processes

MDPI

Review

The Potential of Spectroscopic Techniques in Coffee Analysis—A Review

Leah Munyendo [1,2,*], Daniel Njoroge [2] and Bernd Hitzmann [1]

[1] Department of Process Analytics and Cereal Science, University of Hohenheim, Garbenstr. 23, 70599 Stuttgart, Germany; bernd.hitzmann@uni-hohenheim.de
[2] Institute of Food Bioresources Technology, Dedan Kimathi University of Technology, Nyeri 10143, Kenya; daniel.njoroge@dkut.ac.ke
* Correspondence: leah.munyendo@uni-hohenheim.de; Tel.: +49-1573-2373354

Abstract: This review provides an overview of recent studies on the potential of spectroscopy techniques (mid-infrared, near infrared, Raman, and fluorescence spectroscopy) used in coffee analysis. It specifically covers their applications in coffee roasting supervision, adulterants and defective beans detection, prediction of specialty coffee quality and coffees' sensory attributes, discrimination of coffee based on variety, species, and geographical origin, and prediction of coffees chemical composition. These are important aspects that significantly affect the overall quality of coffee and consequently its market price and finally quality of the brew. From the reviewed literature, spectroscopic methods could be used to evaluate coffee for different parameters along the production process as evidenced by reported robust prediction models. Nevertheless, some techniques have received little attention including Raman and fluorescence spectroscopy, which should be further studied considering their great potential in providing important information. There is more focus on the use of near infrared spectroscopy; however, few multivariate analysis techniques have been explored. With the growing demand for fast, robust, and accurate analytical methods for coffee quality assessment and its authentication, there are other areas to be studied and the field of coffee spectroscopy provides a vast opportunity for scientific investigation.

Keywords: spectroscopy; coffee; chemometrics; prediction; adulterants; defective beans; discrimination; roasting

Citation: Munyendo, L.; Njoroge, D.; Hitzmann, B. The Potential of Spectroscopic Techniques in Coffee Analysis—A Review. *Processes* **2022**, *10*, 71. https://doi.org/10.3390/pr10010071

Academic Editor: Dariusz Dziki

Received: 30 November 2021
Accepted: 27 December 2021
Published: 30 December 2021

1. Introduction

Coffee is one of the most widely traded commodities worldwide and is produced in over 50 countries. According to ICO [1], world green coffee production reached 10.52 million tonnes in 2020, a 6.3% increase compared to 2019. Generally, production has been increasing year by year, clearly indicating an upward trend in coffee consumption. This can be attributed to coffees' desirable sensory properties, the stimulant effects of caffeine, as well as its benefits on human health. More than 100 species in the genus *Coffea* have been identified; however, only two species, namely, *Coffea canephora* and *Coffea arabica*, commonly known as robusta and arabica, respectively, are traded commercially. Arabica coffee accounts for more than 60% of the world's coffee production with a higher commercial value than robusta coffee due to its superior aroma characteristics and lower caffeine content [2].

Assessment of coffee along its supply chain, i.e., from farm to cup, is crucial in ensuring consistent supply of quality products to consumers. The overall quality of a cup of coffee is directly linked to careful control of a number of parameters including the quality of cherries, raw materials (species, origin, presence of adulterants and defective beans, type of processing method (wet or dry), chemical composition etc.), roasting process, storage conditions, and preparation parameters and methods. In the past years, different

analytical techniques have been developed and used to authenticate and assess the quality of coffee. They may involve target analyses, in which chemical profiling is done and the identified compounds are assessed against a set limit established by law or stated on the label. Non-target analyses, in contrast, provide a fingerprint of the whole sample that can then be utilized to detect possible adulteration. Physicochemical and chromatographic analytical methods are the predominant methods in the evaluation of coffee species, origin, and chemical composition [3,4]. Although very objective, these methods usually require experienced personnel, and are labor intensive, expensive, and time consuming. Visual coffee inspection is also a common practice that is done to check for the presence of defective beans, adulterants, degree of roast as well as to distinguish coffee species based on the colour, shape, and size of the bean. This is a subjective method that depends mainly on the experience and skills of the inspector and hence could be subject to erroneous conclusions.

The global coffee market is increasingly stressing the importance of the quality and chemical characteristics of coffee, which calls for appropriate analytical methods. Spectroscopy is a promising technique in this industry because it is rapid, non-destructive, requires no/minimal sample preparation, and integrates easily into other processes. Infrared spectral techniques have gained considerable interest in evaluation of coffee with respect to its chemical composition, species, geographical origin, presence of adulterants and defective beans, roast degree, sensory attributes, etc. [5–10]. The potential of Raman and fluorescence spectroscopy in coffee analysis have also been investigated [11–14].

Potential application of spectroscopy techniques in the food industry has gained interest over the last few years. Therefore, the aim of this review is to provide a summary of the recent studies on applications of near infrared, fluorescence, mid infrared, and Raman spectral techniques in coffee assessment. It is hoped that this will provide an overview to interested parties on what can be achieved using spectroscopy techniques in coffee analysis. Additionally, aspects that still require clarification or further investigations are identified. However, this review does not address all the aspects related to coffee assessment. Important topics such as prediction of coffee processed using different methods, differentiation of organic from inorganic coffee, monitoring of coffee quality during storage, among others, were not considered. Limited information is available, and therefore these are areas that could be considered in the future given their significant impact on the quality of the beverage. In this review, first a brief introduction to the basic concepts of the aforementioned spectroscopic methods is provided followed by discussion of the latest research carried out to develop new approaches based on these methods.

2. Overview of Spectroscopic Techniques

Spectroscopic techniques have gained interest in the food industry over the past years due to their potential in assessing the chemical composition and quality of foods. These techniques involve interaction between matter and electromagnetic radiation, which provides useful information about food samples and can be used for both qualitative and quantitative analysis. These methods present some advantages over conventional analytical methods given their rapidity, provision of on-line analysis, minimal or no sample preparation, and the fact that they are non-destructive [15]. Fluorescence spectroscopy techniques utilizes a beam of light, which excites the electrons in molecules of certain compounds commonly known as fluorophores, causing them to emit light. At a specific wavelength, fluorophores will absorb energy in the form of light and later liberate this energy in the form of light emission at a longer wavelength.

The processes that occur between light absorption and emission are clearly explained in Figure 1 for fluorescence, NIR, and Raman spectroscopy. In fluorescence (Figure 1a) visible light is involved, where the electronic state of a molecule is changed. In the figure, S_0 and S_1 denote the singlet ground and first excited electronic state, respectively. The fluorescence process starts with the excitation where a molecule absorbs light and gets excited from ground state (S_0) to a higher vibrational level in S_1. This is followed by rapid relaxation of molecules from an upper vibrational level to the lowest level in the

first electronic excited state S_1, without any radiation. This process is known as internal conversion and normally happens within 10^{-12} s or less. Finally, emission occurs usually 10^{-8} s after excitation, when the molecule goes back to any vibrational state of its electronic ground state, S_0. In modern food analysis, the fluorescence spectroscopy method has been extensively used in the evaluation of food components, adulterants, contaminants, and additives, thanks to its high specificity and sensitivity [16,17].

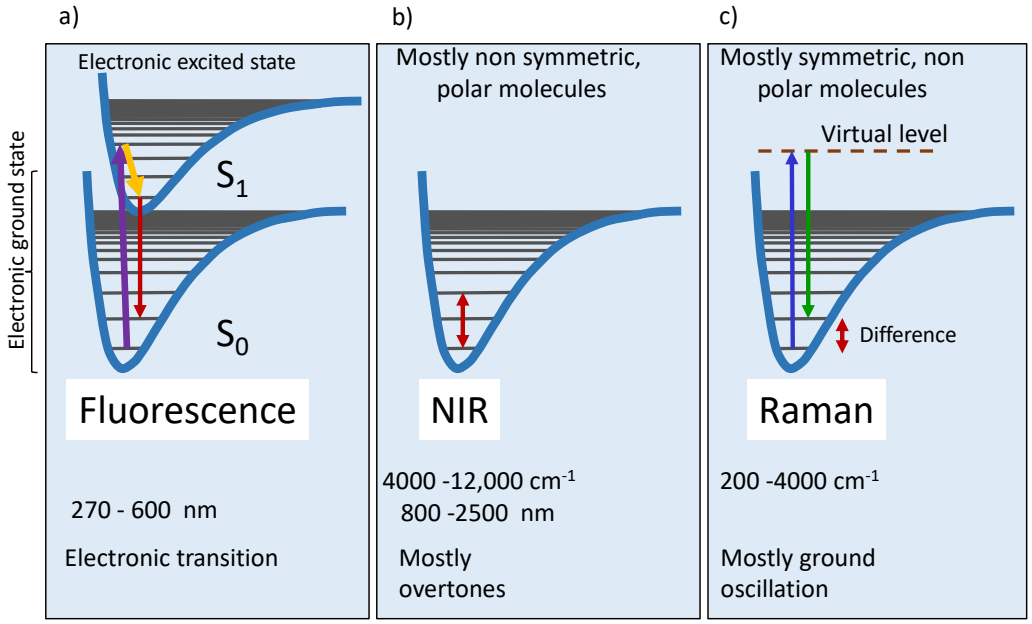

Figure 1. Simplified potential curves to illustrate different spectroscopic methods discussed here; just vibrational states are presented: (**a**) fluorescence, (**b**) NIR, (**c**) Raman. The red arrow indicates the energy involved in the corresponding radiation.

Infrared (IR) spectroscopy is among the most common vibrational spectroscopic methods used in the food industry. It is usually divided into three regions of wavelength: far infrared (15 µm–1 mm or 400–10 cm^{-1}), mid infrared (MIR: 4000–400 cm^{-1} or 2500–25,000 nm), and near infrared (NIR:13,333–4000 cm^{-1} or 800–2500 nm) [18]. IR spectroscopy is based on the principle that every chemical bond vibrates with specific frequency, which corresponds to certain amount of energy. These vibrations induce absorption bands, useful in the qualitative and quantitative analyses of many molecules in samples. Vibrations of certain functional groups, for instance C=O, NH_2, –OH, $–CH_3$, C_6H_5– etc., always produce bands in the infrared spectrum within frequency ranges that are well-defined irrespective of the molecule having the functional group. Identification and designation of these bands to particular chemical groups provide specific information about the product being investigated [17,19].

Broad bands rising from combinations and overtones of the fundamental vibrations relating to O–H, N–H, and C–H chemical bonds characterize spectra in the NIR region (Figure 1b). Samples with different chemical bonds, i.e., chemical composition, will record different spectra, which are unique to them acting as a 'fingerprint'. NIR spectroscopy has been widely used for analysing the quality and chemical composition of foods including meat, fish, milk, cereals, fruits, cereals, wine, beer, eggs, and their related products [17,19,20]. Despite its great potential in food analysis, the interpretation of the spectra in NIR analysis is challenging due to its broadband nature, which consists of overlapping overtone and combination bands. Specific compounds are not as well determined as they are in the MIR region. This has however, been addressed by data processing using chemometric methods to relate spectral information to properties of the samples [21].

Mid-infrared spectroscopy provides information on the related rotational-vibrational structure as well as fundamental vibrations of the chemical bonds in the samples. Through technological advances, original MIR spectrophotometers, which measured the amount of absorbed energy based upon the fact that the infrared light frequency varies when passing through a monochromator, have been replaced by Fourier transform infrared (FTIR) spectrophotometers. FTIR equipment uses interferometers instead of monochromators, yielding advantages such as high sensitivity and precision, fast analysis, and a high signal-to-noise ratio. Consequently, the term "FTIR spectroscopy" and "MIR spectroscopy" are sometimes interchangeably used [22], although FTIR is also applied for NIR. The mid-infrared spectrum is comprised of four regions: the triple bond, the X–H stretching, the fingerprint, and the double bond regions. Fundamental vibrations from O–H, C–H, and N–H stretching characterize the X–H stretching region. The triple bond stretching region is generally due to C≡N and C≡C bonds vibrations, whereas in the double bond region, absorption bands characterized by C=N, C=C, and C=O occur. Bending and skeletal vibrations characterize the fingerprint region [17]. Utilization of MIR spectroscopy has been used as an effective analytical method in the food industry [10,23]. The far infrared (far-IR) spectral region (10–400 cm^{-1}) is the range of wavenumbers characterized by large-amplitude anharmonic vibrations. The technique based on the far-IR has shown to be more suitable in providing information on dynamics of proteins and analyzing structures that are highly ordered, for instance, the formation of fibrillar, due to its sensitivity to vibrational modes resulting from hydrogen bonds and peptide skeletons [24].

In Raman spectroscopy (Figure 1c), a virtual excited energy state is involved. Comparing Raman spectroscopy to IR spectroscopy, the information provided by Raman is due to inelastic light scattering, as opposed to IR, which relies on elastic scattering. In contrast to IR, Raman spectra are obtained from symmetric and non-polar molecules. Raman scattering results from inelastic scattering of the incident photons through which energy is received from or transferred to the molecule owing to the changes in the vibrational or rotational modes of sample molecules. This causes a change in the energy, and thus the frequency of the scattered light [25].

A Raman spectrum is a plot of scattering intensity vs. wavenumber or wavelength consisting of bands, which corresponds to a Raman shift that occurs because of the difference in excitation and emission light energy (see Figure 1c). So-called Stokes–Raman lines are due to smaller emission energy than excitation, and for Anti-Stokes–Raman lines this behavior is reversed, but with much smaller intensity. The obtained Raman bands characterize specific functional groups and/or chemical bonds (C–X (X = Cl, I, F, or Br), C–S, C–NO_2, S–S, C=N, C=C) of materials and by using them, it is possible to obtain fingerprints of materials with Raman spectroscopy. As the bands' intensity is directly proportional to the analyte concentration of the materials, quantitative analyses using Raman spectroscopy is possible [26]. Analyses of safety, quality, and composition of foods are among the applications of Raman spectroscopy in the food industry [11,27,28].

In summary, the potential of spectroscopic methods in the food industry is enormous and broad. With regard to coffee, there is much literature on the utilization of these techniques in determining coffees authenticity, quality, geographical origin, species, and composition and detecting misrepresentation and adulteration. Table 1 shows reported absorption bands of coffees chemical components in the IR and Raman region.

Table 1. Reported absorption bands of coffees chemical components in the IR and Raman region.

	Wavenumber (cm^{-1}) or Wavelength (nm)	Vibrational Modes	Associated Compounds	References
NIR				
	4750 and 6900 and 5200 cm^{-1}	-	Proteins and water, respectively	[29]
	4200 to 4400 cm^{-1}	-	Carbohydrates	
	1450 nm	1st overtone of O–H stretching	Water	[30]
	1100 to 1250 nm	2nd overtone of C–H stretching	Carbohydrates, quinic acid, and lipids	
	1300–1350 nm	-	Caffeine	
	1550 nm	-	CGA, carbohydrates, and amino acids	
	5000–5200 cm^{-1} and 6800–7200 cm^{-1}	-	Water	[31]
	4000–5000 cm^{-1}	C-H bonds vibrations	CGA, carbohydrates, proteins, and trigonelline	
	900–1000 nm	3rd overtone of the CH, CH_2, and CH_3 groups	Ferulic and coumaric acids	[32]
	1400–1500 nm	1st overtone of the OH functional group	CGA, water, and carbohydrates	
	6750–6950 cm^{-1} and 5100–5200 cm^{-1}	1st overtone of O–H stretching and O–H deformation + O–H stretching combination bands	Water	[33]
	5680–5850 and 4760 cm^{-1}	-	Lipids and CGA, proteins, caffeine, and carbohydrates, respectively	
	1850–1950 nm	1st overtone of C=O bonds vibrations	CGA, proteins, lipids, caffeine, water, and carbohydrates	[34]
	4000–12,000 cm^{-1}	Overtone and combination bands of the NH, C–H, SH and OH bonds	Proteins (amino acids) and lipids	[35]
	4000–5600 cm^{-1}	C-H bonds vibrations	Caffeine, fatty acids, amino acids, and lignin	[36]
	8200–8300 cm^{-1}	C-H bonds vibrations	Fatty acids, amino acids, and lignin	
	10,000, 6800, 4400 and 4800 cm^{-1}	O-H and C-H bond vibrations	Cellulose	
	1410, 1742, 1904, and 2318 nm	-	Lipids	[37]
	1410, 1728, 1904, 2306 and 2348 nm	-	Caffeine	
	1436, 1880, 2312, 2324 and 2350 nm	-	CGA	
	5000–5200 and 6800–7200 cm^{-1}	-	Water	[38]
	5325, 5140, 5075, 5410–5385 and 4980–5000 cm^{-1}	1st overtone and combination band of C=O and O-H bonds, 2nd overtone of C=O	Carbohydrates, proteins, and organic acids	

Table 1. *Cont.*

	Wavenumber (cm^{-1}) or Wavelength (nm)	Vibrational Modes	Associated Compounds	References
	1868 nm	-	5-CQA	[39]
	1741 and 2252 nm 2383–2488, 1977–2266, 1638–1793 and 1297–1474 nm	-	Caffeine	
	2306–2312 and 1218–1242 nm	C–H + CC combination bands and 2nd overtone of C–H, respectively	Sucrose and other carbohydrates	[40]
	1472–1478 nm	1st overtone of N–H	Phenols and CGA	
	2128–2132 and 2190–2192 nm	O–H 1st overtone and combination bands of N–H	Proteins and CGA	
	1706–1714, 2436–2475 and 2480–2488 nm	Combination bands of C–H + C–H and C–H + CC and C-H 1st overtone	Lipids	
	1440–1480 and 1930–1950 nm	O–H stretching 1st overtone and combination band of O–H deformation + O–H stretching	Water	[41]
	1715–1760 and 2300–2350 nm	-	Lipids	
	1180–1262 nm	-	Caffeine and cellulose	[5]
	1480–1882 nm and 1896–2180 nm	-	CGA, caffeine, lipids, proteins, and carbohydrates	
	2260–2498 nm	-	Caffeine, carbohydrates, and protein	
	1208 nm	C-H stretching 2nd overtone	Sucrose, lipids, and amino acids	[42]
	1800–2000, 1400, 2200 and 2300 nm	1st overtones of O–H and C-H	Water	
	Around 1500 nm	1st overtone of aromatic structures C–H	Phenolic compounds (CGA)	
	1700 and 1800 nm	C-H 1st overtone	Lipids, caffeine, and carbohydrates	
	2300–2400 nm	C-H + CH_2 combination bands	Lipids	
	6896 and 5154 cm^{-1}	O-H deformation and stretching combination bands	Water	[43]
	4100–4400 cm^{-1}	-	CGA and quinic acid	
MIR/ FTIR				
	2840–2940 cm^{-1}	Stretching of CH bonds in CH_2 and CH_3 groups	Lipids and caffeine	[41]
	1747 cm^{-1}	C=O stretch of aliphatic ester groups	Lipids	
	900–1400 cm^{-1}	-	Carbohydrates	

Table 1. *Cont.*

Wavenumber (cm^{-1}) or Wavelength (nm)	Vibrational Modes	Associated Compounds	References
1740 and 1660 cm^{-1}	Stretching of C=O and C=C bonds	Carbohydrates and lipids, respectively	[44]
1600, 1700, 1260, 1160 and 1060 cm^{-1}	-	CGA	
1330 and 600–1300 cm^{-1}	-	Trigonelline and Pyridine	
1020 cm^{-1}	-	Carbohydrates	
2922, 2852 and 1743 cm^{-1}	Asymmetric and symmetric stretching of CH_2, and stretching of C=O, respectively	Lipids	[45]
1153 and 950–1130 cm^{-1}	-	Polysaccharides and other carbohydrates, respectively	
1643 cm^{-1}	C=O stretching vibrations	Caffeine	
650–900 cm^{-1}	-	Proteins and amino acids	[46]
1000 and 1300 cm^{-1}	-	Carbohydrates, CGA, proteins, amino acids	
1700–1750 and 1600–1680 cm^{-1}	C=O stretching	Carbohydrates and lipids	
3470 cm^{-1}	O-H stretching	CGA and Water	[47]
3008 and 2855–2920 cm^{-1}	Stretching of the C=C and C-H bond, respectively	Lipids and polysaccharides (lignin), respectively	
1746 and 1704 cm^{-1}	Stretching and vibration of the ester group OC=O and carbonyl group C=O	Quinic acids and fatty acids, respectively	
1608 cm^{-1}	Vibrations of the C-N group	Trigonelline and caffeine	
1745 cm^{-1}	-	Triglyceride	[48]
835, 911 and 1061 cm^{-1}	-	Polysaccharides	
1268, 1291, 1322, 1337, 1374, 1396, 1398, 1406, 1418, and 1499 cm^{-1}	-	Proteins and organic acids	
1656 and 2800–3000 cm^{-1}	-	Caffeine	[49]
1743–1741 and 1543 cm^{-1}	-	Lipids and caffeine or/and trigonelline	
1744, 1654, and 1603 cm^{-1}	-	Lipid and caffeine	[50]
1285, 900–1200 and 1400–1500 cm^{-1}	-	CGA and carbohydrates	
804, 991, and 1020 cm^{-1}	-	Carbohydrates	[10]
1661, 1744 and 2922 cm^{-1}	-	Caffeine, triglycerides and lipids, respectively	
1600–1650 and 1150–1450 cm^{-1}	-	Caffeine and CGA, respectively	[8]
600–1700 cm^{-1}	-	CGA, carbohydrates, and trigonelline	

Table 1. *Cont.*

	Wavenumber (cm^{-1}) or Wavelength (nm)	Vibrational Modes	Associated Compounds	References
	3356 and 1067 cm^{-1}	-	CGA and pyruvic acid, pyridine, and quinic acid, respectively	[51]
	1000 and 1750 cm^{-1}	-	Caffeine and trigonelline	
	1744 and 1285 cm^{-1}	-	Esters and CGA, respectively	[52]
	1600–1650 and 900–1200, 1400–1500 cm^{-1}	-	Caffeine and carbohydrates, respectively	
Raman				
	1500 and 1567 and 1478 cm^{-1}	C=C stretch vibrations	Cafestol and Kahweol, respectively	[53]
	1600 cm^{-1}	Phenyl ring stretch		[11]
	1630 cm^{-1}	C=C stretch vibrations		
	1120 cm^{-1}	CH and COH bending vibration	Lipids, CGA, and proteins	
	1200 cm^{-1}	Phenyl ring bending vibration		
	1604 and 1630 cm^{-1}	Aromatic and C=C stretching	Polyphenols, e.g., CGA	[54]
	1690 and 1656 cm^{-1}	Amide I band stretching (structures of R-helix and β-sheet)	Proteins	
	1479 and 1567 cm^{-1}	-	Kahweol	
	1630 and 1605 cm^{-1}	C=C ethylenic stretch and Phenyl ring stretch vibrations, respectively	CGA	[14]
	1000–1750 cm^{-1}	-	CGA	
	2934 and 2905 cm^{-1}	CH_2 asymmetric and symmetric vibrations	Lipids	
	2700–3050 cm^{-1}	-	Lipid bands	
	1507 and 1485 and 1570 cm^{-1}	-	Cafestol and Kahweol, respectively	[55]
	1606, 1637, 1657 and 1680 cm^{-1}	C=C and C=O bonds vibration	CGA	[56]
	1567 and 1478 cm^{-1}	-	Kahweol	
	2900 cm^{-1}	C-H bond symmetric and asymmetric stretching	Fatty acids	
	1441, 1304 and 1265 cm^{-1}	C=C deformation vibrations of CH, CH_2, and CH_3 bonds	Fatty acids	
	1502 and 1442 cm^{-1}	-	Cafestol and lipids, respectively	

3. Application of Spectroscopy Techniques in Coffee Analysis

3.1. Coffee Roasting and Monitoring

Roasting is a significant step in coffee processing where physical, chemical, and structural transformations occur. Moreover, there is production of melanoidins and thousands of flavour compounds, all of which are accountable for the characteristic colour, aroma, and flavour of the beverage [57]. Therefore, roasting is a step with a significant impact

on the quality of the final product, thus warranting its control. Several studies on monitoring and controlling the coffee roasting process using spectroscopic methods have been conducted (Table 2). Bertone et al. [58] studied the ability of Fourier transform (FT-NIR) spectroscopy to determine roast degree (best on colour) and varietal composition of coffee simultaneously. In the study, 130 blends of roasted-ground coffee containing both arabica and robusta species at different levels were used. The partial least square regression (PLS) algorithm was used, and validation was done with a completely independent external set. The developed model had a root mean square error of prediction (RMSEP) of 1.28 Au for colour and 4.34% (*w*/*w*) for arabica content in the blends, thus demonstrating a fast and reliable assessment of important coffee characteristics.

Table 2. Summary of reported studies on potential applications of NIR, FTIR/MIR, fluorescence and Raman spectroscopy in the coffee industry.

Application	Aim	Spectroscopy Technique	Multivariate Analysis	References
Coffee roasting and monitoring	Determine roast degree (based on color, acidity, cracks)	NIR	PLS	[6,7,58,59]
	Online acidity monitoring during coffee roasting	NIR	PLS	[38]
	In-line coffee roasting process monitoring	NIR	PLS	[31]
	Real-time detection of faults during coffee roasting process	NIR	PCA	[60]
Adulterants and defective beans detection	Identify and quantify adulterants (corn, barley, peels, sticks, coffee husks, soy, ort, rice, sticks, soil, and robusta) in roasted-ground coffee	NIR	PCA, PLS, MCR-ALS and SIMCA	[32,47,61,62]
	Detect and quantify adulterations of roasted-ground coffee with corn	FTIR/MIR	PCA and PLS	[63]
	Quantify defective beans (black, immature, dark sour and light sour) in roasted coffees	FTIR/MIR and NIR	PLS and Elastic net	[41]
	Discriminate between defective (sour, immature, insect damaged, and black) and nondefective green coffee beans	NIR	PCA, PLS, CEM-SVM and CNN	[64,65]
	Discrimination of mature, sour, black, and immature green coffee beans	FTIR/MIR	AHC, LDA and PCA	[49,66]
	Evaluation of defects in roasted-ground coffee	MIR/FTIR-PAS	PLS-DA	[51]
Prediction of specialty coffee quality and sensory attributes of coffee	Determine sensory attributes of coffee	NIR	PLS, SNM, DCNN	[40,67,68]
	Prediction of specialty coffee quality and its discrimination from ordinary coffee	NIR	PLS, FFBPANN, CACHAS	[33,69,70]
	Prediction of coffee cup quality	FTIR/MIR	PCA, PLS-DA	[45,71]
	Quantitative evaluation of sensory characteristics of specialty coffees and its discrimination	FTIR/MIR	PLS, FA	[8,44]
	Discrimination of specialty coffee	Fluorescence	PCA and SIMCA	[12]

Table 2. *Cont.*

Application	Aim	Spectroscopy Technique	Multivariate Analysis	References
Discrimination of coffee based on species, variety, and geographical origin	Distinguish between robusta and arabica coffee species and their varieties	NIR	Potential function class-modelling, PCA, SOM, SIMCA, PLS-DA and SVM	[72–74]
	Discriminate coffee of different origin	NIR	PLS-DA, SNM, SIMCA, ANN (SOM)	[5,34,36,37,52,73]
	Coffee variety identification	FTIR/MIR	ELM, BPNN, RBFNN, RVM, SIMCA and SVM	[48]
	Discriminate coffee of different origin	FTIR/MIR	PLS, SVM, RBFs, MLP, SIMCA	[50,52,73,75]
	Discrimination of coffee species (arabica and robusta) and cultivars	Raman	PCA, MDA, QDA, RDA, PLS–DA, SIMCA, LDA	[11,14,53–56]
	Geographical and phenotypic discrimination of coffees	Fluorescence	PARAFAC, NPLS-DA, UPLS-DA, MLR and LDA	[13,76]
Prediction of coffee chemical composition	Predict moisture content, soluble solids, total reducing sugars, lipids, proteins, trigonelline, 5-CQA, caffeine, and sucrose content of coffee	NIR	PLS	[9,30,35,39,42,77–79]
	Predict CGA isomer composition and caffeine content	FTIR/MIR	PLS	[80,81]
	Determine trigonelline and caffeine content	Fluorescence	-	[81,82]

Online monitoring of acidity during coffee roasting has also been studied using FT-NIR techniques [38]. Coffee samples (arabica and robusta) were roasted using fast and slow roasting designs at 220 °C for 17 min and 183 °C for 25 min, respectively. Online acquisition of the spectra was done in the 10,000 and 4000 cm^{-1} range. In order to develop a calibration model between titratable acidity and online-acquired NIR spectra, important wavelength regions were selected (5150–5950 cm^{-1}, 4000–4760 cm^{-1}, 6320–7080 cm^{-1}, and 7470–9400 cm^{-1}) and used in the development of a PLS model for acidity. These selected regions are characteristic of 5-caffeoylquinic acid and were highlighted by Ribeiro et al. [40], who aimed to link perceived acidity of beverages with arabica coffees chemical composition. The proposed model had a RMSEP and range-error-ration (RER) of 0.16 and 11, respectively, indicating its fairness in acidity prediction during coffee roasting. Yergenson and Aston [7] carried out a similar study with the aim of determining coffee roast degree online by controlling acidity. Coffee was roasted using four different temperature-time treatments and NIR spectra acquired online in the 4000–7405 cm^{-1} range. Measurement of titratable acidity over a varied range of roast times and temperatures followed by modeling the findings allowed for the quantification of the relationships between degree of roast, acidity, and roast time.

In another study, Santos et al. [31] explored the ability of FT-NIR spectroscopy in in-line monitoring of coffee roasting process by measuring sucrose and colour. PLS regression was utilized to develop predictive models by regressing spectral data against content of sucrose and L*, a*, and b* colour values. For the case of sucrose, important wavenumbers with significant contribution included 4980–5000, 5385–5410, 5325, 5075, and 5140 cm^{-1}. The C=O second overtone region, first overtone and combination band region of O-H and C=O bonds characterize the selected wavenumbers. Thus, the region could be associated with carbohydrates, organic acids, and proteins. Colour characterization of coffee beans was

based on L* parameters where wavenumbers 6100, 5300, 5200, and 4965 cm^{-1} were selected as having the highest influence. The regions capture vibrations from C=O (5300 cm^{-1}), combination band vibrations from O-H (5200 cm^{-1}), first overtone vibrations of C-H (6100 cm^{-1}), and combination vibrations of N-H (4965 cm^{-1}). The authors concluded that free amino acids, carbohydrates, and chlorogenic acids could be the main compounds associated with colour during coffee roasting because they are involved in important chemical reactions to form colour pigments. The developed PLS models showed a strong prediction ability, with a co-efficient of determination of 0.85 and RER higher than 10.

Yergenson and Aston [6] explored the use of in-situ NIR spectroscopy in the prediction of cracking events (start and end) during coffee roasting to provide a more vigorous technique for controlling degree of roast based on the cracks. During coffee roasting, two sets of popping sounds occur (first and second cracks) that are important roast degree indicators for determining the roast endpoint. In this study, coffee samples were roasted using different time-temperature profiles. Based on the PLS regression (PLSR) with audio recordings from coffee roasting, in-situ NIR spectroscopy proved to be a reliable method in predicting the start and end times of first and second crack events. Models exhibited a good relationship between predicted and reference values with RMSEP values of 0.0068, 0.0091, 0.0041, and 0.0070 Au for the start of the first crack, end of the first crack, start of the second crack, and end of the second crack, respectively.

Pires et al. [59] successfully employed multivariate calibration and NIR spectroscopy as a substitute to the Agtron method to predict roast degrees in ground coffees and coffee beans. With PLS the method, the development of mathematical models was based on the relationship between data sets of NIR spectra and Agtron reference results to predict Agtron values of new coffee samples. In order to build representative models, all profiles of Agtron roasting were considered. The following NIR spectra regions were important for modeling in the whole bean coffee model: 6770–6992, 7482–8000, 5774–5992, 4136–4391, and 4806–5257 cm^{-1}. For the ground coffee model, the NIR spectra regions included 7447–8000, 6783–6922, 5317–5445, 5034–5219, and 4173–4380 cm^{-1}. The developed models presented promising results for roasting profiles prediction in roasted whole coffee beans as well as ground coffees with RMSEP values of 4.48 and 3.67, respectively.

Real-time detection of faults during the coffee roasting process by NIR spectroscopy and multivariate statistical process control (MSPC) based on principal component analysis (PCA) has also been evaluated. In one study, five batches with enforced disturbances (non-nominal batches) were prepared to mimic anomalous conditions of roasting and to authenticate the detection capability of the MSPC method. The non-nominal batches had deviations from nominal batches, i.e., higher amount of robusta coffee in the blend, roasting with lower and higher roasting power, and use of higher and lower initial mass of green coffee beans. PCA analysis of the acquired in-line NIR spectra demonstrated clear differences among non-nominal batches and between nominal batches. The best result was provided by a modelling method based on a time sliding window in terms of differentiating batches with and without disturbances, resourcing to typical MSPC charts: squared predicted error and Hotelling's T^2 statistics. In addition, a PCA model including a four minute time window with three principal components efficiently detected all the disturbances studied [60].

3.2. Prediction of Specialty Coffee Quality and Sensory Attributes

Sensory analysis of coffee is a complex process owing to the disparities of aromas, flavours, as well as diversity of compounds produced during coffee roasting. The cupping method is by far the most common professional technique for evaluating coffee, which entails professional tasters evaluating coffee based on physical quality (size, odor, colour, and shape of the beans) and cup quality (fragrance, aroma, astringency, acidity, flavour, body, bitterness, aftertaste, and overall quality) [83]. However, this method is somewhat subjective and requires professional personnel. Additionally, it is associated with various shortcomings: it is time consuming, there is disparity in sensory perception amongst

evaluators, a lack of reproducibility, and inadequate conclusions about samples, etc. In view of this, spectroscopy techniques may be used as alternative tools in the sensory evaluation of coffee as they are rapid and have the capability to reproduce results (Table 2).

Baqueta et al. [67] studied the ability of NIR spectroscopy in association with the PLS method to determine coffees sensory characteristics. Coffee samples with different variations including species, production region, variety, drying conditions, transport, postharvest process, storage times, coffee blend, coffee composition, and roasting process were used. Professional cuppers evaluated the samples based on the fragrance, acidity, aroma, bitterness, body, flavour, astringency, overall quality, and aftertaste. NIR spectra of the same samples were also taken. The performance of PLS models for each parameter were validated using the following merit parameters: sensitivity, accuracy, linearity, residual prediction deviation, fit, quantification, and detection limits. The developed models were suitable to quantify, detect, differentiate, and predict sensory characteristics of coffee samples, as all sensory characteristics were predicted with acceptable values consistent with the merit parameters.

Similarly, Ribeiro et al. [40] carried out a study using NIR in conjunction with PLS with the aim of developing predictive models for a reproducible and objective sensory analysis of coffee. Additionally, the correlation between beverage sensory characteristics and chemical composition of the green coffee bean was determined. Experts assessed the coffee samples based on flavour, acidity, bitterness, body, overall quality, and cleanliness, assigning a score to each attribute. PLS models were constructed for each attribute and they exhibited good correlation between estimated values and those supplied by the experts with RMSEP values of 0.37, 0.30, 0.25, 0.37, 0.42, and 0.30 for bitterness, acidity, flavour, cleanliness, overall quality, and body, respectively. The researchers observed that, in roasted beans, lipids and proteins were closely related to body attribute of the beverage, chlorogenic acids to acidity, caffeine and chlorogenic acids to bitterness, and finally caffeine, trigonelline, chlorogenic acid, polysaccharides, sucrose, and protein to flavour, cleanliness, and overall quality.

The market for specialty coffee is growing constantly due to the changing preferences of consumers. Specialty coffees have a higher quality than commercial coffees. They are characterized by unique flavours, have a known geographical origin, and commonly comprise coffee beans certified as organic, or with rainforest alliance and fair-trade certifications, etc. On the global market, specialty coffees are traded with a price premium of approximately 20–50% when compared to regular coffees [84]. Therefore, there is a need for more objective and reliable techniques for authenticating such coffees. Tolessa et al. [69] developed a model to predict the quality of specialty coffee based on green beans using NIR spectroscopy in conjunction with PLS analysis. Samples consisted of coffee from different environments (low, mid, and high altitudes in Ethiopia) and processed differently (washed, dry, and semi-washed methods). Professional coffee testers evaluated the samples based on body, acidity, overall cup preference, aroma, uniformity, flavour, aftertaste, cup cleanness, balance, and sweetness. The proposed models provided reliable predictions for the quality of a specialty cup as well as its different quality characteristics with RMSEP values of 1.04, 0.27, 0.22, 0.27, and 0.24 for total quality, acidity, overall preference for the cup, aftertaste, and body, respectively.

NIR spectroscopy has also been explored as an alternative to traditional methods of sensory evaluation (cupping tests) to predict flavours of specialty coffees. Models were developed using deep convolutional neural networks (DCNN) and the support vector machine (SVM) method. Training was then done and used for the prediction of several specialty coffee flavours (as presently being done by professional evaluators) utilizing readings from NIR spectra as the inputs. Descriptions of flavour were classified into nine flavour classes: fruity, floral, fermented/sour, vegetable/green, spices, roasted, sweet, cocoa/nutty, and other. The two methods yielded comparable performance, having the recall and accuracy of 70–73% for SVM and 75–77% for DCNN. This study offered an innovative and objective method in the prediction of intricate flavours present in specialty coffees [68]. A study by Arboleda [70] discriminated civet coffee from ordinary processed

coffee utilizing NIR spectroscopy with the help of feed forward back propagation artificial neural networks. The following wavelengths showed major differences among the two sample groups: 1088 nm, 907 nm, 1650 nm, and 1540 nm. A good model was developed with classification scores of 95–100%.

De-Araújo et al. [33] proposed a new non-destructive methodology based on NIR spectroscopy and CACHAS (chemometrics-assisted colour histogram-based analytical systems) for the verification of roasted-ground gourmet coffees (specialty coffee) with no sample preparation. NIR spectra of the samples (36 traditional, 10 superior, and 44 gourmet) were acquired in the 4000–10,000 cm^{-1} range. Digital images of the samples were made using a scanner module from an HP Deskjet multifunctional printer. Histograms in the hue-saturation-intensity (HSI), red-green-blue (RGB), and grayscale (GS) colour systems from each sample were acquired from a circular region of interest (ROI). In this study, one-class partial least squares (OC-PLS) and data-driven soft independent modeling of class analogy (DD-SIMCA) were employed as one-class classifiers (OCC). Models were developed using different pre-processing techniques and combinations of colour histograms for NIR spectroscopy and CACHAS, respectively. Performance of the constructed models were assessed with regard to sensitivity and specificity. DD-SIMCA using RGB histogram for CACHAS and offset correction for the NIR spectra reported the best results as they both correctly recognized 100% of the studied samples in the training set and test set.

The potential of MIR spectroscopy in association with chemometrics in predicting the cup quality of coffees roasted to different roast degrees were examined (Table 2) [45]. Coffee samples were first classified based on cup quality by skilled cuppers from worst to best. The samples were then roasted to dark, medium, and light roast degrees. PCA was used to show variability among the data as well as to detect probable outliers, whereas classification models were developed using PLS discriminant analysis (PLS-DA). Different preprocessing techniques of the raw data were compared and, with respect to PCA, the best clustering of the samples was achieved with second derivative application to the wavenumber ranges of 1680–1800, 2800–2960, and 1130–1249 cm^{-1}. Models were developed based on a two-level PLS-DA hierarchical strategy. In the first level, coffee was classified as of low or high quality followed by separation based on quality of the cup in the second level. Spectral regions 2918, 1749, 1038, and 806 cm^{-1} were the most important for the differentiation between coffees of low or high quality, whereas in the classification based on cup quality there was no particular band region that characterized one particular class. However, there were differences among the intensities. Similarly, Belchior et al. [8] investigated FTIR potential for quantitative valuation of the sensory attributes of specialty coffees. Coffee samples were evaluated by professional cuppers according to the Specialty Coffee Association of America (SCAA) protocol for sensory analysis of coffee. Based on the scores, PLS regression models were then constructed to establish and predict a sensory profile. The models showed a good relationship between predicted and experimental data with low RMSEP values of 0.23. However, there were no specific peaks significant for the coffee classification determination when correlated to the SCAA scores for the entire spectra. This can be explained by the fact that different compounds affecting the sensory profile of coffee absorb throughout the entire spectrum.

Abreu et al. [44] compared the performance of ultraviolet-visible spectroscopy (UV-Vis), high-performance liquid chromatography (HPLC), and MIR to discriminate between traditional and specialty coffees using exploratory factor analysis (FA). Specialty and traditional roasted ground coffee samples were extracted using the best solvents (water and dichloromethane), i.e., those that produced the most divergent metabolic fingerprints. Among the methods investigated, HPLC and UV-Vis spectral fingerprints of the water extracts provided more relevant discriminatory information than did MIR. Fatty acids, chlorogenic acids, organic acids, trigonelline, and lipids were the most important compounds for this discrimination. A method to discriminate espresso coffees based on their sensory characteristics using attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR) and PLS-DA models was developed by Belchior et al. [71]. Samples

were evaluated based on their flavour, aroma, acidity, aftertaste, and body, and models were developed for each attribute. The models classified the samples based on these attributes with good values of specificity and sensitivity for both calibration and validation sets. Proteins, lipids, trigonelline, caffeine, carbohydrates, chlorogenic acids, and carboxylic acids were linked to the body, flavour, and aftertaste. The acidity was linked to the presence of chlorogenic acid, carboxylic acids, and alcohols, whereas esters, ketones, overall acids, and aldehydes contributed to the aroma of the beverage.

Suhandy and Yulia [12] studied the feasibility of fluorescence spectroscopy in discriminating Indonesian specialty coffees (Pea berry, Pagar Alam, and Civet). Soft independent modeling of class analogy (SIMCA) and PCA were used as chemometric methods. The PCA approach allowed grouping of the samples into three clusters where an excitation of 370 nm was associated with this observation. In contrast, SIMCA models showed a high degree of specificity and sensitivity for calibration and prediction data sets.

3.3. Detection of Defective Beans and Adulterants in Coffee

Coffee has a high market value and as a result, its adulteration has become an extensive practice for economic gain. It can include the addition of spent coffee grounds, chicory, coffee husks, cereals (e.g., barley, corn, wheat, and rice) as well as the substitution of arabica coffee, which is more expensive, with robusta coffee [85]. It is important to emphasize that some countries have allowed the use of specific adulterants up to a certain level. However, it is illegal if the information is not declared on the label. Therefore, fast and reliable techniques are needed to protect consumers from fraudulent practices, and this has necessitated development of new techniques based on spectroscopy (Table 2). Correia et al. [32] proposed a methodology for identification and quantification of adulterants in roasted coffee based on NIR spectroscopy and multivariate calibration by PLS and PCA. Green coffee beans were roasted to different roast levels (dark, medium, and light), ground, and mixed with adulterants (robusta coffee, corn, and peels/sticks) in a range of 1 to 100% w/w. The PCA approach allowed for recognition of adulterated and pure samples with first and second components accounting for over 90% of the total spectral variance. PLS models with good sensitivity were constructed. However, the model for quantification of corn in coffee samples was the best with RMSEP and coefficient of determination (R^2) of 4% and 0.9788, respectively, compared to that of the peels/sticks. In regard to roast levels, light roasted samples gave the best RMSEP values (2.8%).

Ebrahimi-Najafabadi et al. [61] studied the adulteration of ground-roasted coffee with barley utilizing NIR spectroscopy and chemometrics methods. A widely applicable model was constructed using nine varieties of coffee samples with their mixtures adulterated with four types of barley, all, roasted to different roast degrees. Barley was mixed at levels ranging from 2 to 20% w/w. A PLS regression was developed to obtain a quantitative model for prediction of adulteration levels. The model proposed provided dependable predictions of barley adulteration at levels as low as 2% w/w with low values of root mean square errors (RMSE). Selection of variables was done using genetic algorithms, and the following spectral regions were selected as the most significant in identifying coffee adulterated with barley: 6032, 5748, 4788–4880, 4628–4688, 4336, and 4276 cm^{-1}. The regions are ascribed to the first overtone of N–H, C=O, C–H, O–H, and S–H functional groups of ROH, ArOH, H_2O, RNH_2, and CONHR. In a similar approach, Winkler-Moser et al. [62] evaluated NIR spectroscopy in conjunction with PLS regression analysis for corn adulteration detection in ground-roasted coffee samples. Different samples were prepared by mixing corn with coffee at concentrations of 0, 1, 5, 10, 15, and 20% w/w. Based on the confidence intervals of predicted percent corn at specific levels of adulteration, the developed model seemed not to be reliable in the detection of all coffee samples adulterated at levels below 4%. This contrasts with the finding of Ebrahimi-Najafabadi et al. [61], whose model was able to detect coffee adulteration with barley at 2% levels.

The capability of NIR hyperspectral imaging in combination with multivariate curve resolution-alternating least squares (MCR-ALS) to detect and quantify adulterants in

roasted-ground coffee has also been reported [86]. Four coffee adulterants were studied, i.e., coffee husks, wood sticks, soil, and corn kernels. They were added to the roasted-ground coffee in levels of 1.0 to 40% (w/w). Quantification of adulterants was based on the utilization of the MCR correlation constraint, whereas detection was achieved through comparison of the reference spectrum of each adulterant with the spectra recovered by the MCR model utilizing only the adulterated samples. For detection, the values of the coefficient of correlation of the studied adulterants ranged from 0.90 to 0.99. However, the absolute errors for the quantification of adulterants in the developed models were below 4%. The findings pointed out the practicability in the application of the methodologies proposed. Flores-Valdez et al. [47] successfully employed FT-NIR spectroscopy coupled with chemometrics for identification and quantification of adulterants in coffee. Coffee samples were adulterated with soy, corn, barley, oat, coffee husks, and rice, in proportions ranging from 1 to 30% in increments of 1%. They developed a SIMCA model based on the percentage rejection (selectivity) and recognition (sensitivity), which allowed the discrimination of adulterated and unadulterated samples with an accuracy of 100%. The quantitative model with the highest performance corresponded to the PLS1 algorithm with an $R^2 \geq 0.99$, standard error of prediction (SEP) of 0.45–0.94, and standard error of calibration (SEC) of 0.39–0.82. To build this model, the most significant spectral range considered were 800–1800 and 2800–3500 cm^{-1} because they showed the highest variability of the spectra, signifying the change in absorbance as well as increase in the levels of adulterants in the samples. In another study, Brondi et al. [63] compared differential scanning calorimetry (DSC) and attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy methodologies together with chemometric analysis for detection and quantification of corn in roasted-ground coffee. Adulterated samples were prepared mixing coffee roasted to light, medium, and dark roasts with ground roast corn in levels ranging from 0.5 to 40% (m/m). PLS and PCA models were constructed to quantify adulteration levels and to distinguish non-adulterated samples from adulterated ones, respectively. PCA models built with FTIR and DSC data were able to detect adulteration of coffee with corn in concentrations below 1%. Conversely, PLS models demonstrated a good correlation between reference concentrations and estimated values with RMSEP of 1.53% for FTIR and 3.94% for DSC.

The presence of defective beans (i.e., insect damaged, broken, sour, black, mouldy, or immature) significantly affects the price and quality of the batch as well as the beverage. According to Franca et al. [87], defective beans roast differently, acquiring a different colour after roasting and imparting undesirable organoleptic characteristics to the beverage, which decreases cup quality. Therefore, it is vital to develop techniques for rapid and accurate assessment of coffee quality. Craig et al. [41] successfully developed a technique based on FTIR and NIR to quantify defects in roasted coffees. Quantitative models based on PLS regression were developed with the aim of predicting defective bean percentage in mixtures with non-defective beans. The mixtures were prepared by mixing each type of defective bean (immature, black, light sour, and dark sour) with non-defective ones, with the defect amounts ranging from 3 to 30% in steps of 3%. PLS regression models provided accurate predictive results with correlation coefficients and root mean square error of validation (RMSEV) of 0.891 and 0.032, respectively, for FTIR, and 0.953 and 0.026 for the NIR technique. Comparing the two methods, quantitative models from NIR were considerably more robust as compared to the ones based on FTIR for quantifying defective beans in roasted coffees. With the employment of a new statistical approach, Elastic net, the same authors evaluated the performance of NIR and ATR-FTIR for the discrimination of roasted defective (immature, black, dark sour, and light sour) and non-defective coffees. The developed models based on Elastic net allowed for the correct classification of non-defective and defective coffees with amino acids/proteins, carbohydrates, lipids, chlorogenic acids, and caffeine being the chief chemical descriptors that characterized the samples. Generally, both NIR and ATR-FTIR proved to be reliable methods for the discrimination of roasted defective and non-defective coffees.

In another study, Santos et al. [64] assessed the ability of NIR spectroscopy in discriminating defective beans (sour, immature, and black) from non-defective ones using PCA and PLSR analysis. Although PCA modelling did not allow the separation of all the studied defects, PLS regression linking the NIR spectrum to the mass fraction of non-defective and defective beans showed relative errors of about 5%. This is an indication that with the proposed methodology, it may be possible to do quantification of non-defective quality against other qualities in a batch. Chen et al. [65] studied the capability of VISIBLE-NEAR INFRARED (Vis-NIR) hyperspectral imaging for insect-damaged beans detection. The experimental samples included 570 and 569 defective and healthy beans, respectively. A Vis-NIR push-broom hyperspectral imaging camera was used to attain the beans images, a hyperspectral insect damage detection algorithm (HIDDA) was used for selection of bands, and a convolutional neural network (CNN) and constrained energy minimization-support vector machine (CEM-SVM) was used for identification. Classification accuracy of defective beans was 93% and that of healthy beans was 96.4%. It was observed that the wavelength range of 850–950 nm was important in accurately identifying insect damaged and healthy beans.

Application of diffuse reflectance (DR) and attenuated total reflectance (ATR) Fourier transform infrared spectroscopy for discrimination of mature and immature green coffees has also been studied. Agglomerative hierarchical clustering (AHC) and PCA analysis of the normalized ATR and DR spectra clearly separated the samples into two groups, mature and immature. Spectral ranges that significantly contributed to sample clustering were as follows 2074–3100, 1800–1995, 1477–800, and 1725–1762 cm^{-1} for DR spectra and 3030–3080 and 1533–1535 cm^{-1} for ATR spectra. Linear discriminant analysis (LDA) classification models based on the absorbance readings after normalization at four wavenumbers (1141, 1741, 2852, and 2921 cm^{-1}) were developed. The results provided by the models presented recognition as well as prediction capabilities of 100%, for both ATR and DR data [49]. The same research group also employed similar techniques with an aim of characterizing and discriminating between non-defective and defective (sour, immature, black) coffee beans prior to roasting. Similar observations were reported with PCA and AHC of the reflectance spectra showing discrimination between non-defective and defective beans [66].

Another study focused on the application of Fourier transform infrared-photoacoustic spectroscopy (FTIR-PAS) in a quantitative assessment of defects in roasted ground coffee [51]. Mixtures of different defects (broken, sour, black, peels, and woods) were mixed with non-defective coffee of robusta and arabica species in precise ratios to form different classes of blends. All samples were medium roasted and ground. With PCA, it was possible to predict the fraction/amount as well as the defects nature in blends. PLS-DA provided information on the similarities between the blends. The developed model classified samples into four classes with 100% accuracy and specificities of higher than 0.9.

3.4. Discrimination of Coffee Based on Species, Variety, and Geographical Origin

Coffee identification has become an important aspect of the global coffee market considering the great variability in quality and selling price based on its geographic origin, species, and variety. This has necessitated the development of fast and reliable methods for discrimination of coffee based on the aforementioned parameters to avoid fraudulence. Arabica coffees are usually considered to be of better quality than robusta coffees in terms of flavour and taste. NIR spectroscopy in combination with the direct orthogonal signal correction (DOSC) preprocessing technique was utilized to develop an analytical technique to discriminate between pure robusta and arabica species and their blends. Classification models from raw and corrected NIR spectra were developed using the potential function class-modelling technique. It was concluded that the combination of DOSC, NIR spectroscopy, and the potential functions method could be utilized as an ideal approach to distinguish between pure robusta and arabica coffee varieties and to differentiate between blends and pure varieties of the two species [72]. In another study, Zhang et al. [48] sought to establish the best pattern recognition method that could be

employed in conjunction with MIR spectroscopy for coffee variety identification. PCA successfully categorized each coffee variety with the first four principal components (PCs), explaining over 99% of the total variance. Twenty-nine best wavenumbers were selected by the loadings of the first four PCs and used in the classification models development utilizing ten different pattern recognition techniques. The extreme learning machine (ELM), radial basis function neural network (RBFNN), back propagation neural network (BPNN), relevance vector machine (RVM), SIMCA, and SVM models were categorized as highly effective techniques, with classification accuracies of more than 95% in the test and training set. In contrast, naive Bayes classifier and PLS-DA, *K*-nearest neighbors, and random forest were categorized as techniques of low effectiveness and medium effectiveness with classification accuracy below 80% and above 80% in the test and training, respectively.

Marquetti et al. [37] evaluated the potential of NIR spectroscopy and PLS-DA in distinguishing green arabica coffee samples based on their geographical origin and genotypic characteristics. Coffee samples of four genotypes (IA 59, IPR 106, IPR 99, and IPR 105) and from four different cities were studied. The best PLS-DA models were developed from the pretreated spectra (Savitzky–Golay second derivative and multiplicative scatter correction) and were capable of distinguishing coffee samples both genotypically and geographically. However, the specificity and sensitivity for prediction and calibration sets for geographic discrimination were greater than for the genotypic one. The best model for both parameters correctly identified 94.4% of validation samples. Similarly, Giraudo et al. [36] used NIR spectroscopy and multi-variate data analysis to classify coffee cultivated in different countries and continents. Interval PLS-DA and laboratory-independent partial least square-discriminant analysis models were developed by following a ranked method, i.e., by first putting into consideration the continent followed by the country of origin as the rule of discrimination. For the continent and country of origin, the best classification models correctly identified over 98% and 100% of validation samples, respectively.

Medina et al. [73] compared the capability of nuclear magnetic resonance (^{1}H-NMR), NIR, and ATR-MIR spectroscopic methods in discriminating coffee based on its species (arabica and robusta) and geographic origin (Colombian coffee versus other origins). Considering species, samples were successfully discriminated as expected by all the techniques using the built PLS classification models. However, for the case of origin determination, ATR-MIR and ^{1}H-NMR exhibited comparable capability to discriminate Colombian coffee samples, but weak results were observed with the NIR spectroscopy technique. In contrast, Bona et al. [52] evaluated the ability of NIR and FTIR (MIR) spectroscopy in conjunction with support vector machines (SVM) for geographical classification of different arabica coffee genotypes and reported better results for the NIR method. The technique exhibited greater performance as compared to FTIR, with a specificity and sensitivity of 100% in discriminating coffee samples from different regions. Okubo and Kurata [34] proposed an approach for geographical classification of different species of coffee (arabica and robusta) exploiting NIR spectroscopy and SIMCA as the classification methods. Models were developed and their distance from SIMCA evaluated to establish if they were different. Although some samples were partially classified into several categories, a good classification result was obtained with over 73% correct classification rate. Model distance values from SIMCA were relatively similar among samples, which could explain the misclassification.

Wongsaipun et al. [5] applied an artificial neural network technique (a self-organizing map (SOM)) with the aim of tracing geographical origin of arabica coffee beans grown in three provinces in northern Thailand using NIR spectroscopy. The three provinces included Chiang Mai, Lampang, and Mae Hong Son. The self-organizing map discrimination index (SOMDI) was applied to identify important parameters of the spectroscopic data. SOM analysis of the NIR spectra separated the samples from the three different regions. Based on the SOMDI results, the coffee samples from Chiang Mai could be well discriminated using the following NIR spectral ranges: 2260–2498, 1896–2180, 1254–1326, and 880–1182 nm. Luna et al. [74] evaluated different chemometric methods for the classification of robusta coffee cultivars using FT-NIR spectroscopy. Five coffee cultivars were studied. SOM,

SIMCA, PLS-DA, and SVM were explored for the cultivars classification, whereas PCA was used for identification of the clusters. PCA analysis of the pre-processed spectra discriminated samples into different groups, which were correlated to the presence of water, caffeine, lipids, sugars, chlorogenic acids, carbohydrates, and proteins. SOM presented the best results, providing 100% correct identification of the validation samples, whereas SIMCA, PLS-DA, SVM using 4 PCs and SVM using 3 PCs provided 99.6, 82.9, 99.6, and 82.9%, respectively.

A study by Link et al. [50] distinguished coffee samples of different geographical origins and genetic make-up using FTIR spectroscopy in tandem with radial-basis function networks (RBFs). Optimization of some parameters was done using sequential simplex in order to select the best neural network. The optimized RBFs correctly classified the samples both genotypically (94.44%) and geographically (100%). This chemometric method exhibited superior performance compared to multilayer perceptron (MLP) and SIMCA developed for classification of coffee as reported by the same research group. A feasibility study using ATR-FTIR spectroscopy and chemometric analysis with the aim of discriminating coffees from different topographical origins as well as of different roast degrees was conducted by Wang et al. [75]. Coffee samples from different origins (Kenya, Colombia, Ethiopia, and Costa Rica) were roasted to two roast degrees (medium and dark) and extracted using six organic solvents (hexane, ethyl acetate, dichloromethane, acetic acid, ethanol, and acetone) and a mixture of an equal volume of the solvent and water. Extraction using water and the solvent resulted in clean extracts, which provided good spectral information important for sample discrimination. Developed classification models based on SIMCA and of dark roasted coffee correctly classified samples (100%) based on their origin when ethyl acetate solvent was used.

The potential of Raman spectroscopy in coffee species and cultivar discrimination has been investigated. A study by Rubayiza and Meurens [53] discriminated arabica, liberica, and robusta coffees based on their lipid fraction using Fourier transform (FT) Raman spectroscopy. Raman measurements of the lipid fraction from roasted and green coffee beans was taken using a laser emissions at a wavelength of 1064 nm and resolution of 4 cm^{-1}. PCA analysis of the spectra revealed clustering of the samples based on the species with the first principal component (PC1), explaining 93% of the spectral variance. This was associated with the concentration of kahweol, which manifests at two specific scattering bands at 1478 and 1567 cm^{-1}. Keidel et al. [54] reported similar findings after analyzing green coffee beans belonging to robusta and arabica species from different geographical origins (South America, Asia, and Africa). Measurements were taken using a 1064 nm laser and a spectral resolution of 4 cm^{-1}. It was possible to classify coffee samples based on the specific kahweol content. In a similar approach, Wermelinger et al. [55] evaluated Raman spectroscopy for quantification of robusta coffee fraction in blends by analyzing the lipid fraction. Pure robusta and arabica and their mixtures with contents of robusta of 5, 10, 25, 33, 50, and 75 wt.% were used. Lipids were extracted from roasted coffee beans and their Raman spectra collected at 532 nm. It was possible to determine the robusta content in studied mixtures by examining the intensity ratio between the Raman peaks of fatty acids at 1665 or 1460 cm^{-1} and the kahweol peak at 1570 cm^{-1}.

Luna et al. [11] compared different chemometric approaches that could be used to classify robusta coffee cultivars using Raman spectroscopy. Raman spectra of five coffee cultivars were collected at 785 nm with a laser excitation power of 100 mW. The following classification methods were studied: quadratic discriminant analysis (QDA), mixture discriminant analysis (MDA), PLS–DA with Bayesian inference, SIMCA, regularized discriminant analysis (RDA), and LDA. Two preprocessing methods, multiplicative scatter correction (MSC) and mean centering (MC), were also compared, with MSC yielding more accurate results for all the classification methods. Using multiplicative scatter correction, RDA, MDA, QDA, SIMCA, and PLS–DA methods correctly classified 100% of the samples, whereas LDA correctly classified 98.7% of the samples. Conversely, using MC, correct classification of the samples was 70.7%, 62.7%, 62.7%, 62.7%, 61.3%, and 97.3% for MDA,

LDA, RDA, QDA, PLS-DA, and SIMCA, respectively. A study by Figueiredo et al. [56] differentiated green coffee beans of different genotypes using Raman spectroscopy in conjunction with chemometric methods (PLS-DA and PCA). Acquisition of the Raman spectra were made using an excitation laser with λ = 1064 nm at a spectral resolution of 4 cm^{-1}. It was possible to distinguish the coffee genotypes based on spectral bands that are typical of kahweol (1567 and 1479 cm^{-1}) and fatty acids (1442 and 1302 cm^{-1}) using PCA. The developed PLS-DA models demonstrated a good association between the training and test set of the studied samples indicating its suitability in performing the task.

El-Abassy et al. [14] proposed a method capable of discriminating between robusta and arabica coffee species based on their lipid and chlorogenic acid (CGA) contents using Raman spectroscopy. Measurements of the samples were made using a 514.5 nm laser with an excitation power of 10 mW. PCA analysis of the whole Raman spectra presented a clear discrimination between robusta and arabica coffee with 93% of the total spectral variation. The most important spectral range that contributed to this observation was 1000–1750 cm^{-1}, an area dominated with Raman bands of CGA. Discrimination of the two coffee species was also observed when PCA analysis was restricted to the spectral range between 2700–3050 cm^{-1}, which is dominated by bands of lipids. In this respect, the first two PCs constituted 85% of explained variation in the spectra. Studies on fluorescence spectroscopy as a valuable technique to discriminate coffee based on their geographical origin and phenotypic characteristics are also present in the literature. Botelho et al. [13] developed a supervised classification technique using chemometrics tools with the ability to discriminate coffee of different origins (four) using fluorescence spectroscopy. Discrimination models were constructed by employing N-way partial least squares (NPLS-DA), parallel factor analysis (PARAFAC), and unfolded partial least squares (UPLS-DA) methods. Among the three methods studied, UPLS-DA models showed the best results, with f-scores of coffees from two regions above 0.8, for both test and training sets. In another study, Dankowska et al. [76] compared the ability of UV-Vis spectroscopies, fluorescence, and the low- and mid-level data fusion of both spectroscopies to quantify concentrations of roasted robusta and arabica in coffee blends. Classification models were developed using multiple linear regression (MLR) and LDA. The best prediction capability of the MLR models was achieved with application of the mid-level data fusion model of fluorescence and UV-Vis intensities at 60 nm wavelength interval with RMSEV of 7.9% and RMSEC of 3.6%. Better discrimination ability was obtained with data fusion with the highest classification accuracy (above 96.0%) being achieved for the low-level LDA model with fluorescence intensities at a 60 nm wavelength interval.

3.5. Prediction of Coffee Chemical Composition

Coffees chemical composition affects the quality of the beverage, making its analysis a necessity for coffee processors. The use of known conventional methods to determine these compounds in coffee is laborious and expensive due to sample preparation. Spectroscopic analysis of ground or whole coffee beans is a potential alternative to overcome these limitations. Studies on the application of NIR spectroscopy for the quantification of the chemical components of green coffee beans have been done (Table 2). Tugnolo et al. [30] compared the potential of NIR spectroscopy and thermogravimetric analysis in measuring moisture content of roasted beans and their grounds. The developed PLS models from NIR data demonstrated a good relationship between estimated values and reference concentrations with errors in prediction below 0.15% and coefficients of determination higher than 0.95. To compare the measurements of the two techniques, the passing-Bablok regression method was performed. Even though there were significant differences, the proposed residual dispersion index (RDI%) showed higher predictive accuracy of NIR-based predictions (RDI% = 5.93) with regard to thermogravimetric analysis measurements (RDI% = 9.68). Escobar et al. [77] developed a rapid method for moisture content prediction in unroasted coffee beans using NIR spectroscopy. Among the samples analyzed (two accessions of liberica coffee and four varieties of arabica coffee), the calibration model for typica coffee

(arabica variety) exhibited the best performance in predicting the moisture content of studied samples as evidenced by its high cross-validated coefficient of determination (0.8556) and cross-validated root mean square error of prediction (0.6355).

Kyaw et al. [78] reported promising results in relation to the moisture content prediction of ground unroasted coffee beans using NIR spectroscopy. Good accuracy for the prediction of moisture content was obtained from the spectral data pretreated with second derivative and Kubelka-Munk (K/S) data transformation (correlation coefficient (r) = 0.87 and accuracy = 99%). Zhu et al. [35] proposed a rapid method for lipid and protein content determination in unroasted coffee beans using NIR spectroscopy and chemometrics. Orthogonal signal correction (OSC) and different spectral pretreatment methods (standard normal variate (SNV), MSC, 1st or 2nd derivative, and Savitzky-Golay smoothing) were compared during the process of building a PLS regression model. Model quality was enhanced by the MSC, SNV, and OSC pretreatment methods. On the contrary, the 1st and 2nd derivative reduced the model quality. Important variable selection significantly enhanced the PLS models, with OSC-PLS models standing out as the most robust for prediction of lipids and proteins with an RMSEP of less than 0.106 and a coefficient of determination for prediction of more than 0.982. In a similar approach, Macedo et al. [79] proposed a methodology based on the use of NIR spectroscopy associated with PLS regression to estimate some chemical properties (soluble solids, total and reducing sugars, and moisture content) in intact green coffee samples. The highest R^2 obtained for the samples in the validation set included 0.781, 0.810, 0.694, and 0.516 for reducing sugars, moisture content, total sugar, and soluble solids, respectively.

In another study, NIR spectroscopy coupled to chemometrics (PLS regression) was investigated to predict 5-caffeoylquinic acid (5-CQA), trigonelline, and caffeine content in green coffee beans. The prediction models for 5-CQA, trigonelline, and caffeine were constructed using 5, 6, and 7 latent variables, giving RMSEPs of 0.27, 0.07, and 0.08 and correlation coefficients of cross validation (R_{cv}) of 0.96, 0.96, and 0.98, respectively [39]. An analytical technique based on NIR spectroscopy for prediction of caffeine, proteins, lipids, CGA, phenolic compounds, sucrose, and total sugars was investigated in grounds of green coffee beans [42]. A modified PLS regression was employed in the development of prediction models. Prediction models for lipids (ratio of performance deviation (RPD) = 2.77, R^2 = 0.87), phenolic compounds (RPD = 2.62, R^2 = 0.86), total sugars (RPD = 2.55, R^2 = 0.85), and sucrose (RPD = 2.44, R^2 = 0.84) were not very accurate. Models for proteins (RPD = 4.09, R^2 = 0.94), caffeine (RPD = 4.16, R^2 = 0.92), and chlorogenic acids (RPD = 4.16, R^2 = 0.94), presented the best prediction capabilities. Likewise, Safrizal et al. [9] reported calibration models for the prediction of caffeine, lipid, and CGA content with an r-value and RPD above 0.7 and 2, respectively.

The literature also contains investigations on the feasibility of FTIR/MIR spectroscopy in the prediction of chemical components in coffee. Liang et al. [80] used ATR-FTIR spectroscopy together with PLS regression with the aim of developing a rapid method to determine antioxidant activity and CGA isomer composition in roasted and unroasted coffee beans. FTIR spectral data were processed (i.e., baseline corrected and mean centered) prior to model development. For the case of CGA isomer concentrations, the models presented a good relationship between HPLC reference values and estimated values with R_{cv} of more than 0.92 for all the isomers. Good models were also constructed with the ability to predict antioxidant capacities in coffee beans. In another study, Weldegebreal et al. [81] successfully developed rapid methods for direct caffeine content determination in aqueous solution of unroasted coffee beans using fluorescence, NIR, and ATR-FTIR spectroscopy. Caffeine content of the samples was determined using the three types of equipment and the results were compared to the standard method (UV-Vis spectroscopy) for validation. The results for the newly developed techniques were similar to the results achieved by the standard method. In a similar approach, Yisak et al. [82] developed a reliable and sensitive technique for simultaneous trigonelline and caffeine determination in the aqueous extract of unroasted coffee beans with relative standard deviations below 4%. Evaluation of the

developed analytical method was done by spiking green coffee beans with trigonelline and caffeine standards. Average recoveries of 99 ± 2% for both the alkaloids was reported.

4. Future Perspectives and Conclusions

The information presented in this review clearly confirms the potential of spectroscopic techniques in coffee assessment. However, most of these studies are purely academic. Mendes and Duarte [23] highlighted issues that must be addressed for the developed spectroscopic analytical methods to be applied as official methods of analysis in the coffee industry. They include guidelines on the validation of the methods, use of samples from credible sources (known origin) for method development, use of a satisfactory number of representative samples in order to cover all sample variances for development of a robust model, and choosing the most appropriate chemometric technique for a particular objective. The findings of this review demonstrate the potential of spectroscopic methods to evaluate coffee for different parameters. However, there are other aspects that need further investigation, including prediction of coffees processed using different methods, differentiation of organic from inorganic coffee, and monitoring of coffee quality during storage, of which limited scientific information is available. Additionally, Raman and fluorescence spectroscopy received very little attention, with most applications being focused on NIR spectroscopy. Available information in the field of coffee spectroscopy primarily focused on IR spectroscopy. To broaden this scope, this review includes recent studies based on IR, Raman, and fluorescence spectroscopy.

In conclusion, coffee composition and quality assessment are an important practice in the global coffee trade market. Conventional methods that are most frequently used to perform these tasks are often time-consuming, complex, expensive, and require several steps in sample preparation. Spectroscopic techniques (NIR, MIR/FTIR, Raman, and fluorescence), as evidenced by the reviewed literature, could be used to evaluate coffee for different parameters along the production process and in the near future replace the conventional methods. The vast majority of applications are based on NIR spectroscopy; only a few examples with Raman and fluorescence spectroscopy are available in the literature. In the future, applications using these methods will likely become more common, as valuable information can also be obtained from these spectra. The major advantages of these techniques are that they are rapid, require no or simple sample preparation, are non-destructive, and can be easily integrated into processes. Nevertheless, it is important for laboratory studies to upscale to industrial applications if these advantages are to be realized.

Author Contributions: Conceptualization, writing—review and editing, B.H.; writing—original draft preparation, L.M.; writing—review and editing, D.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Academic Exchange Service (DAAD), funding ID 57524989. However, there was no funding of the publication costs.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. ICO Trade Statistcis. Available online: http://www.ico.org/trade_statistics.asp?section=Statistics (accessed on 31 August 2021).
2. Belitz, H.-D.; Grosch, W.; Schieberle, P. Coffee, tea, cocoa. In *Food Chemistry*, 4th ed.; Belitz, H.-D., Grosch, W., Schieberle, P., Eds.; Springer: Berlin, Germany, 2009; pp. 938–951.
3. Yashin, A.; Yashin, Y.; Xia, X.; Nemzer, B. Chromatographic Methods for Coffee Analysis: A Review. *J. Food Res.* **2017**, *6*, 60. [CrossRef]

4. Alonso-Salces, R.M.; Serra, F.; Reniero, F.; Héberger, K. Coffea arabica and Coffea canephora): Chemometric evaluation of phenolic and methylxanthine contents. *J. Agric. Food Chem.* **2009**, *57*, 4224–4235. [CrossRef] [PubMed]
5. Wongsaipun, S.; Theanjumpol, P.; Muenmanee, N.; Boonyakiat, D.; Funsueb, S.; Kittiwachana, S. Application of Artificial Neural Network for Tracing the Geographical Origins of Coffee Bean in Northern Areas of Thailand Using Near Infrared Spectroscopy. *Chiang Mai J. Sci.* **2021**, *48*, 163–175.
6. Yergenson, N.; Aston, D.E. Monitoring coffee roasting cracks and predicting with in situ near-infrared spectroscopy. *J. Food Process Eng.* **2020**, *43*, 13305. [CrossRef]
7. Yergenson, N.; Aston, D.E. Online determination of coffee roast degree toward controlling acidity. *J. Near Infrared Spectrosc.* **2020**, *28*, 175–185. [CrossRef]
8. Belchior, V.; Botelho, B.G.; Casal, S.; Oliveira, L.S.; Franca, A.S. FTIR and Chemometrics as Effective Tools in Predicting the Quality of Specialty Coffees. *Food Anal. Methods* **2020**, *13*, 275–283. [CrossRef]
9. Safrizal; Sutrisno; Pujantoro, L.E.N.; Ahmad, U. Samsudin Predicting Lipid, Caffeine and Chlorogenic Acid Contents of Arabica Coffee Using NIRS. *Int. J. Sci. Technol. Res.* **2019**, *8*, 367–370.
10. Reis, N.; Botelho, B.G.; Franca, A.S.; Oliveira, L. Simultaneous Detection of Multiple Adulterants in Ground Roasted Coffee by ATR-FTIR Spectroscopy and Data Fusion. *Food Anal. Methods* **2017**, *10*, 2700–2709. [CrossRef]
11. Luna, A.S.; DaSilva, A.P.; DaSilva, C.S.; Lima, I.C.A.; DeGois, J.S. Chemometric methods for classification of clonal varieties of green coffee using Raman spectroscopy and direct sample analysis. *J. Food Compos. Anal.* **2019**, *76*, 44–50. [CrossRef]
12. Suhandy, D.; Yulia, M. Discrimination of several Indonesian specialty coffees using Fluorescence Spectroscopy combined with SIMCA method. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *334*, 012059. [CrossRef]
13. Botelho, B.G.; Oliveira, L.S.; Franca, A.S. Fluorescence spectroscopy as tool for the geographical discrimination of coffees produced in different regions of Minas Gerais State in Brazil. *Food Control* **2017**, *77*, 25–31. [CrossRef]
14. El-Abassy, R.M.; Donfack, P.; Materny, A. Discrimination between Arabica and Robusta green coffee using visible micro Raman spectroscopy and chemometric analysis. *Food Chem.* **2011**, *126*, 1443–1448. [CrossRef]
15. Zettel, V.; Ahmad, M.H.; Beltramo, T.; Hermannseder, B.; Hitzemann, A.; Nache, M.; Paquet-Durand, O.; Schöck, T.; Hecker, F.; Hitzmann, B. Supervision of food manufacturing processes using optical process analyzers—An overview. *ChemBioEng Rev.* **2016**, *3*, 219–228. [CrossRef]
16. Ahmad, M.H.; Sahar, A.; Hitzmann, B. Fluorescence spectroscopy for the monitoring of food processes. *Adv. Biochem. Eng. Biotechnol.* **2017**, *161*, 121–152.
17. Nawrocka, A.; Lamorska, J. Determination of Food Quality by Using Spectroscopic Methods. In *Advances in Agrophysical Research*; Grundas, S., Stepniewski, A., Eds.; Intechopen: London, UK, 2013; pp. 348–363.
18. Lin, M.; Rasco, B.A.; Cavinato, A.G.; Al-Holy, M. Infrared (IR) Spectroscopy—Near Infrared Spectroscopy and Mid-Infrared Spectroscopy. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Sun, D.-W., Ed.; Academic Press: London, UK, 2009; pp. 119–143.
19. Dufour, E. Principles of Infrared spectroscopy. In *Infrared Spectroscopy for Food Quality Analysis and Control*; Sun, C., Ed.; Academic Press: London, UK, 2009; pp. 1–27.
20. Prieto, N.; Pawluczyk, O.; Dugan, M.E.R.; Aalhus, J.L. A Review of the Principles and Applications of Near-Infrared Spectroscopy to Characterize Meat, Fat, and Meat Products. *Appl. Spectrosc.* **2017**, *71*, 1403–1426. [CrossRef]
21. Reich, G. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv. Drug Deliv. Rev.* **2005**, *57*, 1109–1143. [CrossRef]
22. Nelson, D.L. *Spectroscopic Methods in Food Analysis*, 1st ed.; Franca, A.S., Nollet, L., Eds.; CRC Press: Boca Raton, FL, USA, 2018.
23. Mendes, E.; Duarte, N. Mid-Infrared Spectroscopy as a Valuable Tool to Tackle Food Analysis: A Literature Review on Coffee, Dairies, Honey, Olive Oil and Wine. *Foods* **2021**, *10*, 477. [CrossRef]
24. Stehle, C.U.; Abuillan, W.; Gompf, B.; Dressel, M. Far-infrared spectroscopy on free-standing protein films under defined temperature and hydration control. *J. Chem. Phys.* **2012**, *136*, 075102. [CrossRef] [PubMed]
25. Wang, Q. Raman Spectroscopic Characterization and Analysis of Agricultural and Biological Systems. 2013. Available online: https://dr.lib.iastate.edu/handle/20.500.12876/27208 (accessed on 12 October 2021).
26. McCreery, R.L. *Calibration and Validation in Raman Spectroscopy for Chemical Analysis*; Wiley-Interscience: New York, NY, USA, 2000.
27. Petersen, M.; Yu, Z.; Lu, X. Application of Raman Spectroscopic Methods in Food Safety: A Review. *Biosensors* **2021**, *11*, 187. [CrossRef] [PubMed]
28. Boyaci, I.H.; Temiz, H.T.; Genis, H.E.; Soykut, E.A.; Yazgan, N.N.; Guven, B.; Uysal, R.S.; Bozkurt, A.G.; Ilaslan, K.; Toruna, O.; et al. Dispersive and FT-Raman spectroscopic methods in food analysis. *R. Soc. Chem.* **2015**, *5*, 56606–56624. [CrossRef]
29. Khuwijitjaru, P.; Boonyapisomparn, K.; Huck, C.W. Near-infrared spectroscopy with linear discriminant analysis for green 'Robusta' coffee bean sorting. *Int. Food Res. J.* **2020**, *27*, 287–294.
30. Tugnolo, A.; Giovenzana, V.; Malegori, C.; Oliveri, P.; Casson, A.; Curatitoli, M.; Guidetti, R.; Beghi, R. A reliable tool based on near-infrared spectroscopy for the monitoring of moisture content in roasted and ground coffee: A comparative study with thermogravimetric analysis. *Food Control* **2021**, *130*, 108312. [CrossRef]
31. Santos, J.R.; Viegas, O.; Páscoa, R.N.M.J.; Ferreira, I.M.P.L.V.O.; Rangel, A.O.S.S.; Lopes, J.A. In-line monitoring of the coffee roasting process with near infrared spectroscopy: Measurement of sucrose and colour. *Food Chem.* **2016**, *208*, 103–110. [CrossRef]

32. Correia, R.M.; Tosato, F.; Domingos, E.; Rodrigues, R.R.T.; Aquino, L.F.M.; Filgueiras, P.R.; Lacerda, V.; Romão, W. Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta* **2018**, *176*, 59–68. [CrossRef] [PubMed]
33. DeAraújo, T.K.L.; Nobrega, R.O.; Fernandes, D.D.S.; DeAraújo, M.C.U.; Diniz, P.H.G.D.; DaSilva, E.C. Non-destructive authentication of Gourmet ground roasted coffees using NIR spectroscopy and digital images. *Food Chem.* **2021**, *364*, 130452. [CrossRef]
34. Okubo, N.; Kurata, Y. Nondestructive Classification Analysis of Green Coffee Beans by Using Near-Infrared Spectroscopy. *Foods* **2019**, *8*, 82. [CrossRef]
35. Zhu, M.; Long, Y.; Chen, Y.; Huang, Y.; Tang, L.; Gan, B.; Yu, Q.; Xie, J. Fast determination of lipid and protein content in green coffee beans from different origins using NIR spectroscopy and chemometrics. *J. Food Compos. Anal.* **2021**, *102*, 104055. [CrossRef]
36. Giraudo, A.; Grassi, S.; Savorani, F.; Gavoci, G.; Casiraghi, E.; Geobaldo, F. Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. *Food Control* **2019**, *99*, 137–145. [CrossRef]
37. Marquetti, I.; Link, J.V.; Lemes, A.L.G.; Scholz, M.B.S.; Valderrama, P.; Bona, E. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. *Comput. Electron. Agric.* **2016**, *121*, 313–319. [CrossRef]
38. Santos, J.R.; Lopo, M.; Rangel, A.O.S.S.; Lopes, J.A. Exploiting near infrared spectroscopy as an analytical tool for on-line monitoring of acidity during coffee roasting. *Food Control* **2016**, *60*, 408–415. [CrossRef]
39. Ribeiro, J.S.; Salva, T.J.G.; Silvarolla, M.B. Prediction of a wide range of compounds concentration in raw coffee beans using NIRS, PLS and variable selection. *Food Control* **2021**, *125*, 107967. [CrossRef]
40. Ribeiro, J.S.; Ferreira, M.M.C.; Salva, T.J.G. Chemometric models for the quantitative descriptive sensory analysis of arabica coffee beverages using near infrared spectroscopy. *Talanta* **2011**, *83*, 1352–1358. [CrossRef]
41. Craig, A.P.; Franca, A.S.; Oliveira, S.L.; Irudayaraj, J.; Ileleji, K. Fourier transform infrared spectroscopy and near infrared spectroscopy for the quantification of defects in roasted coffees. *Talanta* **2015**, *134*, 379–386. [CrossRef]
42. Scholz, M.B.S.; Kitzberger, C.S.G.; Pereira, L.F.P.; Davrieux, F.; Pot, D.; Charmetantd, P.; Leroy, T. Application of near infrared spectroscopy for green coffee biochemical phenotyping. *J. Near Infrared Spectrosc.* **2014**, *22*, 411–421. [CrossRef]
43. Alessandrini, L.; Romani, S.; Pinnavaia, G.; Rosa, M.D. Near infrared spectroscopy: An analytical tool to predict coffee roasting degree. *Anal. Chim. Acta* **2008**, *625*, 95–102. [CrossRef] [PubMed]
44. Abreu, M.B.; Marcheafave, G.G.; Bruns, R.E.; Scarminio, I.S.; Zeraik, M.L. Spectroscopic and Chromatographic Fingerprints for Discrimination of Specialty and Traditional Coffees by Integrated Chemometric Methods. *Food Anal. Methods* **2020**, *13*, 2204–2212. [CrossRef]
45. Craig, A.P.; Botelho, B.G.; Oliveira, L.S.; Franca, A.S. Mid infrared spectroscopy and chemometrics as tools for the classification of roasted coffees by cup quality. *Food Chem.* **2018**, *245*, 1052–1061. [CrossRef] [PubMed]
46. Fioresi, D.B.; Pereira, L.L.; Oliveira, E.C.S.; Moreira, T.R.; Ramos, A.C. Mid infrared spectroscopy for comparative analysis of fermented arabica and robusta coffee. *Food Control* **2021**, *121*, 107625. [CrossRef]
47. Flores-Valdez, M.; Meza-Márquez, O.G.; Osorio-Revilla, G.; Gallardo-Velázquez, T. Identification and Quantification of Adulterants in Coffee (*Coffea arabica* L.) Using FT-MIR Spectroscopy Coupled with Chemometrics. *Foods* **2020**, *9*, 851. [CrossRef]
48. Zhang, C.; Wang, C.; Liu, F.; He, Y. Mid-Infrared Spectroscopy for Coffee Variety Identification: Comparison of Pattern Recognition Methods. *J. Spectrosc.* **2016**, *2016*, 7927286. [CrossRef]
49. Craig, A.P.; Franca, A.S.; Oliveira, L.S. Discrimination between Immature and Mature Green Coffees by Attenuated Total Reflectance and Diffuse Reflectance Fourier Transform Infrared Spectroscopy. *J. Food Sci.* **2011**, *76*, 2011. [CrossRef]
50. Link, J.V.; Lemes, A.L.G.; Marquetti, I.; Scholz, M.B.S.; Bona, E. Geographical and genotypic classification of arabica coffee using Fourier transform infrared spectroscopy and radial-basis function networks. *Chemom. Intell. Lab. Syst.* **2014**, *135*, 150–156. [CrossRef]
51. Dias, R.C.E.; Valderrama, P.; Março, P.H.; Scholz, M.B.S.; Edelmann, M.; Yeretzian, C. Quantitative assessment of specific defects in roasted ground coffee via infrared-photoacoustic spectroscopy. *Food Chem.* **2018**, *255*, 132–138. [CrossRef]
52. Bona, E.; Marquetti, I.; Link, J.V.; Makimori, G.Y.F.; da Costa Arca, V.; Lemes, A.L.G.; Ferreira, J.M.G.; dos Santos Scholz, M.B.; Valderrama, P.; Poppi, R.J. Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. *LWT–Food Sci. Technol.* **2017**, *76*, 330–336. [CrossRef]
53. Rubayiza, A.B.; Meurens, M. Chemical Discrimination of Arabica and Robusta Coffees by Fourier Transform Raman Spectroscopy. *J. Agric. Food Chem.* **2005**, *53*, 4654–4659. [CrossRef]
54. Keidel, A.; Stetten, D.; Rodrigues, C.; Guas, C.; Hildebrandt, P. Discrimination of Green Arabica and Robusta Coffee Beans by Raman Spectroscopy. *J. Agric. Food Chem.* **2010**, *58*, 11187–11192. [CrossRef]
55. Wermelinger, T.; D'Ambrosio, L.; Klopprogge, B.; Yeretzian, C. Quantification of the Robusta Fraction in a Coffee Blend via Raman Spectroscopy: Proof of Principle. *J. Agric. Food Chem.* **2011**, *59*, 9074–9079. [CrossRef]
56. Figueiredo, L.P.; Borém, F.M.; Almeida, M.R.; Oliveira, L.F.C.; Oliveira, L.F.C.; Alves, A.P.C.; Santos, C.M.; Rios, P.A. Raman spectroscopy for the differentiation of Arabic coffee genotypes. *Food Chem.* **2019**, *288*, 262–267. [CrossRef]
57. Sunarharum, W.B.; Williams, D.J.; Smyth, H.E. Complexity of coffee flavor: A compositional and sensory perspective. *Food Res. Int.* **2014**, *62*, 315–325. [CrossRef]
58. Bertone, E.; Venturello, A.; Giraudo, A.; Pellegrino, G.; Geobaldo, F.G. Simultaneous determination by NIR spectroscopy of the roasting degree and Arabica/Robusta ratio in roasted and ground coffee. *Food Control* **2016**, *59*, 683–689. [CrossRef]

59. Pires, F.C.; Pereira, R.G.F.A.; Baqueta, M.R.; Valderrama, P.; Rocha, R.A. Near-infrared spectroscopy and multivariate calibration as an alternative to the Agtron to predict roasting degrees in coffee beans and ground coffees. *Food Chem.* **2021**, *365*, 130471. [CrossRef] [PubMed]
60. Catelani, T.A.; Santos, J.R.; Páscoa, R.N.M.J.; Pezza, L.; Pezza, H.R.; Lopes, J.A. Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study. *Talanta* **2018**, *179*, 292–299. [CrossRef]
61. Ebrahimi-Najafabadi, H.; Leardi, R.; Oliveri, P.; Casolino, M.C.; Jalali-Heravi, M.; Lanteri, S. Detection of addition of barley to coffee using near infrared spectroscopy and chemometric techniques. *Talanta* **2012**, *99*, 175–179. [CrossRef] [PubMed]
62. Winkler-Moser, J.K.; Singh, M.; Rennick, K.A.; Bakota, E.L.; Jham, G.; Liu, S.X.; Vaughn, S.F. Detection of Corn Adulteration in Brazilian Coffee (Coffea arabica) by Tocopherol Profiling and Near-Infrared (NIR) Spectroscopy. *J. Agric. Food Chem.* **2015**, *63*, 10662–10668. [CrossRef]
63. Brondi, A.M.; Torres, C.; Garcia, J.S.; Trevisan, M. Differential scanning calorimetry and infrared spectroscopy combined with chemometric analysis to the determination of coffee adulteration by corn. *J. Braz. Chem. Soc.* **2017**, *28*, 1308–1314. [CrossRef]
64. Santos, J.R.; Sarraguça, M.C.; Rangel, A.O.S.S.; Lopes, J.A. Evaluation of green coffee beans quality using near infrared spectroscopy: A quantitative approach. *Food Chem.* **2012**, *135*, 1828–1835. [CrossRef]
65. Chen, S.; Chang, C.; Ou, C.; Lien, C. Detection of Insect Damage in Green Coffee Beans Using VIS-NIR Hyperspectral Imaging. *Remote Sens.* **2020**, *12*, 2348. [CrossRef]
66. Craig, A.P.; Franca, A.S.; Oliveira, L.S. Evaluation of the potential of FTIR and chemometrics for separation between defective and non-defective coffees. *Food Chem.* **2012**, *132*, 1368–1374. [CrossRef]
67. Baqueta, M.R.; Coqueiro, A.; Valderrama, P. Brazilian coffee blends: A simple and fast method by near-infrared spectroscopy for the determination of the sensory attributes elicited in professional coffee cupping. *J. Food Sci.* **2019**, *84*, 1247–1255. [CrossRef]
68. Chang, Y.T.; Hsueh, M.C.; Hung, S.P.; Lu, J.M.; Peng, J.H.; Chen, S.F. Prediction of specialty coffee flavors based on near-infrared spectra using machine and deep-learning methods. *J. Sci. Food Agric.* **2021**, *101*, 4705–4714. [CrossRef]
69. Tolessa, K.; Rademaker, M.; Baets, B.D.; Boeckx, P. Prediction of specialty coffee cup quality based on near infrared spectra of green coffee beans. *Talanta* **2016**, *150*, 367–374. [CrossRef] [PubMed]
70. Arboleda, E.R. Discrimination of civet coffee using near infrared spectroscopy and artificial neural network. *Int. J. Adv. Comput. Res.* **2018**, *8*, 324–334. [CrossRef]
71. Belchior, V.; Botelho, B.G.; Oliveira, L.S.; Franca, A.S. Attenuated Total Reflectance Fourier Transform Spectroscopy (ATR-FTIR) and chemometrics for discrimination of espresso coffees with different sensory characteristics. *Food Chem.* **2019**, *273*, 178–185. [CrossRef]
72. Esteban-Díez, I.; González-Sáiz, J.M.; Saenz-Gonzalez, J.M.; Pizarro, C.I. Coffee varietal differentiation based on near infrared spectroscopy. *Talanta* **2007**, *71*, 221–229. [CrossRef] [PubMed]
73. Medina, J.; Caro Rodríguez, D.; Arana, V.A.; Bernal, A.; Esseiva, P.; Wist, J. Comparison of Attenuated Total Reflectance MidInfrared, Near Infrared, and 1H-Nuclear Magnetic Resonance Spectroscopies for the Determination of Coffee's Geographical Origin. *Int. J. Anal. Chem.* **2017**, *2017*, 7210463. [CrossRef]
74. Luna, A.S.; DaSilva, A.P.; Alves, E.A.; Rocha, R.B.; Lima, I.C.A.; DeGois, J.S. Evaluation of chemometric methodologies for the classification of Coffea canephora cultivars via FTNIR spectroscopy and direct sample analysis. *Anal. Methods* **2017**, *9*, 4255. [CrossRef]
75. Wang, N.; Fu, Y.; Lim, L. Feasibility Study on Chemometric Discrimination of Roasted Arabica Coffees by Solvent Extraction and Fourier Transform Infrared Spectroscopy. *J. Agric. Food Chem.* **2011**, *59*, 3220–3226. [CrossRef]
76. Dankowska, A.; Domagała, A.; Kowalewski, W. Quantification of Coffea arabica and Coffea canephora var. robusta concentration in blends by means of synchronous fluorescence and UV-Vis spectroscopies. *Talanta* **2017**, *172*, 215–220. [CrossRef] [PubMed]
77. Escobar, M.M.; Torres, A.E.L.; Rodriguez, N.J.M. Non-Destructive Prediction of Moisture Content of Philippine Coffea arabica and Coffea liberica Green Beans Using Locally-Developed NIR Spectroscopy Instrument. *Mindanao J. Sci. Technol.* **2020**, *18*, 208–223.
78. Kyaw, E.M.; Budiastra, I.W.; Samsudin; Sutrisno. Estimation of moisture content in Liberica coffee by using near infrared spectroscopy. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *542*, 012013. [CrossRef]
79. Macedo, L.L.; Araújo, C.D.S.; Vimercati, W.C.; Hein, P.R.G.; Pimenta, C.J.; Saraiva, S.H. Evaluation of chemical properties of intact green coffee beans using near-infrared spectroscopy. *J. Sci. Food Agric.* **2020**, *101*, 3500–3507. [CrossRef]
80. Liang, N.; Lu, X.; Hu, Y.; Kitts, D.D. Application of Attenuated Total Reflectance-Fourier Transformed Infrared (ATR-FTIR) Spectroscopy To Determine the Chlorogenic Acid Isomer Profile and Antioxidant Capacity of Coffee Beans. *J. Agric. Food Chem.* **2016**, *64*, 681–689. [CrossRef]
81. Weldegebreal, B.; Redi-Abshiro, M.; Chandravanshi, B.S. Development of new analytical methods for the determination of caffeine content in aqueous solution of green coffee beans. *Chem. Cent. J.* **2017**, *11*, 126. [CrossRef]
82. Yisak, H.; Redi-Abshiro, M.; Chandravanshi, B.S. New fluorescence spectroscopic method for the simultaneous determination of alkaloids in aqueous extract of green coffee beans. *Chem. Cent. J.* **2018**, *12*, 59. [CrossRef] [PubMed]
83. Bressanello, D.; Liberto, E.; Cordero, C.; Rubiolo, P.; Pellegrino, G.; Ruosi, M.R.; Bicchi, C. Coffee aroma: Chemometric comparison of the chemical information provided by three different samplings combined with GC–MS to describe the sensory properties in cup. *Food Chem.* **2017**, *214*, 218–226. [CrossRef]
84. Alcantara, G.M.R.N.; Dresch, D.; Melchert, W.R. Use of non-volatile compounds for the classification of specialty and traditional Brazilian coffees using principal component analysis. *Food Chem.* **2020**, *360*, 130088. [CrossRef] [PubMed]

85. Toci, A.T.; Farah, A.; Pezza, H.R.; Pezza, L. Coffee Adulteration: More than Two Decades of Research. *Crit. Rev. Anal. Chem.* **2016**, *46*, 83–92. [CrossRef] [PubMed]
86. Forchetti, D.A.P.; Poppi, R.J. Detection and Quantification of Adulterants in Roasted and Ground Coffee by NIR Hyperspectral Imaging and Multivariate Curve Resolution. *Food Anal. Methods* **2020**, *13*, 44–49. [CrossRef]
87. Franca, A.S.; Oliveira, L.S.; Mendonc, J.C.F.; Silva, X.A. Physical and chemical attributes of defective crude and roasted coffee beans. *Food Chem.* **2005**, *90*, 84–89. [CrossRef]