges in rapid food analysis.

2.2. Experimental Design

The first consideration when designing a non-destructive spectroscopic analysis is the sampling plan, which can include certified or authentic food samples. Nowadays, most experimental studies that deal with food authentication are conducted to verify the labelling information about a specific food property (e.g., a certain geographic origin, variety, or composition) and to ensure its purity and quality. As a result, the absence of libraries of authentic samples (unlike the pharma industry where agencies such as Pharmacopeia

exist) leads the literature authors to purchase food samples from unreliable sources, such as the retail market, for calibrating and validating their methods [20]. This compromises the integrity of the results since the authenticity of retail samples is not confirmed [21]. Additionally, the experimental calibration data used in food studies mainly focus on a specific source of variation for the development of classification models [22]. Furthermore, only a few studies capture some of the variability in the original samples introduced by different influential natural factors (e.g., climate, temperature, and geographical location) and other factors such as processing and, in most cases, storage conditions. The building of comprehensive databases of certified reference food materials for authenticity will allow for accurate and reliable food studies, which constitutes not only an important first step but also a considerable challenge [23].

To extend the use of non-destructive spectroscopic methods, the variability in classification models, from the acquisition of data to the prediction of the sample's properties, needs careful consideration of factoring in differences derived from instruments, sampling, and analyst operations. The experimental design, from conception to data acquisition and beyond, changes significantly depending on the setting, with the research/academic environment having more control over the conditions (temperature, light, moisture, and motion) compared to industry. Also, sample conditions can dynamically change with the sample's journey from industry to the laboratory, where most of the literature studies have been conducted. Some studies obtained samples under certain processing steps, while other studies scanned packaged samples [24], which will influence the model's performance on newly presented samples. Naturally, the question arises as to whether the lack of integrating this variability into the experimental design can cause data (in)consistency problems and therefore very low model applicability.

For this purpose, it is evident that the calibration data should encompass as many sources of variation as possible for a well-designed study. Specifically, the incorporation of not only natural (e.g., seasonal, cultivar, environmental, food varieties or types, etc.) and processing (production parameters and storage) variability into a food sample but also other secondary types of real-world experimental variation (instrumental, sampling, and human factors) is necessary for the creation of universal and robust chemometric models and their transition to industrial online applications [25].

2.3. AI/ML/Chemometrics Pipelines

Over the last decade, both established and newly developed chemometrics techniques have been applied in food applications to tackle various emerging problems (Table 1). Moreover, chemometric approaches have been used for multi-analytical methods that are outside the field of non-destructive analysis, including untargeted HPLC [26], nuclear magnetic resonance [27], elemental metabolomics [28], mass spectrometry [29], and stable isotope analysis [30]. In this regard, there are specific chemometric techniques that have found wide applicability compared to others in the literature. In particular, principal component analysis (PCA), partial least-squares–discriminant analysis (PLS-DA), and partial least-squares regression (PLS-R) are by far the most widely used techniques in exploratory, classification, and regression analysis, respectively [31]. Nonetheless, these techniques are not always effective for solving non-linear analytical problems due to their assumption of linearity [32]. Additionally, the performance of the PLS-DA classifier is worsened by the increase in the number of model classes. This hinders its application to more complicated multi-class classification problems [33].

Table 1. Development of chemometrics in food applications using non-destructive spectroscopic tools.

Chemometrics	Technique	Applications	Performance (as Reported)	Refs
LDA	NIR	Alcohol degree of Chinese liquor	$R^2 = 0.96$	[34]
LDA	Raman	Adulterants (fructose corn syrup and maltose syrup) in honey	Accuracy = 91%	[35]

Table 1. Cont.

Chemometrics	Technique	Applications	Performance (as Reported)	Refs
QDA	HSI	Detection of anthracnose in mango fruits	KAPPA = 90%	[36]
KNN	NIR	Geographic origin discrimination of millet	Discrimination rate = 0.99	[37]
KNN	Raman	Detection of adulteration of extra virgin olive oil	Classification rate = 0.79	[38]
KNN	MIR	Monitoring of soluble pectin content in orange juice	Accuracy = 85%	[39]
SIMCA	NIR	Identification of the types of fat added to feed	Sensitivities = 100%	[40]
SIMCA	Raman	Detection of milk powder adulteration	Accuracy = 97%	[41]
PCR	Raman	Honey adulteration	Accuracy = 96.54%	[42]
MLR	NIR	Soluble solids content of tea soft drinks	R ² = 0.98	[43]
MLR	MIR	Mycotoxin deoxynivalenol (DON) in wheat	R ² = 0.99	[44]
PLS	NIR	Aflatoxigenic fungal contamination in rice	R ² = 0.67	[45]
PLS	SERS	Identification of food processing bacteria	Accuracy = 99%	[46]
PLS	MIR	Quality characteristics in pomegranate kernel oil	R ² = 0.91	[47]
PLS	THz	Detection of moisture content for <i>Ginkgo biloba</i> fruit	R ² = 0.78	[48]
PLS	HSI	Identification of sun-dried and sulphur-fumigated herbals	Sensitivity = 96%	[49]
SVM	NIR	Adulteration of food products (extra-virgin-olive oil, honey, milk, and yogurt)	Accuracy = 0.90–1.00	[50]
SVM	Raman	Adulteration of extra virgin olive oil	R ² = 0.99	[51]
SVM	MIR	Discrimination of wild Paris Polyphylla Smith var. yunnanensis	Accuracy = 87%	[52]
SVM	THz	Use in dried tangerine peels	Accuracy = 94%	[53]
SVM	HSI	Authentication of Theobroma cacao bean hybrids	Prediction error = 3.8–23.1%	[54]
DT	NIR	Poultry quality classification	Precision = 0.74	[55]
DT	Raman	Identification of foodborne pathogenic bacteria	Correct recognition ratio = 0.98	[56]
DT	MIR	Rapid screening of aflatoxin-contaminated peanut oil	Sensitivity = 100%	[57]
DT	THz	Moisture content in fruits and vegetables	Accuracy > 94%	[58]
DT	HSI	Evaluation of subjective tea quality	R ² = 93%	[59]
RF	NIR	Determination of the food dye indigotine in cream	R ² = 0.94; RPD = 4.09	[60]
RF	Raman	Classification of milk products from cows, buffalos, and goats	Accuracy > 94%	[61]
RF	MIR	Identification of the geographical origin of black tea	Accuracy = 100%	[62]
RF	THz	Identification of rice powder mixtures	Accuracy = 98%	[63]
RF	HSI	Quantification of <i>Clostridium sporogenes</i> spores in food products	Accuracy = 80%	[64]
FNN	THz	Prediction of gelatin of various animal origins	Accuracy = 100%	[65]
FNN	HSI	Measurement of firmness and soluble solids content for apples	R ² = 0.76, 0.79	[66]

Table 1. Cont.

Chemometrics	Technique	Applications	Performance (as Reported)	Refs
ANN	NIR	Measurement of carbohydrates and moisture in rice	$R^2 = 0.98, 0.97$	[67]
ANN	Raman	Rapid analysis of sugars in honey	$R^2 > 0.96$	[68]
ANN	THz	Prediction of the freshness of pork	RMSEP = 9.9%	[69]
ANN	HSI	Prediction of moisture content in Lonicerae Japonicae Flos	RPD = 4.42	[70]
CNN	NIR	Determination of the soluble solid content of crown pear	$R^2 = 0.96$	[71]
CNN	SERS	Quantification of thiram and pymetrozine in tea	$R^2 = 0.99, 0.98$	[72]
CNN	MIR	Identification of sugar adulteration in honey	Accuracy = 100%	[73]
CNN	THz	Classification of wheat grain varieties	Accuracy = 98%	[74]
CNN	HSI	Quantitative adulteration in Atlantic salmon	$R^2 = 0.99$	[75]
ELM	NIR	Detection of fennel origin	Accuracy = 100%	[76]
ELM	Raman	Identification of infant rice cereal	Accuracy = 99%	[77]
ELM	THz	Identification of adulterated rice seeds	Accuracy = 100%	[78]
ELM	HSI	Prediction of the cadmium content in rape leaf	$R^2 = 0.98$	[79]

Note: LDA: linear discriminant analysis; QDA: quadratic discriminant analysis; KNN: K-nearest neighbour; SIMCA: soft independent modelling of class analogy; PCR: principal component regression; MLR: multiple linear regression; PLS: partial least squares; SVM: support vector machine; DT: decision Tree; RF: random forest; ELM: extreme learning machine; FNN: feedforward neural network; ANN: artificial neural networks; CNN: convolutional neural network; NIR: near-infrared; MIR: mid-infrared; SERS: surface-enhanced Raman spectroscopy; THz: terahertz; HSI: hyperspectral imaging; R^2 : coefficient of determination; RPD: ratio of performance to deviation; RMSEP: root mean square error of prediction.

Regarding non-linear modelling methods, some studies have used SVM as a classifier, which aids in coping with multi-factorial problems using a mathematical concept called the kernel trick. However, this approach needs careful design consideration with the right number of support vectors and adequate hyperparameters. Also, SVMs are normally single outputs (meaning that they provide binary outputs for classification). Practically, the detailed setup of SVM algorithms is not mentioned very often in the published studies in which several commercial software packages were used directly for data analysis. Besides SVM, ANNs have found rapid development and wide applicability in various fields of science as a non-linear modelling method, thanks to their high predictability and practicability [80]. The ANN is an advanced calibration method, applicable in both classification and regression problems [81]. However, the ANN requirement for a large number of samples to achieve the appropriate modelling of the non-linear data and to reduce the risk of overfitting must be considered by anyone who uses this method, and this could, in theory, limit its application [82]. Another concern is the inherent complexity and lack of “explainability” (i.e., interpretation) of the ANN models, even if there are enough samples supplied to the algorithm. Addressing or mitigating these issues, certain types of ANNs, including FNNs, CNNs, and recurrent neural networks, when co-developed with computer and food scientists, have the potential to address complex food authentication analytical problems. On the other hand, the inevitable requirement of samples to cover the comprehensive variability will lead to the creation of huge databases with expensive proprietary software and hardware. Database management implies several challenges such as complexity, privacy, and storage problems [83]. This provides an additional challenge since important issues concerning computation time and complexity, exceeding memory demands, and low classification performance can be faced due to the increasing number of calibration samples [84]. Inherent to this issue, the use of batch learning methods requiring full recalibration after the addition of each new sample is inappropriate for the

building of these universal chemometric models. Hence, there is a growing demand for accurate open-source chemometric tools to allow for the evolution or incremental learning of existing models and, at the same time, minimise computational and spatial costs [85]. In parallel, privacy and ethical concerns related to data sharing and user rights may hinder the development of large-scale general models when different proprietary databases are distributed across different institutions, private organisations, or countries. In this regard, the rise in federated learning (a decentralised ML approach to model building) as a novel privacy-preserving paradigm brings new opportunities to the sector [86].

To address the increasing complexity of real-world food applications, the building of more robust multi-class chemometric models is necessary [87]. For this purpose, the development of new chemometric methods that will enable the flexible modelling and representation of these multi-class datasets and their synergy with other pattern recognition techniques are two crucial requirements going forward with this challenge. Besides this, these methods should be highly sensitive to the detection of adulterants/contaminants at very low percentages and the simultaneous determination of multiple components in complex mixtures. Both are considered difficult real-world analytical problems.

2.4. Model Validation

Model validation is an important step in the construction of a chemometric model for assessing the accuracy, reliability, robustness, and predictive power of a model for unknown samples, optimising the model parameters, and preventing model overfitting or underfitting [88]. The phenomenon of overfitting is observed when a model has been trained with limited data and is dependent on these data and, therefore, has lost its generalisation ability [89]. Moreover, in underfitting, a model that has been created with too few components/variables or original examples (samples, in this case) cannot cover the variability in real-world data [90]. In addition, it is crucial to ensure that the same samples are not used to build the model and to evaluate the predictive power of the model [14]. Due to the importance of the validation step, two cross-validation schemes (“Leave-One-Out” (LOO) and k-fold cross-validation) are mainly used in the relevant literature [91,92]. LOO cross-validation produces overestimated results and potentially conceals overfitting when small validation datasets are used, resulting in classifying unrepresentative samples as part of the cross-validation set [93]. In addition, most published studies follow a single-laboratory validation strategy instead of an inter-laboratory validation method where the same experimental trial is performed and assessed in different laboratories to assess the transferability of the chemometric models.

Therefore, obtaining significant results requires not only meaningful analysis but also appropriate validation strategies, which are crucial in providing proof that the developed method has realistic and unbiased responses for future unknown samples [94]. Apart from cross-validation and external validation using independent testing samples/data, the application of inter-laboratory validation studies to the developed chemometric models to evaluate their prediction ability in other laboratories is essential. However, this is limited in food science studies due to the necessity of an adequate number of training samples incorporating the variability of multiple laboratories in different instrumental conditions [95]. This also has the potential to increase cross-lab collaboration, exploit federated chemometric models, and make data and model sharing a common practice, a practice that is currently lacking in the field [86].

2.5. Analyst Perspective

The human influence in decision-making is substantial. In theory, non-destructive rapid food analysis provides an unbiased, automatic, and ‘impersonalised’ result, removing the human element due to the high complexity of the data structure and the “black box” nature of the modelling methods. There are examples in industry where such offline instruments (FOSS Milkoscan) are in place. In these cases, the human is the user of the method.

On the contrary, in small-scale studies published in the literature, the analyst is a person with sufficient knowledge of data analysis and who is more involved in the process as both the user and the developer. Results from modelling could be reasonably adjusted or overfitted, and these adjustments may not always be in favour of the generality of the model. In some cases, one can argue that the same dataset can produce either mixed or controversial results depending on the analyst in charge [96]. Therefore, upholding integrity when dealing with spectral data is vital, especially in high-risk applications. Model validation is a critical strategy to mitigate the risk of producing overrated prediction performance. Debates may also arise concerning the size and proprietary nature of the experimental dataset. Therefore, open-source data and modelling methods can implicitly provide another form of validation, as human interference cannot dominate the distortion of the results.

3. How to Address the Challenges

3.1. Sensor/Hardware Development

Despite the achievements in analytical instrumentation over the last few years, the development of new types of cost-effective instrumental techniques combining spectroscopic techniques and other instrumental analytical techniques is a challenging task due to technological limitations (in optics, detector technology, etc.) [97]. In addition, the complexity of the data acquired increases exponentially [98]. The development of new non-destructive sampling procedures that will enable the frequent collection of representative and adequate samples, independently of their physical nature (solid, liquid, or gas), for highly accurate and reliable posterior analysis also presents a great challenge [99].

There is penetration of rapid, easy-to-use spectroscopy-based instruments in industry based on existing benchtop or portable sensor designs. The feed industry has already replaced chemical analysis for basic feed composition with near-infrared spectroscopy and chemometrics, and several analytical targets (total protein, fats, carbohydrates, etc.) have been tested [100]. Moreover, rapid methods developed using vibrational spectroscopy and chemometrics are already used in quality control/quality assurance (QC/QA) in the dairy industry, such as the Milkoscan[®] and FoodScan[®] from FOSS Instruments (Hilleroed, DK). Benchtop devices, such as the DA 7250 NIR analyser from PerkinElmer (Waltham, MA, USA), the TANGO FT-NIR from Bruker (Ettlingen, Germany), and handheld devices, e.g., MicroNIR OnSite-W from VIAVI Solutions (Milpitas, CA, USA), are designed specifically for analysis in the food and agriculture industries. To support these instruments and the analyses conducted, either the instrumental manufacturer or a third party provides calibration models. Besides this, third parties, such as AuNIR Ltd. (Towcester, UK) and Sagitto Ltd. (Cambridge, New Zealand), specialise in chemometric models that support the agri-food industry by developing, maintaining, and updating calibration models that extend the functionality of existing spectral instruments to provide more precise analysis.

3.2. Dedicated Chemometrics Approaches for Further Analysis

As discussed, the current limitations of chemometrics methods are that present algorithms lack the capability to integrate varied datasets, which limits their effectiveness. A potential interim solution is the construction of an ensemble of models each trained on different datasets. This ensemble, functioning related to a decision support system, could collect insights from multiple models to deliver a comprehensive analysis.

Moreover, each spectroscopic instrument provides individual information for different food analyses, making assessing food products challenging, of which multi-sensor systems hold considerable promise [101]. For instance, NIR alone cannot holistically decipher the structural information of food products. Therefore, combining NIR with other techniques, such as MIR or SERS, can lead to an encompassing evaluation of food by integrating multiple information sources. Data fusion from complementary sensors has been employed for authentication and quality assessment, significantly boosting the performance of individual instruments [102,103]. Regarding HSI data, modelling and wavelength selection methods

for spectral data are not entirely effective at solving the issue of 3D data. The spatial information of HSI also needs analysis, leading to the application of data fusion techniques to maximise the use of HSI data [103]. Data fusion achieves convincing results by separately evaluating the spectral and spatial information from HSI data. Kucha et al. [104] applied three data fusion methods to predict the intramuscular Fat content in pork samples, and data fusion resulted in a higher capacity for predicting this parameter.

Recent advances in neural network models have also marked a significant shift in interpreting complex chemical characteristics [105]. These models, particularly optimised for real-world applications in food analysis, reveal the potential field of artificial general intelligence (AGI) in food products. Despite their resource-intensive nature, these models promise a systematic approach to data analysis, pending the availability of sufficient computing power and annotated data to support their real-world application. Additionally, the suggestion to apply fuzzy-related and Bayesian-based models could potentially mitigate the opacity of NN and SVM, providing an interpretable analysis. A fuzzy system is a method for modelling and processing imprecise and vague information, which is characterised by membership functions that assign degrees of belonging to each variable [106]. Moreover, explainable and trustworthy AI are two recent and emerging topics that aim to address these challenges with the development of new algorithms. These proposed approaches would respond to several key issues and facilitate incremental and federated learning—concepts not exclusively pertinent to the food sector but universally recognised as challenges in diverse fields.

A critical observation in this paper is the necessity to design the chemometric pipeline prior to delving into modelling methods. Current studies often disproportionately emphasise sophisticated classification or regression techniques such as optimised decision tree models, neglecting the importance of a systematic chemometric process with simpler approaches. This process should span from preprocessing to the application of modelling methods. Regarding preprocessing, the integration of additional tools and methods is recommended. While PCA is a standard approach, it is not without its limitations. Moreover, reimagining the outputs of instruments like FTIR as time-series datasets opens possibilities for the application of decomposition techniques (without the Fourier transformation algorithm). These strategies might hold the key to refining the approach, underscoring the potential that lies in reevaluating and enhancing preprocessing stages before exploring new modelling schemes. To describe future rapid spectroscopic analysis, one could cite a message from the NIR 2023 conference: “There will be no causal interpretation without prediction validation and no prediction without attempted causal interpretation”.

The concept of multi-class and multi-task problems, involving more than two classes or objectives, further highlights the potential of using multiple models in parallel. By combining the outcomes of these models, an ensemble approach can be realised. This methodology aligns with the foundational principles of ensemble methods, underscoring their relevance and applicability in addressing complex challenges in chemometric analysis [107]. The multidisciplinary approach, encompassing algorithm development, preprocessing enhancement, and ensemble methods, presents a comprehensive strategy for advancing chemometric methodologies in the food industry and beyond [107].

3.3. Artificial Intelligence-Driven Methodology for Rapid Food Analysis

Several review papers have highlighted the importance of the sub-field of AI for rapid food analysis, including machine learning and deep learning, and focused on the implementation of ML/AI in a practical way [108]. The application of data-driven models, particularly different types of neural network models, in food manufacturing underscores a growing reliance on ML/AI to elucidate complex input–output relationships [109]. This data-centric approach could be beneficial for characterising systems where mechanistic knowledge is limited, thereby enhancing the predictive accuracy of quality and safety parameters for food products. However, the methodology of applying AI in future analytical studies lacks overviewing. On a theoretical level, the role of AI extends beyond modelling

and simulation, offering benefits in the predictive maintenance of key factory equipment, the design of novel, cost-effective spectroscopy instruments, factory operational efficiency, and risk management [12]. Future studies under discussion should highlight the critical success factors for AI adoption in the food supply chain, including technology readiness, security, privacy, customer satisfaction, regulatory compliance, and the importance of information sharing among partners [110].

To meet real-world demand, various AI-driven approaches facilitate a collaborative environment where food scientists and computational experts can jointly tackle manufacturing challenges, directing the way for sophisticated multi-scale models that surpass traditional methods. This proactive stance is further exemplified in the development of leading indicators for food safety, where AI algorithms analyse behavioural data to anticipate potential safety issues, thus shifting the focus from reactive to preventive measures [111].

3.4. Model Validation

In the realm of chemometrics, the importance of robust data validation methods like k-fold cross-validation cannot be overstated. However, the effectiveness of these methods hinges on the availability of sufficient and high-quality data, as well as its ground truthing to facilitate training and evaluation of the chemometric models. A key aspect of ensuring data quality involves the detection and removal of outliers. Additionally, the creation of “synthetic” datasets has emerged as a viable solution to enhance the performance of classifiers, especially when dealing with limited or skewed data [112]. Besides removing observed outliers, data augmentation is a disruptive method used to enhance model robustness for developing and validating rapid food analysis methods, involving artificially enlarging the training dataset by creating modified versions of the existing datasets [113]. Data augmentation fundamentally enhances model validation: a more diversified dataset is used for training; the occurrence of overfitting is minimised; the training process becomes more efficient by gathering more information at each epoch. As a result, there is the enhancement of the model’s generalisability with exposure to increasing numbers of datasets [114]. In particular, data augmentation is carried out by integrating noise, such as Gaussian noise, or in combination with ensemble methods [112]. Additionally, there are more innovative techniques employing deep learning for data augmentation, including generative adversarial networks (GANs) and semi-supervised generative adversarial networks to create new data points [114]. In practice, data augmentation presents itself as an attractive low-cost strategy for robust temperature compensation in NIR calibration and prediction, challenging traditional methods that integrate non-relevant spectral variations by measuring samples under different conditions [115]. Two types of approaches have been proposed: augmenting the calibration matrix with simulated noise and a correction method to eliminate non-relevant variations from new spectra. Continuous development of data augmentation is a related framework for chemometric analysis to enhance the classification results; synthetic spectroscopic data from vegetable oil samples were used to assist the model’s prediction performance without the need for extensive ‘real’ sample datasets [112]. This study highlighted its potential in addressing challenges related to sample variability and instrument differences in spectroscopic analysis. Regarding the deep learning approach, a GAN was used for predicting the oil content of a single maize kernel by generating artificial spectra after many iterations, resulting in very promising results [116].

Overall, the integration of such methodologies into a standardised workflow is crucial for maintaining data integrity in chemometric analysis. This standardisation arguably minimises manual interference and prevents the generation of overly optimistic results stemming from unfit analytical strategies. From the perspective of utilising spectroscopy instruments in the industrial production line, the addition of historical data to the regular validation process could provide a forecast on operations and the decision-making process in a timely manner. The integration of blockchain technology for data sharing in validating data analysis steps represents innovative approaches to ensuring data transparency and

reliability [9]. Moreover, the establishment of open repositories for spectroscopic data can democratise access to valuable datasets, further boosting research outputs.

4. Conclusions

Nowadays, there is an increasing need for rapid, easy, non-destructive, and low-cost analytical solutions with minimal sample preparation and accurate results in the food field to control the quality and safety of raw materials and intermediate and final products and identify any issues such as food adulteration (Figure 3). Vibrational spectroscopic methods are untargeted analytical techniques capable of meeting these demands, at the cost of producing complex multidimensional data, containing plenty of useful information for analysis but whose interpretation is commonly a difficult, laborious, and time-consuming task. Chemometrics offers a wide spectrum of data analysis methods that can enable the deep and extensive analysis of the acquired spectra, as well as its automation, by allowing their elucidation, analysis, and further modelling thanks to its interdisciplinary nature. However, there are some practical issues related to the use of vibrational spectroscopy associated with chemometrics that limit its online industrial applications. Generally, chemometric methods demand a spectral library of high quality and representative spectra to build accurate, sensitive, robust, generalised chemometric models. Concerning sample representativity, vibrational spectroscopy requires sample homogeneity over the sampling area to produce a representative spectrum of an examined sample. This fact limits its application to heterogeneous and complex food samples where better sampling techniques are needed. Moreover, conducting the analysis in a laboratory is different than in a real-world industrial environment. A laboratory has well-controlled conditions for sample handling, preparation, and subsequent analysis, with little variation affecting data acquisition, whereas the same analysis in the field will create more noise, i.e., more variation that the models have not been exposed to during development [117]. This means that spectroscopic acquisition can be affected due to variations in experimental or sample conditions. For this reason, it would be preferable if the developed model incorporates all the possible sources of sample variation, such as natural and experimental variation. Nonetheless, the inclusion of all relevant sources of sample variation seems impractical due to limited resources since it requires time-consuming and costly processes as well as exceeding computer processing and memory demands or fulfilling the required privacy and legal constraints. Moreover, apart from the practical limitations, the complexity of the developed chemometric model may be increased rapidly, and low classification performance may be faced due to the increasing number of training samples and model classes.

Except for the issues related to the production of realistic solutions for food analysis, current classification and regression techniques meet challenges when they are used for the detection of adulterants at low concentration levels and the characterisation of complex food blends, where many compounds must be identified and quantified [118]. This can be explained by the difficulty associated with these analytical problems due to the subtle spectral differences between authentic and adulterated mixtures, as well as the overlapping peaks of compounds in mixtures. Thus, even though chemometrics offers automatic results, current methods are limited for low-level adulterations and multi-component analysis and current models are not general enough and often overfitted when and if extended to real-world applications.

On the other hand, there is no significant “year-over-year” or even “decade-over-decade” scientific progress in the efficiency of these algorithmic methods to tackle even more complex food analysis problems. If one considers the example of the speech recognition and natural language processing research area, the rise in powerful novel ML algorithms such as deep learning systems and follow-up concepts such as reinforcement learning, federated learning, trustworthy AI, etc., have “transformed the scene” leading to digital assistants on our phones (Siri and Google Assistant), AI dominance in the game Go (DeepMind) [119], and, of course, the recent introduction of large language models trained on generative pre-trained transformers (OpenAI and ChatGPT). It remains to be

seen, however, as to how these hyper-advanced methods will be translated to the field of chemometric-based rapid food analysis. At the top, there is a limited amount of research, produced in academic studies, that finds its way into industry and the official methods of regulatory bodies, which highlights a gap in technology transfer as well as the applicability of the results.

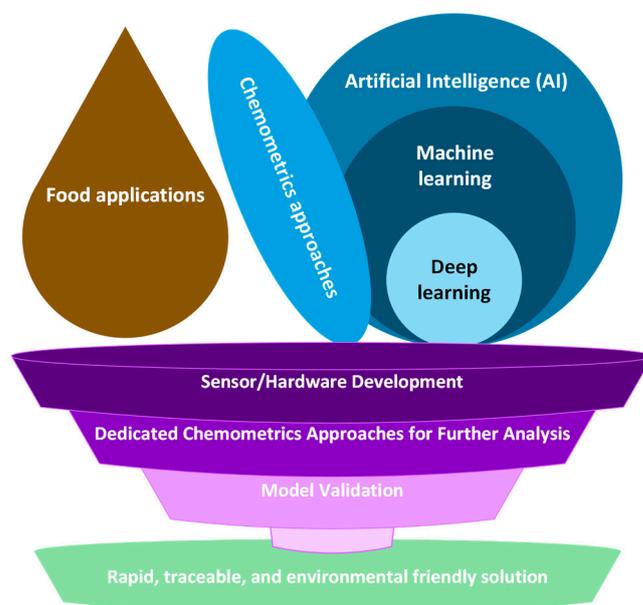


Figure 3. Overview of AI-driven non-destructive spectroscopic tools for rapid food analysis.

Funding: This research was supported by funding from the Department of Learning and Employment Northern Ireland (DELNI) (Ph.D. studentship block grant). Work at LLNL was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract DEAC52-07NA27344. LLNL-JRNL-861315.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

References

- Schulze, B.; Jeon, Y.; Kaserzon, S.; Heffernan, A.L.; Dewapriya, P.; O'Brien, J.; Gomez Ramos, M.J.; Ghorbani Gorji, S.; Mueller, J.F.; Thomas, K.V.; et al. An Assessment of Quality Assurance/Quality Control Efforts in High Resolution Mass Spectrometry Non-Target Workflows for Analysis of Environmental Samples. *TrAC Trends Anal. Chem.* **2020**, *133*, 116063. [[CrossRef](#)]
- Kutsanedzie, F.Y.H.H.; Guo, Z.; Chen, Q. Advances in Nondestructive Methods for Meat Quality and Safety Monitoring. *Food Rev. Int.* **2019**, *35*, 536–562. [[CrossRef](#)]
- McGrath, T.F.; Haughey, S.A.; Patterson, J.; Fauhl-Hassek, C.; Donarski, J.; Alewijn, M.; van Ruth, S.; Elliott, C.T. What Are the Scientific Challenges in Moving from Targeted to Non-Targeted Methods for Food Fraud Testing and How Can They Be Addressed?—Spectroscopy Case Study. *Trends Food Sci. Technol.* **2018**, *76*, 38–55. [[CrossRef](#)]
- Sørensen, K.M.; van den Berg, F.; Engelsen, S.B. *NIR Data Exploration and Regression by Chemometrics—A Primer*; Springer: Singapore, 2021; ISBN 9789811586484.
- Frank, I.E.; Kowalski, B.R. Chemometrics. *Anal. Chem.* **1982**, *54*, 232–243. [[CrossRef](#)]
- Akulich, F.; Anahideh, H.; Sheyyab, M.; Ambre, D. Explainable Predictive Modeling for Limited Spectral Data. *Chemom. Intell. Lab. Syst.* **2022**, *225*, 104572. [[CrossRef](#)]
- Zhou, L.; Zhang, C.; Liu, F.; Qiu, Z.; He, Y. Application of Deep Learning in Food: A Review. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18*, 1793–1811. [[CrossRef](#)] [[PubMed](#)]
- Reinholds, I.; Bartkevics, V.; Silvis, I.C.J.; van Ruth, S.M.; Esslinger, S. Analytical Techniques Combined with Chemometrics for Authentication and Determination of Contaminants in Condiments: A Review. *J. Food Compos. Anal.* **2015**, *44*, 56–72. [[CrossRef](#)]

9. Jia, W.; van Ruth, S.; Scollan, N.; Koidis, A. Hyperspectral Imaging (HSI) for Meat Quality Evaluation across the Supply Chain: Current and Future Trends. *Curr. Res. Food Sci.* **2022**, *5*, 1017–1027. [[CrossRef](#)] [[PubMed](#)]
10. Szymańska, E.; Gerretzen, J.; Engel, J.; Geurts, B.; Blanchet, L.; Buydens, L.M.C. Chemometrics and Qualitative Analysis Have a Vibrant Relationship. *TrAC Trends Anal. Chem.* **2015**, *69*, 34–51. [[CrossRef](#)]
11. Rothe, S.; Kudsus, B.; Söfker, D. Does Classifier Fusion Improve the Overall Performance? Numerical Analysis of Data and Fusion Method Characteristics Influencing Classifier Fusion Performance. *Entropy* **2019**, *21*, 866. [[CrossRef](#)]
12. Xu, Y.; Liu, X.; Cao, X.; Huang, C.; Liu, E.; Qian, S.; Liu, X.; Wu, Y.; Dong, F.; Qiu, C.W.; et al. Artificial Intelligence: A Powerful Paradigm for Scientific Research. *Innovation* **2021**, *2*, 100179. [[CrossRef](#)] [[PubMed](#)]
13. Danezis, G.P.; Tsagkaris, A.S.; Camin, F.; Brusica, V.; Georgiou, C.A. Food Authentication: Techniques, Trends & Emerging Approaches. *TrAC Trends Anal. Chem.* **2016**, *85*, 123–132. [[CrossRef](#)]
14. Ropodi, A.I.; Panagou, E.Z.; Nychas, G.J.E. Data Mining Derived from Food Analyses Using Non-Invasive/Non-Destructive Analytical Techniques; Determination of Food Authenticity, Quality & Safety in Tandem with Computer Science Disciplines. *Trends Food Sci. Technol.* **2016**, *50*, 11–25. [[CrossRef](#)]
15. Gallo, M.; Ferranti, P. The Evolution of Analytical Chemistry Methods in Foodomics. *J. Chromatogr. A* **2016**, *1428*, 3–15. [[CrossRef](#)]
16. Aparicio, R.; Aparicio-Ruiz, R. Authentication of Vegetable Oils by Chromatographic Techniques. *J. Chromatogr. A* **2000**, *881*, 93–104. [[CrossRef](#)] [[PubMed](#)]
17. Capuano, E.; Boerrigter-Eenling, R.; van der Veer, G.; van Ruth, S.M. Analytical Authentication of Organic Products: An Overview of Markers. *J. Sci. Food Agric.* **2013**, *93*, 12–28. [[CrossRef](#)]
18. Do, T.K.T.; Hadji-Minaglou, F.; Antoniotti, S.; Fernandez, X. Authenticity of Essential Oils. *TrAC Trends Anal. Chem.* **2015**, *66*, 146–157. [[CrossRef](#)]
19. Cozzolino, D. Recent Trends on the Use of Infrared Spectroscopy to Trace and Authenticate Natural and Agricultural Food Products. *Appl. Spectrosc. Rev.* **2012**, *47*, 518–530. [[CrossRef](#)]
20. Wójcicki, K.; Khmelinskii, I.; Sikorski, M.; Sikorska, E. Near and Mid Infrared Spectroscopy and Multivariate Data Analysis in Studies of Oxidation of Edible Oils. *Food Chem.* **2015**, *187*, 416–423. [[CrossRef](#)]
21. Black, C.; Chevallier, O.P.; Elliott, C.T. The Current and Potential Applications of Ambient Mass Spectrometry in Detecting Food Fraud. *TrAC Trends Anal. Chem.* **2016**, *82*, 268–278. [[CrossRef](#)]
22. Bassbasi, M.; De Luca, M.; Ioele, G.; Oussama, A.; Ragno, G. Prediction of the Geographical Origin of Butters by Partial Least Square Discriminant Analysis (PLS-DA) Applied to Infrared Spectroscopy (FTIR) Data. *J. Food Compos. Anal.* **2014**, *33*, 210–215. [[CrossRef](#)]
23. Danezis, G.P.; Tsagkaris, A.S.; Brusica, V.; Georgiou, C.A. Food Authentication: State of the Art and Prospects. *Curr. Opin. Food Sci.* **2016**, *10*, 22–31. [[CrossRef](#)]
24. Washburn, K.E.; Stormo, S.K.; Skjelvareid, M.H.; Heia, K. Non-Invasive Assessment of Packaged Cod Freeze-Thaw History by Hyperspectral Imaging. *J. Food Eng.* **2017**, *205*, 64–73. [[CrossRef](#)]
25. Han, Q.; Peller, J.; Erasmus, S.W.; Elliott, C.T.; van Ruth, S.M. Interpreting the Variation in Particle Size of Ground Spice by High-Resolution Visual and Spectral Imaging: A Ginger Case Study. *Food Res. Int.* **2023**, *170*, 113023. [[CrossRef](#)] [[PubMed](#)]
26. Sun, X.; Zhang, M.; Wang, P.; Chen, J.; Yang, S.; Luo, P.; Gao, X. Detection and Quantitation of Adulterated Paprika Samples Using Second-Order HPLC-FLD Fingerprints and Chemometrics. *Foods* **2022**, *11*, 2376. [[CrossRef](#)] [[PubMed](#)]
27. Musio, B.; Todisco, S.; Antonicelli, M.; Garino, C.; Arlorio, M.; Mastroilli, P.; Latronico, M.; Gallo, V. Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production. *Appl. Sci.* **2022**, *12*, 2583. [[CrossRef](#)]
28. Kalogiouri, N.P.; Manousi, N.; Klaoudatos, D.; Spanos, T.; Topi, V.; Zachariadis, G.A. Rare Earths as Authenticity Markers for the Discrimination of Greek and Turkish Pistachios Using Elemental Metabolomics and Chemometrics. *Foods* **2021**, *10*, 349. [[CrossRef](#)]
29. Rivera-Pérez, A.; Romero-González, R.; Garrido Frenich, A. Fingerprinting Based on Gas Chromatography-Orbitrap High-Resolution Mass Spectrometry and Chemometrics to Reveal Geographical Origin, Processing, and Volatile Markers for Thyme Authentication. *Food Chem.* **2022**, *393*, 133377. [[CrossRef](#)]
30. Hai, C.; Chen, H.; Suo, Y.; Guan, Y.; Wang, S.; Lan, W.; Long, W.; Yang, X.; Yang, J.; Fu, H. Geographical Origin and Species Identification of Lili Bulbus Using C/N/H/O Stable Isotopes and Multi-Elemental Combined Chemometrics. *J. Food Compos. Anal.* **2023**, *116*, 105062. [[CrossRef](#)]
31. Nunes, C.A.; Alvarenga, V.O.; de Souza Sant’Ana, A.; Santos, J.S.; Granato, D. The Use of Statistical Software in Food Science and Technology: Advantages, Limitations and Misuses. *Food Res. Int.* **2015**, *75*, 270–280. [[CrossRef](#)]
32. Abdel-Rahman, A.I.; Lim, G.J. A Nonlinear Partial Least Squares Algorithm Using Quadratic Fuzzy Inference System. *J. Chemom.* **2009**, *23*, 530–537. [[CrossRef](#)]
33. Reid, L.M.; O’Donnell, C.P.; Downey, G. Recent Technological Advances for the Determination of Food Authenticity. *Trends Food Sci. Technol.* **2006**, *17*, 344–353. [[CrossRef](#)]
34. Li, Z.; Wang, P.P.; Huang, C.C.; Shang, H.; Pan, S.Y.; Li, X.J. Application of Vis/NIR Spectroscopy for Chinese Liquor Discrimination. *Food Anal. Methods* **2014**, *7*, 1337–1344. [[CrossRef](#)]
35. Li, S.; Shan, Y.; Zhu, X.; Zhang, X.; Ling, G. Detection of Honey Adulteration by High Fructose Corn Syrup and Maltose Syrup Using Raman Spectroscopy. *J. Food Compos. Anal.* **2012**, *28*, 69–74. [[CrossRef](#)]

36. Velásquez, C.; Prieto, F.; Palou, L.; Cubero, S.; Blasco, J.; Aleixos, N. New Model for the Automatic Detection of Anthracnose in Mango Fruits Based on Vis/NIR Hyperspectral Imaging and Discriminant Analysis. *J. Food Meas. Charact.* **2023**, *18*, 560–570. [[CrossRef](#)]
37. Kabir, M.H.; Guindo, M.L.; Chen, R.; Liu, F. Geographic Origin Discrimination of Millet Using Vis-NIR Spectroscopy Combined with Machine Learning Techniques. *Foods* **2021**, *10*, 2767. [[CrossRef](#)] [[PubMed](#)]
38. Georgouli, K.; Martinez Del Rincon, J.; Koidis, A. Continuous Statistical Modelling for Rapid Detection of Adulteration of Extra Virgin Olive Oil Using Mid Infrared and Raman Spectroscopic Data. *Food Chem.* **2017**, *217*, 735–742. [[CrossRef](#)] [[PubMed](#)]
39. Bizzani, M.; William Menezes Flores, D.; Alberto Colnago, L.; David Ferreira, M. Monitoring of Soluble Pectin Content in Orange Juice by Means of MIR and TD-NMR Spectroscopy Combined with Machine Learning. *Food Chem.* **2020**, *332*, 127383. [[CrossRef](#)]
40. Daszykowski, M.; Orzel, J.; Wrobel, M.S.; Czarnik-Matusewicz, H.; Walczak, B. Improvement of Classification Using Robust Soft Classification Rules for Near-Infrared Reflectance Spectral Data. *Chemom. Intell. Lab. Syst.* **2011**, *109*, 86–93. [[CrossRef](#)]
41. Karunathilaka, S.R.; Farris, S.; Mossoba, M.M.; Moore, J.C.; Yakes, B.J. Non-Targeted Detection of Milk Powder Adulteration Using Raman Spectroscopy and Chemometrics: Melamine Case Study. *Food Addit. Contam. Part A* **2017**, *34*, 170–182. [[CrossRef](#)]
42. Oroian, M.; Ropciuc, S.; Paduret, S. Honey Adulteration Detection Using Raman Spectroscopy. *Food Anal. Methods* **2018**, *11*, 959–968. [[CrossRef](#)]
43. Li, X.; He, Y.; Wu, C.; Sun, D.W. Nondestructive Measurement and Fingerprint Analysis of Soluble Solid Content of Tea Soft Drink Based on Vis/NIR Spectroscopy. *J. Food Eng.* **2007**, *82*, 316–323. [[CrossRef](#)]
44. Cui, H.; Wang, S.; Yang, X.; Zhang, W.; Chen, M.; Wu, Y.; Li, S.; Li, L.; Cai, D.; Guo, B.; et al. Predictive Models for Assessing the Risk of Fusarium Pseudograminearum Mycotoxin Contamination in Post-Harvest Wheat with Multi-Parameter Integrated Sensors. *Food Chem. X* **2022**, *16*, 100472. [[CrossRef](#)]
45. Dachoupakan Sirisomboon, C.; Putthang, R.; Sirisomboon, P. Application of near Infrared Spectroscopy to Detect Aflatoxigenic Fungal Contamination in Rice. *Food Control* **2013**, *33*, 207–214. [[CrossRef](#)]
46. Kashif, M.; Majeed, M.I.; Nawaz, H.; Rashid, N.; Abubakar, M.; Ahmad, S.; Ali, S.; Hyat, H.; Bashir, S.; Batool, F.; et al. Surface-Enhanced Raman Spectroscopy for Identification of Food Processing Bacteria. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *261*, 119989. [[CrossRef](#)] [[PubMed](#)]
47. Okere, E.E.; Arendse, E.; Nieuwoudt, H.; Perold, W.J.; Opara, U.L. Non-Destructive Evaluation of the Quality Characteristics of Pomegranate Kernel Oil by Fourier Transform Near-Infrared and Mid-Infrared Spectroscopy. *Front. Plant Sci.* **2022**, *13*, 867555. [[CrossRef](#)]
48. Gong, Z.; Deng, D.; Sun, X.; Liu, J.; Ouyang, Y. Non-Destructive Detection of Moisture Content for Ginkgo Biloba Fruit with Terahertz Spectrum and Image: A Preliminary Study. *Infrared Phys. Technol.* **2022**, *120*, 103997. [[CrossRef](#)]
49. Zhang, H.; Wu, T.; Zhang, L.; Zhang, P. Development of a Portable Field Imaging Spectrometer: Application for the Identification of Sun-Dried and Sulfur-Fumigated Chinese Herbals. *Appl. Spectrosc.* **2016**, *70*, 879–887. [[CrossRef](#)]
50. Folli, G.S.; Santos, L.P.; Santos, F.D.; Cunha, P.H.P.; Schaffel, I.F.; Borghi, F.T.; Barros, I.H.A.S.; Pires, A.A.; Ribeiro, A.V.F.N.; Romão, W.; et al. Food Analysis by Portable NIR Spectrometer. *Food Chem. Adv.* **2022**, *1*, 100074. [[CrossRef](#)]
51. Dong, W.; Zhang, Y.; Zhang, B.; Wang, X. Quantitative Analysis of Adulteration of Extra Virgin Olive Oil Using Raman Spectroscopy Improved by Bayesian Framework Least Squares Support Vector Machines. *Anal. Methods* **2012**, *4*, 2772–2777. [[CrossRef](#)]
52. Wu, X.M.; Zuo, Z.T.; Zhang, Q.Z.; Wang, Y.Z. FT-MIR and UV-Vis Data Fusion Strategy for Origins Discrimination of Wild Paris Polyphylla Smith Var. Yunnanensis. *Vib. Spectrosc.* **2018**, *96*, 125–136. [[CrossRef](#)]
53. Yang, S.Z.; Li, C.; Li, C.; Wang, Z.Q.; Huang, L.L.; Ding, Q. The Differences in the Dried Tangerine Peels Stored for Different Years Revealed by Terahertz Spectroscopy and Imaging. *Mod. Food Sci. Technol.* **2020**, *35*, 258–266. [[CrossRef](#)]
54. Cruz-Tirado, J.P.; Fernández Pierna, J.A.; Rogez, H.; Barbin, D.F.; Baeten, V. Authentication of Cocoa (*Theobroma cacao*) Bean Hybrids by NIR-Hyperspectral Imaging and Chemometrics. *Food Control* **2020**, *118*, 107445. [[CrossRef](#)]
55. Barbon, S.; da Costa Barbon, A.P.A.; Mantovani, R.G.; Barbin, D.F. Machine Learning Applied to Near-Infrared Spectra for Chicken Meat Classification. *J. Spectrosc.* **2018**, *2018*, 8949741. [[CrossRef](#)]
56. Zhang, J.; Gao, P.; Wu, Y.; Yan, X.; Ye, C.; Liang, W.; Yan, M.; Xu, X.; Jiang, H. Identification of Foodborne Pathogenic Bacteria Using Confocal Raman Microspectroscopy and Chemometrics. *Front. Microbiol.* **2022**, *13*, 874658. [[CrossRef](#)]
57. Yang, Y.; Zhang, Y.; He, C.; Xie, M.; Luo, H.; Wang, Y.; Zhang, J. Rapid Screen of Aflatoxin-Contaminated Peanut Oil Using Fourier Transform Infrared Spectroscopy Combined with Multivariate Decision Tree. *Int. J. Food Sci. Technol.* **2018**, *53*, 2386–2393. [[CrossRef](#)]
58. Ren, A.; Zahid, A.; Zoha, A.; Shah, S.A.; Imran, M.A.; Alomainy, A.; Abbasi, Q.H. Machine Learning Driven Approach towards the Quality Assessment of Fresh Fruits Using Non-Invasive Sensing. *IEEE Sens. J.* **2020**, *20*, 2075–2083. [[CrossRef](#)]
59. Ren, G.; Wang, Y.; Ning, J.; Zhang, Z. Using Near-Infrared Hyperspectral Imaging with Multiple Decision Tree Methods to Delineate Black Tea Quality. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *237*, 118407. [[CrossRef](#)] [[PubMed](#)]
60. Zhang, S.; Tan, Z.; Liu, J.; Xu, Z.; Du, Z. Determination of the Food Dye Indigotine in Cream by Near-Infrared Spectroscopy Technology Combined with Random Forest Model. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *227*, 117551. [[CrossRef](#)]
61. Amjad, A.; Ullah, R.; Khan, S.; Bilal, M.; Khan, A. Raman Spectroscopy Based Analysis of Milk Using Random Forest Classification. *Vib. Spectrosc.* **2018**, *99*, 124–129. [[CrossRef](#)]

62. Li, Y.; Logan, N.; Quinn, B.; Hong, Y.; Birse, N.; Zhu, H.; Haughey, S.; Elliott, C.T.; Wu, D. Fingerprinting Black Tea: When Spectroscopy Meets Machine Learning a Novel Workflow for Geographical Origin Identification. *Food Chem.* **2024**, *438*, 138029. [CrossRef]
63. Friska, J.; Navaneetha Velammal, M.; Rajeshwari, A.; Hannah Blessy, P. Random Forest (RF) Based Identification of Rice Powder Mixture Using Terahertz Spectroscopy. *J. Phys. Conf. Ser.* **2021**, *1979*, 012056. [CrossRef]
64. Soni, A.; Al-Sarayreh, M.; Reis, M.M.; Brightwell, G. Hyperspectral Imaging and Deep Learning for Quantification of Clostridium Sporogenes Spores in Food Products Using 1D- Convolutional Neural Networks and Random Forest Model. *Food Res. Int.* **2021**, *147*, 110577. [CrossRef] [PubMed]
65. Li, Y.; Liu, L.; Wang, Z.; Chang, T.; Li, K.; Xu, W.; Wu, Y.; Yang, H.; Jiang, D. To Estimate Performance of Artificial Neural Network Model Based on Terahertz Spectrum: Gelatin Identification as an Example. *Front. Nutr.* **2022**, *9*, 925717. [CrossRef] [PubMed]
66. Lu, R. Nondestructive Measurement of Firmness and Soluble Solids Content for Apple Fruit Using Hyperspectral Scattering Images. *Sens. Instrum. Food Qual. Saf.* **2007**, *1*, 19–27. [CrossRef]
67. Son, S.; Kim, D.; Choul Choi, M.; Lee, J.; Kim, B.; Min Choi, C.; Kim, S. Weight Interpretation of Artificial Neural Network Model for Analysis of Rice (*Oryza sativa* L.) with near-Infrared Spectroscopy. *Food Chem. X* **2022**, *15*, 100430. [CrossRef] [PubMed]
68. Özbalci, B.; Boyaci, I.H.; Topcu, A.; Kadilar, C.; Tamer, U. Rapid Analysis of Sugars in Honey by Processing Raman Spectrum Using Chemometric Methods and Artificial Neural Networks. *Food Chem.* **2013**, *136*, 1444–1452. [CrossRef]
69. Qi, L.; Zhao, M.; Zhao, J.; Tang, Y. Preliminary Investigation of Terahertz Spectroscopy to Predict Pork Freshness Non-Destructively. *Food Sci. Technol.* **2019**, *39*, 563–570. [CrossRef]
70. Cui, P.; Yu, Y.; Zhao, J.; Miao, P.; Xue, Q.; Liu, C.; Li, Z. Hyperspectral Imaging Combined with Artificial Intelligence Techniques to Explore the Drying Behavior of Natural Lonicerae Japonicae Flos Extracts. *Measurement* **2023**, *218*, 113246. [CrossRef]
71. Qi, H.; Shen, C.; Chen, G.; Zhang, J.; Chen, F.; Li, H.; Zhang, C. Rapid and Non-Destructive Determination of Soluble Solid Content of Crown Pear by Visible/near-Infrared Spectroscopy with Deep Learning Regression. *J. Food Compos. Anal.* **2023**, *123*, 105585. [CrossRef]
72. Li, H.; Luo, X.; Haruna, S.A.; Zareef, M.; Chen, Q.; Ding, Z.; Yan, Y. Au-Ag OHCs-Based SERS Sensor Coupled with Deep Learning CNN Algorithm to Quantify Thiram and Pymetrozine in Tea. *Food Chem.* **2023**, *428*, 136798. [CrossRef]
73. Li, Q.; Zeng, J.; Lin, L.; Zhang, J.; Zhu, J.; Yao, L.; Wang, S.; Du, J.; Wu, Z. Mid-Infrared Spectra Feature Extraction and Visualization by Convolutional Neural Network for Sugar Adulteration Identification of Honey and Real-World Application. *LWT* **2021**, *140*, 110856. [CrossRef]
74. Chen, F.; Shen, Y.; Li, G.; Ai, M.; Wang, L.; Xiong, X.; Tao, D. Classification of Wheat Grain Varieties Using Terahertz Spectroscopy and Convolutional Neural Network. *J. Food Compos. Anal.* **2024**, *19*, 106060. [CrossRef]
75. Li, P.; Tang, S.; Chen, S.; Tian, X.; Zhong, N. Hyperspectral Imaging Combined with Convolutional Neural Network for Accurately Detecting Adulteration in Atlantic Salmon. *Food Control* **2023**, *147*, 109573. [CrossRef]
76. Zuo, E.; Sun, L.; Yan, J.; Chen, C.; Chen, C.; Lv, X. Rapidly Detecting Fennel Origin of the Near-Infrared Spectroscopy Based on Extreme Learning Machine. *Sci. Rep.* **2022**, *12*, 13593. [CrossRef]
77. Zhao, Y.-J.; Zhang, Z.; Zhang, Y.; Ni, X.-F.; Wang, H.-Y. Identification of Infant Rice Cereal Based on Raman Spectroscopy Combined with an Extreme Learning Machine Algorithm. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4539298 (accessed on 1 February 2024).
78. Hou, X.; Jie, Z.; Wang, J.; Liu, X.; Ye, N. Application of Terahertz Spectroscopy Combined with Feature Improvement Algorithm for the Identification of Adulterated Rice Seeds. *Infrared Phys. Technol.* **2023**, *131*, 104694. [CrossRef]
79. Cheng, J.; Sun, J.; Yao, K.; Xu, M.; Wang, S.; Fu, L. Development of Multi-Disturbance Bagging Extreme Learning Machine Method for Cadmium Content Prediction of Rape Leaf Using Hyperspectral Imaging Technology. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *279*, 121479. [CrossRef]
80. Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto, O. Data Fusion Methodologies for Food and Beverage Authentication and Quality Assessment—A Review. *Anal. Chim. Acta* **2015**, *891*, 1–14. [CrossRef] [PubMed]
81. Marini, F. Artificial Neural Networks in Foodstuff Analyses: Trends and Perspectives A Review. *Anal. Chim. Acta* **2009**, *635*, 121–131. [CrossRef] [PubMed]
82. Ni, W.; Nørgaard, L.; Mørup, M. Non-Linear Calibration Models for near Infrared Spectroscopy. *Anal. Chim. Acta* **2014**, *813*, 1–14. [CrossRef] [PubMed]
83. Rokach, L. Ensemble-Based Classifiers. *Artif. Intell. Rev.* **2010**, *33*, 1–39. [CrossRef]
84. Sáiz-Abajo, M.J.; Mevik, B.H.; Segtnan, V.H.; Næs, T. Ensemble Methods and Data Augmentation by Noise Addition Applied to the Analysis of Spectroscopic Data. *Anal. Chim. Acta* **2005**, *533*, 147–159. [CrossRef]
85. Diaz-Chito, K.; Georgouli, K.; Koidis, A.; Martinez del Rincon, J. Incremental Model Learning for Spectroscopy-Based Food Analysis. *Chemom. Intell. Lab. Syst.* **2017**, *167*, 123–131. [CrossRef]
86. Gosselin, R.; Vieu, L.; Loukil, F.; Benoit, A. Privacy and Security in Federated Learning: A Survey. *Appl. Sci.* **2022**, *12*, 9901. [CrossRef]
87. Khashei, M.; Zeinal Hamadani, A.; Bijari, M. A Novel Hybrid Classification Model of Artificial Neural Networks and Multiple Linear Regression Models. *Expert Syst. Appl.* **2012**, *39*, 2606–2620. [CrossRef]
88. López, M.I.; Callao, M.P.; Ruisánchez, I. A Tutorial on the Validation of Qualitative Methods: From the Univariate to the Multivariate Approach. *Anal. Chim. Acta* **2015**, *891*, 62–72. [CrossRef] [PubMed]

89. Berrueta, L.A.; Alonso-Salces, R.M.; Héberger, K. Supervised Pattern Recognition in Food Analysis. *J. Chromatogr. A* **2007**, *1158*, 196–214. [[CrossRef](#)] [[PubMed](#)]
90. Kumar, N.; Bansal, A.; Sarma, G.S.; Rawal, R.K. Chemometrics Tools Used in Analytical Chemistry: An Overview. *Talanta* **2014**, *123*, 186–199. [[CrossRef](#)]
91. Gouvinhas, I.; De Almeida, J.M.M.M.; Carvalho, T.; Machado, N.; Barros, A.I.R.N.A. Discrimination and Characterisation of Extra Virgin Olive Oils from Three Cultivars in Different Maturation Stages Using Fourier Transform Infrared Spectroscopy in Tandem with Chemometrics. *Food Chem.* **2015**, *174*, 226–232. [[CrossRef](#)]
92. Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Calvo, A.; Busto, O. Prediction of Olive Oil Sensory Descriptors Using Instrumental Data Fusion and Partial Least Squares (PLS) Regression. *Talanta* **2016**, *155*, 116–123. [[CrossRef](#)]
93. Baumann, K. Cross-Validation as the Objective Function for Variable-Selection Techniques. *TrAC Trends Anal. Chem.* **2003**, *22*, 395–406. [[CrossRef](#)]
94. Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Ren, D.; Liu, S.; Liang, Y. Chemometric Methods in Data Processing of Mass Spectrometry-Based Metabolomics: A Review. *Anal. Chim. Acta* **2016**, *914*, 17–34. [[CrossRef](#)] [[PubMed](#)]
95. Esslinger, S.; Riedl, J.; Faulh-Hassek, C. Potential and Limitations of Non-Targeted Fingerprinting for Authentication of Food in Official Control. *Food Res. Int.* **2014**, *60*, 189–204. [[CrossRef](#)]
96. Breznau, N.; Rinke, E.M.; Wuttke, A.; Nguyen, H.H.V.; Adem, M.; Adriaans, J.; Alvarez-Benjumea, A.; Andersen, H.K.; Auer, D.; Azevedo, F.; et al. Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2203150119. [[CrossRef](#)] [[PubMed](#)]
97. Kessler, R.W.; Kessler, W.; Zikulnig-Rusch, E. A Critical Summary of Spectroscopic Techniques and Their Robustness in Industrial PAT Applications. *Chem. Ing. Tech.* **2016**, *88*, 710–721. [[CrossRef](#)]
98. Skov, T.; Honoré, A.H.; Jensen, H.M.; Næs, T.; Engelsen, S.B. Chemometrics in Foodomics: Handling Data Structures from Multiple Analytical Platforms. *TrAC Trends Anal. Chem.* **2014**, *60*, 71–79. [[CrossRef](#)]
99. Granato, D.; Ares, G. *Mathematical and Statistical Methods in Food Science and Technology*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2013; pp. 1–513. [[CrossRef](#)]
100. Pérez-Marín, D.C.; Garrido-Varo, A.; Guerrero-Ginel, J.E.; Gómez-Cabrera, A. Near-Infrared Reflectance Spectroscopy (NIRS) for the Mandatory Labelling of Compound Feedingstuffs: Chemical Composition and Open-Declaration. *Anim. Feed. Sci. Technol.* **2004**, *116*, 333–349. [[CrossRef](#)]
101. Dong, J.; Zhuang, D.; Huang, Y.; Fu, J. Advances in Multi-Sensor Data Fusion: Algorithms and Applications. *Sensors* **2009**, *9*, 7771–7784. [[CrossRef](#)]
102. Xu, Y.; Koidis, A.; Tian, X.; Xu, S.; Xu, X.; Wei, X.; Jiang, A.; Lei, H. Bayesian Fusion Model Enhanced Codfish Classification Using Near Infrared and Raman Spectrum. *Foods* **2022**, *11*, 4100. [[CrossRef](#)]
103. Yang, Y.; Wang, W.; Zhuang, H.; Yoon, S.C.; Jiang, H. Fusion of Spectra and Texture Data of Hyperspectral Imaging for the Prediction of the Water-Holding Capacity of Fresh Chicken Breast Filets. *Appl. Sci.* **2018**, *8*, 640. [[CrossRef](#)]
104. Kucha, C.T.; Liu, L.; Ngadi, M.; Gariépy, C. Anisotropic Effect on the Predictability of Intramuscular Fat Content in Pork by Hyperspectral Imaging and Chemometrics. *Meat Sci.* **2021**, *176*, 108458. [[CrossRef](#)] [[PubMed](#)]
105. Gredell, D.A.; Schroeder, A.R.; Belk, K.E.; Broeckling, C.D.; Heuberger, A.L.; Kim, S.Y.; King, D.A.; Shackelford, S.D.; Sharp, J.L.; Wheeler, T.L.; et al. Comparison of Machine Learning Algorithms for Predictive Modeling of Beef Attributes Using Rapid Evaporative Ionization Mass Spectrometry (REIMS) Data. *Sci. Rep.* **2019**, *9*, 5721. [[CrossRef](#)] [[PubMed](#)]
106. Czabanski, R.; Jezewski, M.; Leski, J. Introduction to Fuzzy Systems. *Stud. Fuzziness Soft Comput.* **2017**, *356*, 23–43. [[CrossRef](#)]
107. Peris-Díaz, M.D.; Krężel, A. A Guide to Good Practice in Chemometric Methods for Vibrational Spectroscopy, Electrochemistry, and Hyphenated Mass Spectrometry. *TrAC Trends Anal. Chem.* **2021**, *135*, 116157. [[CrossRef](#)]
108. Liu, Z.; Wang, S.; Zhang, Y.; Feng, Y.; Liu, J.; Zhu, H. Artificial Intelligence in Food Safety: A Decade Review and Bibliometric Analysis. *Foods* **2023**, *12*, 1242. [[CrossRef](#)] [[PubMed](#)]
109. Mishra, P.; Passos, D.; Marini, F.; Xu, J.; Amigo, J.M.; Gowen, A.A.; Jansen, J.J.; Biancolillo, A.; Roger, J.M.; Rutledge, D.N.; et al. Deep Learning for Near-Infrared Spectral Data Modelling: Hypes and Benefits. *TrAC Trends Anal. Chem.* **2022**, *157*, 116804. [[CrossRef](#)]
110. Dora, M.; Kumar, A.; Mangla, S.K.; Pant, A.; Kamal, M.M. Critical Success Factors Influencing Artificial Intelligence Adoption in Food Supply Chains. *Int. J. Prod. Res.* **2022**, *60*, 4621–4640. [[CrossRef](#)]
111. Kudashkina, K.; Corradini, M.G.; Thirunathan, P.; Yada, R.Y.; Fraser, E.D.G. Artificial Intelligence Technology in Food Safety: A Behavioral Approach. *Trends Food Sci. Technol.* **2022**, *123*, 376–381. [[CrossRef](#)]
112. Georgouli, K.; Osorio, M.T.; Martínez Del Rincon, J.; Koidis, A. Data Augmentation in Food Science: Synthesising Spectroscopic Data of Vegetable Oils for Performance Enhancement. *J. Chemom.* **2018**, *32*, e3004. [[CrossRef](#)]
113. Rutan, S.C.; De Noord, O.E.; Andréa, R.R. Characterization of the Sources of Variation Affecting Near-Infrared Spectroscopy Using Chemometric Methods. *Anal. Chem.* **1998**, *70*, 3198–3201. [[CrossRef](#)]
114. Moisés, A.G.; Pascual, I.V.; González, J.J.I.; Zamarreño, C.R. Data Augmentation Techniques for Machine Learning Applied to Optical Spectroscopy Datasets in Agrifood Applications: A Comprehensive Review. *Sensors* **2023**, *23*, 8562. [[CrossRef](#)] [[PubMed](#)]
115. Segtnan, V.H.; Mevik, B.H.; Isaksson, T.; Næs, T. Low-Cost Approaches to Robust Temperature Compensation in Near-Infrared Calibration and Prediction Situations. *Appl. Spectrosc.* **2005**, *59*, 816–825. [[CrossRef](#)] [[PubMed](#)]

116. Zhang, L.; Wang, Y.; Wei, Y.; An, D. Near-Infrared Hyperspectral Imaging Technology Combined with Deep Convolutional Generative Adversarial Network to Predict Oil Content of Single Maize Kernel. *Food Chem.* **2022**, *370*, 131047. [[CrossRef](#)] [[PubMed](#)]
117. Truong, V.K.; Dupont, M.; Elbourne, A.; Gangadoo, S.; Pathirannahalage, P.R.; Cheeseman, S.; Chapman, J.; Cozzolino, D. From Academia to Reality Check: A Theoretical Framework on the Use of Chemometric in Food Sciences. *Foods* **2019**, *8*, 164. [[CrossRef](#)]
118. Okolo, C.A.; Kilcawley, K.N.; O'Connor, C. Recent Advances in Whiskey Analysis for Authentication, Discrimination, and Quality Control. *Compr. Rev. Food Sci. Food Saf.* **2023**, *22*, 4957–4992. [[CrossRef](#)]
119. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems, Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 5999–6009.

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.