Food Research International 172 (2023) 113216

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Authentication and discrimination of new Brazilian Canephora coffees with geographical indication using a miniaturized near-infrared spectrometer

Michel Rocha Baqueta [a,b], Federico Marini [b], Rodrigo Barros Rocha [c], Patrícia Valderrama [d,*], Juliana Azevedo Lima Pallone [a,*]

[a] University of Campinas – UNICAMP, School of Food Engineering, Department of Food Science and Nutrition, Campinas, São Paulo, Brazil
[b] Department of Chemistry, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome, Italy
[c] Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA Rondônia, Porto Velho, Rondônia, Brazil
[d] Universidade Tecnológica Federal do Paraná – UTFPR, Campo Mourão, Paraná, Brazil

ARTICLE INFO

Keywords:
Authentication
Chemometrics
DD-SIMCA
Geographical origin
MicroNIR
PLS-DA

ABSTRACT

New Brazilian Canephora coffees (Conilon and Robusta) of high added value from specific origins have been protected by geographical indication to guarantee their origin and quality. Recently, benchtop near-infrared (NIR) spectroscopy combined with chemometrics has demonstrated its usefulness to discriminate them. It was the first study, however, and therefore the possibility exists to develop a new portable NIR method for this purpose. This work assessed a miniaturized NIR as a cheaper spectrometer to discriminate and authenticate new Brazilian Canephora coffees with certified geographical origins and to differentiate them from specialty Arabica. Discriminant chemometric and class modeling techniques have been applied and have obtained good predictive ability on external test sets. In addition, models with similar classification purpose were compared with those obtained in previous research carried out with benchtop NIR for the same samples, obtaining comparable results. In this context, the portable method was used as a laboratory technique and has the advantage of being cheaper than benchtop NIR spectrometer. Furthermore, it brings a high possibility to be implemented in small coffee cooperatives, industries or control agencies in the future that do not have high economic resources.

1. Introduction

Brazil is the second-largest Canephora producer in the world, ranking only behind Vietnam (ICO, 2021). The most important producing states in the country are Espírito Santo, Rondônia, and Bahia, as illustrated in Fig. 1. They produce two distinct botanical varieties of Canephora: Robusta and Conilon. Conilon is typically cultivated in Espírito Santo (Southeast region) and Bahia (Northeast region), while Robusta is produced in Rondônia (North region) (CONAB, 2023; de Souza et al., 2021).

Conilon and Robusta coffees have historically been undervalued coffees in the world (Liu et al., 2019). However, their quality standards in Brazil have improved currently, and new Canephora have emerged with qualities that are comparable or superior to specialty Arabica (Alves et al., 2020; Filete et al., 2022; Gomes et al., 2022; Lemos et al., 2020; de Souza et al., 2021; Zani Agnoletti et al., 2022). Consequently, Conilon produced in Espírito Santo (Brazil, 2021b) and Robusta Amazônico from Rondônia (Brazil, 2021d) were registered with the geographical indication (GI) (Brazil, 2021c). Robusta Amazônico coffees from Rondônia are cultivated in a sustainable agroforestry system in the Amazon region by indigenous and non-indigenous coffee producers (Brazil, 2021a), with different manufacturers providing a differentiation issue. In turn, Espírito Santo has a consolidated history in the production of the Conilon variety (Brazil, 2021b).

Recently, benchtop near-infrared (NIR) spectroscopy combined with chemometrics has demonstrated its usefulness to discriminate roasted Brazilian Canephora coffees, their origins, cultivars from Western Brazilian Amazon, and specialty Canephora and Arabica in a previous research carried out by the authors of this study (Baqueta et al., 2023). The comprehensive sample set comprised 527 samples analyzed by the benchtop NIR adopting different strategies of multi-class partial least squares with discriminant analysis (PLS-DA) to classify the five available coffee classes. Subsequently, it was applied binary PLS-DA models to discriminate only GI Canephora coffees from those without GI specification, as well as GI Canephora versus GI Canephora coffees (Robusta Amazônico versus Conilon from Espírito Santo). The results indicated

* Corresponding authors.
E-mail addresses: pativalderrama@gmail.com (P. Valderrama), jpallone@unicamp.br (J.A.L. Pallone).

https://doi.org/10.1016/j.foodres.2023.113216
Received 21 March 2023; Received in revised form 27 June 2023; Accepted 29 June 2023
Available online 30 June 2023
0963-9969/

that the discrimination among the three GI Canephora coffees, and between indigenous and non-indigenous producers of Robusta Amazônico was not a straightforward task even for benchtop NIR. It was the first study, however, and therefore the possibility exists to develop a new portable NIR method for this purpose.

Miniaturization of NIR spectroscopy has made substantial progress with the trend towards portability, especially for in-field analysis for natural samples (Beć et al., 2022). However, different studies have introduced the applicability of portable NIR as a laboratory technique to analyze prepared samples, expanding its application and making it versatile and cheaper compared to the benchtop spectrometer. For coffee, measurements directly in the field on raw coffee beans do not reflect the final product that reaches the consumer. In addition, a portable NIR is much more affordable for the reality of many coffee industries and for control agencies that do not have high economic resources.

The application of portable NIR has simplified coffee analyses (Baqueta, Coqueiro, Março, & Valderrama, 2021; Baqueta, Coqueiro, Março, Mandrone, et al., 2021; Correia et al., 2018, 2020; Souza et al., 2022). However, the main limitation with portable NIR is that almost all available instruments are not designed to cover a wide NIR spectral range. Portable sensors with InGaAs-based short wave infrared (SWIR) detector cover 900–1700 nm range and capture partly the 1st and the 2nd overtones of most functional group vibrations. Therefore, in counterpart to the benchtop spectrometer, this is its main drawback, which makes the application of chemometrics a challenging task (Mishra et al., 2021).

This work assessed a miniaturized NIR as a cheaper spectrometer to discriminate and authenticate a very large set of new Brazilian Canephora coffees with certified geographical origins and to differentiate them from specialty Arabica. Exploratory and discriminant chemometric techniques have been applied to evaluate the spectral data and then fairly compared (because the same samples have been analyzed) with the classification results obtained previously with benchtop NIR in Baqueta et al. (Baqueta et al., 2023). Additionally, one class-modeling by data-driven soft independent modeling of class analogy (DD-SIMCA) has been proposed to authenticate GI Canephora. The possibility of developing a new portable NIR method to certify Canephora coffees of high added value from specific Brazilian GIs may increases the interest in this technique.

2. Materials and methods

2.1. Coffee samples

A large sample set composed of 527 roasted and ground coffee samples was analyzed. It comprised the same set previously analyzed (Baqueta et al., 2023) by the authors of this study. They were roasted to a medium degree in a Probat sample roaster with the initial temperature of 160 °C and 190 °C at the end, with a time ranging between 7.5 min and 9 min to obtain the desired profile. Table 1 specifies the number of samples and their origins. From the 99 Robusta Amazônico samples from the indigenous producers, 78 were from "Sete de Setembro" indigenous land and 21 from "Rio Branco" indigenous land. Conilon from São Paulo and low-quality Robusta/Conilon coffees were included for a brief comparison and were not the focus of the study. Robusta Amazônico samples and Conilon from Espírito Santo were provided by the Brazilian Agricultural Research Corporation (EMBRAPA) located in Porto Velho in the state of Rondônia, which guaranteed their GI authenticity.

2.2. Portable NIR spectra collection

The coffee beans were milled and sieved through a 20-mesh sieve for particle size standardization before analysis. The analyses were carried out in the solid roasted coffees, in ground form, to reflect the final product that reaches the consumer. MicroNIR spectra of the milled coffee samples were obtained directly using a portable NIR spectrometer (microNIR™ 1700) from JDSU Uniphase Corporation with a glass cuvette, in reflectance mode. This spectrometer was designed with an InGaAs-based short wave infrared (SWIR) that covers the 906–1676 nm

Table 1
Coffee samples included in the study.

Samples	Varietal types	Origin/Producer
99	Robusta Amazônico	Indigenous producers from Rondônia
133	Robusta Amazônico	Non-indigenous from Rondônia
126	Conilon	Espírito Santo
75	Conilon	Bahia
75	Specialty Arabica	Different origins, qualities, and sensory characteristics
7	Conilon	São Paulo
12	Low-quality Conilon/ Robusta	Unknown origin

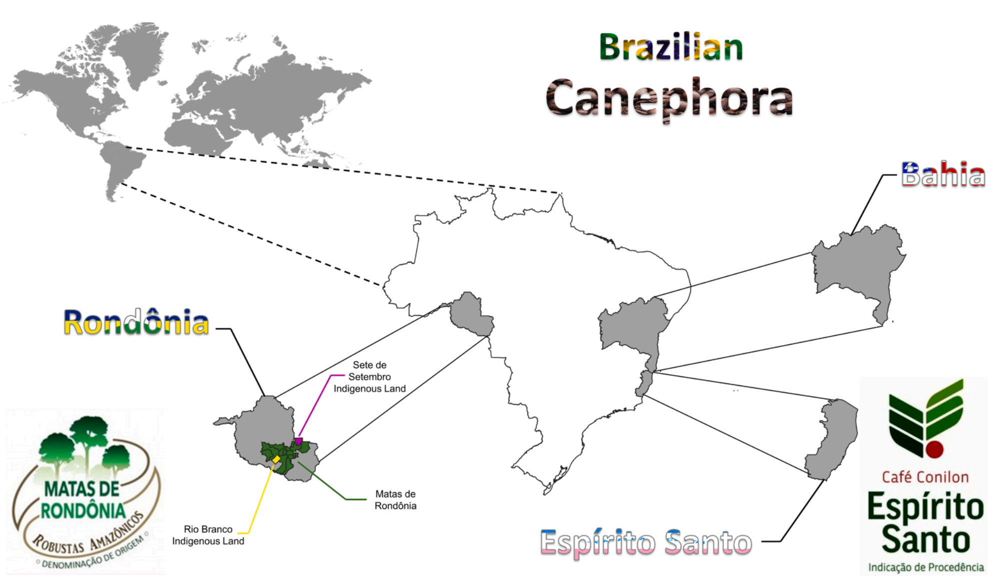

Fig. 1. Graphical representation showing the main Canephora producing states in Brazil (Espírito Santo, Bahia and Rondônia in gray), with highlight for Matas de Rondônia (Robusta) colored in green, where the "Sete de Setembro" Indigenous Land is highlighted in purple, and "Rio Branco" Indigenous Land is highlighted in yellow, as well as GI seals for Robusta Amazônico from Rondônia and Conilon from Espírito Santo. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

range. The samples were randomly analyzed at room temperature (22 °C). Three different sample aliquots were used, and the spectrum of each aliquot was recorded with 50 scans and a resolution of 6.25 nm, resulting in a measurement time of 0.50 s. The blank was evaluated using a standard NIR reflectance (Spectralon™) with a diffuse reflection coefficient of 99%, while a dark reference (zero–to simulate non-reflection) was obtained with the lamp off. The MicroNIR had a diameter of 45 mm and a height of 42 mm, weighing about 60 g.

2.3. Spectral pre-process and exploratory analysis

Each sample's average microNIR spectra were calculated and imported into Matlab R2019a (The Mathworks, Natick, MA) with the PLS_Toolbox 8.6 computational package. The 527 original spectroscopic profiles had 121 variables per spectrum and were transformed into absorbance (log 1/R). Different pre-processing methods were studied before the chemometric analysis. The best one was combining Savitzky–Golay smoothing with a 5-point window (Savitzky & Golay, 1964), and multiplicative scatter correction (MSC) (Geladi et al., 1985). Moreover, the spectra were mean centered after pre-processing. First, PCA was used as an unsupervised method to map patterns in samples, trying to differentiate them in terms of varietal types (Robusta and Conilon), species (Canephora and Arabica), quality (high and low quality Canephora), or geographical origin (Rondônia and its different producers, Espírito Santo, Bahia, and São Paulo).

2.4. Chemometric classification

The samples belonging to each dataset in the supervised models were selected by the Kennard-Stone algorithm (Kennard & Stone, 1969). Calibration datasets (training sets) comprised 75% of the samples selected from each class. Prediction datasets (test sets) composed 25% of

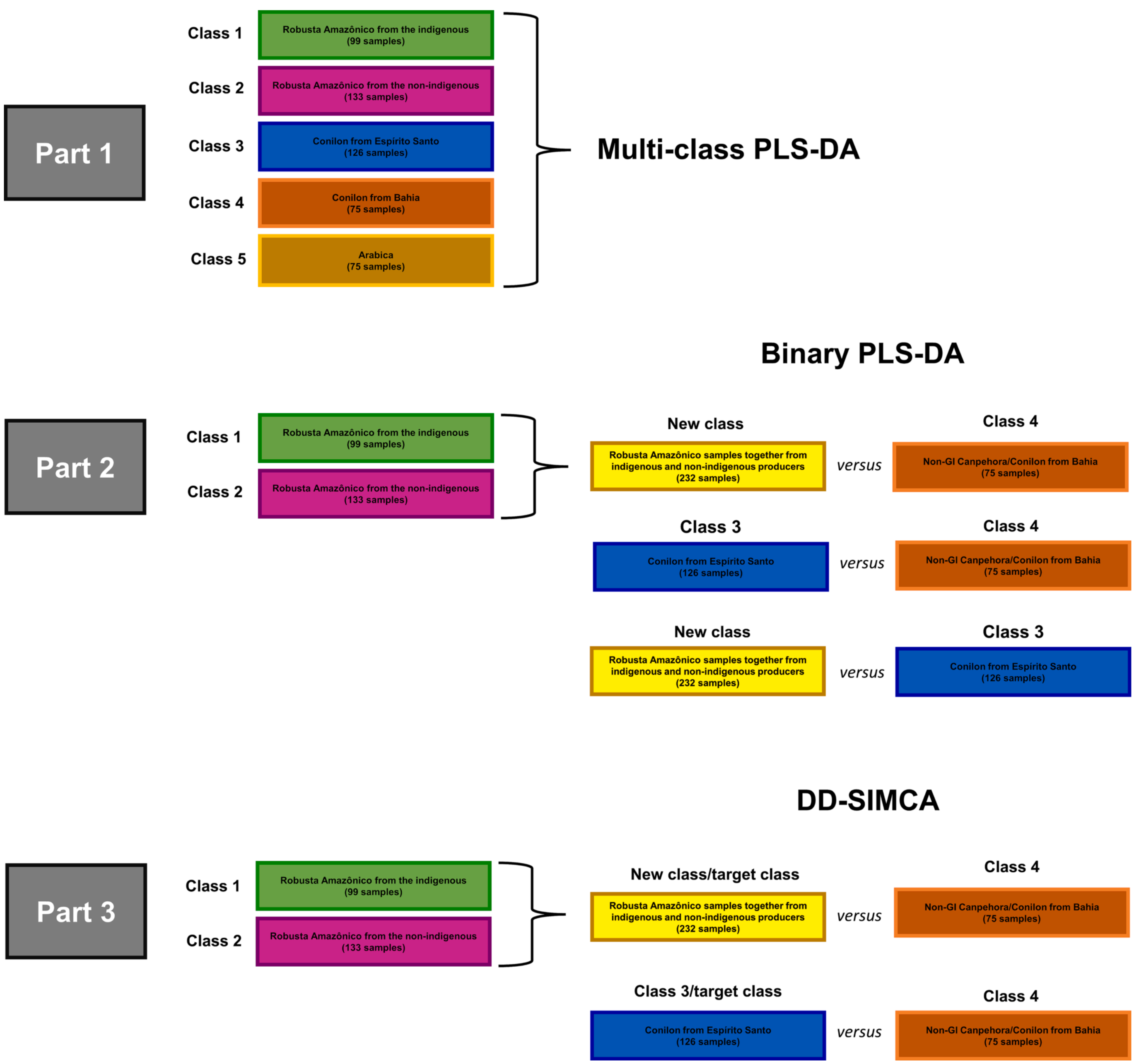

Fig. 2. Scheme applied to perform the 5-class discrimination (multi-class PLS-DA or PLS2-DA), 2-class discrimination (binary PLS-DA or PLS1-DA), and one-class authentication of the coffee samples.

the remaining samples to evaluate the model's predictive ability.

Fig. 2 shows a graphical representation of the data sets and from the chemometric analyses performed divided into 3 parts. In the part 1, a multi-class PLS-DA, which can be called as PLS2-DA, was developed to model and classify five coffee classes of interest: Robusta Amazônico from all indigenous producers in a single class (class 1), Robusta Amazônico from non-indigenous producers (class 2), Conilon from Espírito Santo (class 3), Conilon from Bahia (class 4), and specialty Arabica (class 5). Conilon from São Paulo and low-quality Canephora were not included in the classification/authentication models considering the small number of samples.

In the part 2, three binary PLS-DA, which can be called as PLS1-DA, were built: a first to distinguish between all Robusta Amazônico and non-GI Canephora coffees, one second to differentiate between Conilon from Espírito Santo and non-GI Canephora coffees, and a third to distinguish between Robusta Amazônico and Conilon from Espírito Santo. Non-GI Canephora coffees were represented by Conilon from Bahia that is not currently registered with GI.

The PLS1-DA models were developed taking into account that the multi-class PLS2-DA model had excessive variability (five modeled classes) and a strong variance contribution between the Canephora/ Arabica coffee species that could influence the desired discrimination between only the Canephora groups.

All PLS-DA models were implemented using Matlab R2019a (The Mathworks, Natick, MA) with the PLS_Toolbox 8.6 computational package. The internal validation of the model was carried out through a venetian blinds cross-validation procedure with five samples. The criterion for choosing the optimal number of latent variables (LVs) was determined by figures of merit as sensitivity and specificity and classification errors. In addition, samples showing both high leverage and high Q residuals were detected and removed since they could be outliers. The quality and reliability of the PLS-DA models were evaluated in terms of sensitivity and specificity. The performance of the PLS-DA models was assessed by determining sensitivity and specificity values for calibration (CAL), cross-validation (CV), and prediction (PRED) datasets (Baqueta, Coqueiro, et al., 2021; López et al., 2015). Sensitivity is defined for each class as the percentage of samples of this class, which are correctly recognized as members of this class in the discriminant model. Specificity is defined for each class as the percentage of samples from other classes, which are correctly rejected by the model and therefore attributed as inconsistent with the class of interest. These parameters can assume values from 0 to 1 or 0% to 100%. A perfect sample discrimination occurs when 100% sensitivity and 100% specificity are obtained (Rodionova & Pomerantsev, 2020).

2.5. Chemometric authentication

DD-SIMCA was applied as a one-class classifier (Rodionova, Titova, et al., 2016) to authenticate the GI Canephora coffees. Two DD-SIMCA models were built as shown in Fig. 2: the first with all Robusta Amazônico samples as a single target class, thus creating the training set for model building. Another DD-SIMCA was developed using Conilon from Espírito Santo as the target class, where these samples comprised, the training set for model building. Training and test sets of the target classes were the same used in the PLS-DA models. Non-GI Canephora coffees (those represented by Conilon from Bahia) were exclusively considered for testing the predictive ability of the two DD-SIMCA models and were included in the test set together with the validation samples of each target class. The acceptance significance and the outlier significance with levels of 0.01 were adopted. The number of principal components (PCs) was selected according to their explained variance in PCA and was applied to the training samples (target class). DD-SIMCA is freely available at https://github.com/yzontov/dd-simca, and implementation details can be found in the reference paper (Zontov et al., 2017). The quality and reliability of the DD-SIMCA models were also evaluated in terms of sensitivity and specificity. Sensitivity was determined for training sets, while sensitivity and specificity were verified for the test sets. In the case of SIMCA, sensitivity is obtained exclusively for the target class, while the sensitivity and specificity can be calculated for the non-target class (Rodionova, Oliveri, et al., 2016).

3. Results and discussion

3.1. Visualization of the spectra

The original, pre-processed, and average microNIR spectra of the ground and roasted coffee samples are shown in Fig. 3A-C. In all cases, the differences among classes were not easily observed. The molecular vibration bands occurred similarly throughout the spectra region with high spectral overlap between coffee classes. The spectra were very similar over almost the entire range, except for the region around 1450 nm, where the differences between the spectra were visible.

Even though previous studies have reported a clear effect of coffee variety/species/origin on the NIR spectrum (Giraudo et al., 2019; Monteiro et al., 2018), the current analysis with the portable instrument has shown that the determination of different varietal types, geographical origins, and distinction of producers for the new Brazilian Canephora coffees can be a complex task, even though there are differences in their composition, as already shown in previous studies (Francisco et al., 2021; Oliveira et al., 2020). An adequate spectra analysis requires the application of chemometric methods.

3.2. Exploratory analysis

A PCA brought more information of interest about the samples. The first three principal components (PCs) and respective loadings are shown in Fig. 4. Overall, PC1 (62.94% of explained variance) in Fig. 4A, and the joint observation of PC1 and PC2 (26.00% of explained variance) in Fig. 4B, tended to differentiate between GI Canephora coffees (Robusta Amazônico and Conilon from Espírito Santo) and others (Conilon from Bahia, Arabica, and low-quality Canephora). Conilon from São Paulo showed similar behavior to the GI Canephora coffees. This was even better observed in Fig. 4C with the visualization of PC1 and PC3 with 7.20% of variance explained. However, the analysis of the PCs together also showed a huge overlap of samples, not differentiating all the available classes.

The loadings of PC1 and PC2 (Fig. 4D) were more intense between 1400 and 1600 nm. High absorption bands in this region were previously assigned to the vibration bands of carbohydrates and chlorogenic acids (Barbin et al., 2014; Ribeiro et al., 2011). Previous studies have shown that Arabica contains more sugars than Conilon and Robusta (Caporaso et al., 2018; Lemos et al., 2020; Liu et al., 2019). However, new Brazilian Conilon genotypes have been found with higher concentrations of sugars and even higher than Arabica (Lemos et al., 2020). Furthermore, Robusta and Conilon were associated with higher concentrations of chlorogenic acids than Arabica (Ky et al., 2001; Lemos et al., 2020; Liu et al., 2019). Interestingly, a recent study showed that a Brazilian Conilon genotype and a Brazilian Robusta contained nearly three times as much caffeoylquinic acid as Arabica (Lemos et al., 2020).

There was a region of importance for PC3 between 900 and 1150 nm (Fig. 4D). It was previously associated with the difference absorption of caffeine and trigonelline in coffee species, where caffeine content is typically higher in Canephora than in Arabica, whereas Arabica has more trigonelline than Robusta (Caporaso et al., 2018; ICO, 2022; Lemos et al., 2020; Liu et al., 2019). However, it was shown recently (Lemos et al., 2020) that Conilon with 100% mature cherry fruit can have similar or even higher trigonelline content than Arabica.

3.3. Discrimination of Robusta, Conilon, and Arabica

A multi-class PLS-DA model was developed for distinguishing the five coffee classes of interest in this study: Robusta Amazônico from

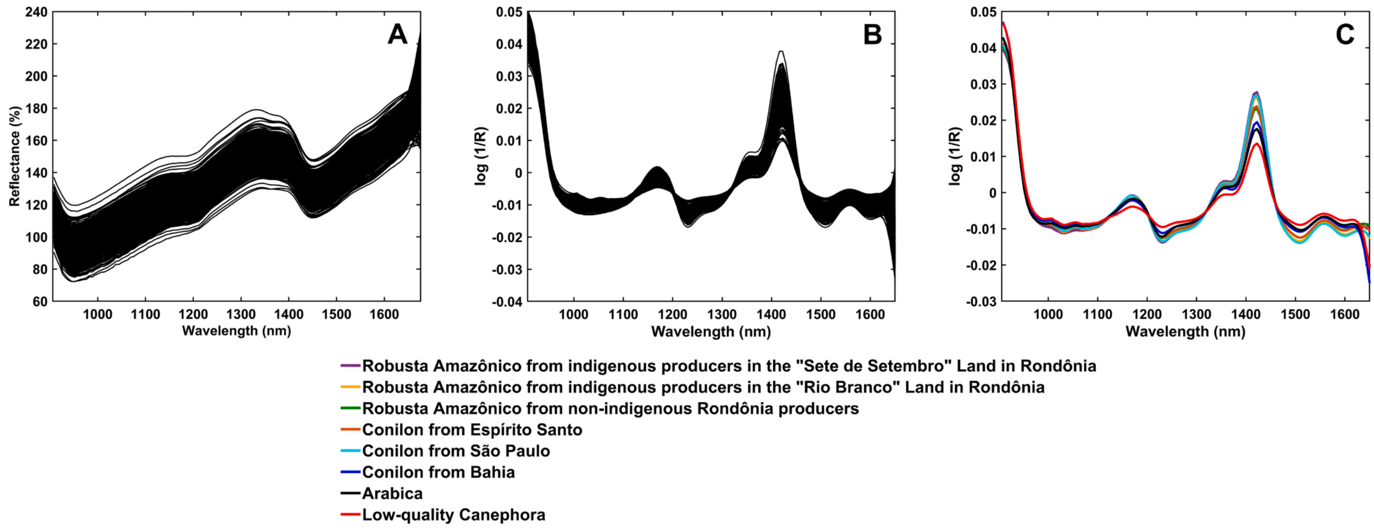

Fig. 3. Original (A), pre-processed (B), and average (C) microNIR spectra of the coffee samples.

indigenous producers (class 1), Robusta Amazônico from non-indigenous Rondônia producers (class 2), Conilon from Espírito Santo (class 3), Conilon from Bahia (class 4), and specialty Arabica coffees (class 5). A total of 19 LVs were used in the multi-class PLS-DA.

The current PLS-DA model dealt with a typical case reported in multi-class modeling, which requires a high number of variables to obtain reasonable results. Each class takes around 2 to 3 LVs for the internal modeling, and another 1 to 2 LVs are necessary to describe the external links between classes, not meaning an overfitting (Pomerantsev & Rodionova, 2018). In the 5-class discrimination, this means at least 17 LVs. With the 19 LVs selected for the PLS-DA, it was possible to obtain about 72% of the explained variation in Y, which is low and has been reported previously for a similar case study (Baqueta, Coqueiro, et al., 2021). The attempt to discriminate the five classes of roasted coffee in a single model is challenging in this work and reflects it in the model complexity.

Sensitivity and specificity rates obtained in the multi-class PLS-DA are shown in Table 2a. The model discriminated perfectly the Conilon from Bahia that still has predominantly an intermediate quality, and Arabica. It correctly classified all these samples, showing 100 % sensitivity and specificity in the prediction. However, the main interest was in discriminating high value-added Canephora coffees from specific Brazilian geographical indications or producers: Robusta Amazônico from indigenous producers, Robusta Amazônico from non-indigenous producers, and Conilon from Espírito Santo. For these three classes, the sensitivity and specificity were in the range of 84.0–97.0 % in the prediction. This may indicate that GI Canephora share characteristics from a chemical point of view, creating misclassifications. No Arabica or Conilon samples produced in Bahia were wrongly assigned to the other classes.

Robusta Amazônico coffees are produced in the delimitation of the GI “Matas de Rondônia” shown in Fig. 1, which includes the two indigenous lands. However, a relatively high discrimination between them was obtained and this may be associated with the specific soil type of the protected indigenous regions (that maintains native forest) and also the beans processing carried out by each group. It is known that indigenous people produce their coffee mostly with induced fermentation techniques.

Few cases in the literature have addressed real problems with more than three classes available, especially using portable equipment that often has less information contained in the spectra. The results found there are consistent with previous studies that have faced the same challenge (Baqueta, Coqueiro, et al., 2021; dos Santos et al., 2021).

3.4. Discrimination of GI and non-GI Canephora

After multi-class classification, binary PLS-DA models were applied to discriminate specifically GI Canephora coffees from those without GI. Even though to some extent this discrimination was partially tested in the previous multi-class model, here the variation between coffee species was not considered. Furthermore, it is worth noting that a two-class model is simpler and less challenging than a five-class model.

The results are shown in Table 2b. The model to discriminate Robusta Amazônico coffees from non-GI Canephora and the model to discriminate Conilon from Espírito Santo from non-GI Canephora both had 100% correct classification in calibration, cross-validation, and prediction. Both models were built using 2 LVs. These results demonstrated that the models perfectly discriminate GI Canephora from those without the GI certification, inferring that the geographical origin and other factors relevant to a food with GI protection (quality of raw product and production system, for example) had effect in this discrimination.

When applying a third model to try to differentiate between Robusta Amazônico and Conilon from Espírito Santo (GI Canephora versus another GI Canephora), sensitivity and specificity range between 93.1 and 93.5%, and 19 LVs were necessary to achieve the best compromise between sensitivity and specificity values. This indicates that PLS-DA model for discriminating GI Canephora from a non-GI Canephora performs better than trying to discriminate two GI Canephora.

3.5. Comparison of classification results obtained between portable and benchtop NIR

In order to explore the advantages associated to the use of a portable spectrophotometer for Canephora coffee classification, the performance of the portable NIR and benchtop NIR was compared. For this purpose, PLS-DA results developed with identical classification purposes using benchtop NIR for the same samples in a previous research carried out by the authors of this study (Baqueta et al., 2023) were considered. The results obtained in the prediction sets of PLS-DA models from both instruments were examined by comparing the percentage of sensibility and specificity in Table 3.

The sensitivity and specificity percentages for the multi-class models obtained with NIR instruments (Table 3a) showed that the results were comparable/distinct depending on the predicted class. The 100 % correct classification was repeated for Conilon from Bahia and Arabica. The results obtained for binary PLS-DA models were similar for both NIR instruments (Table 3b), exhibiting always 100 %. In general, it is

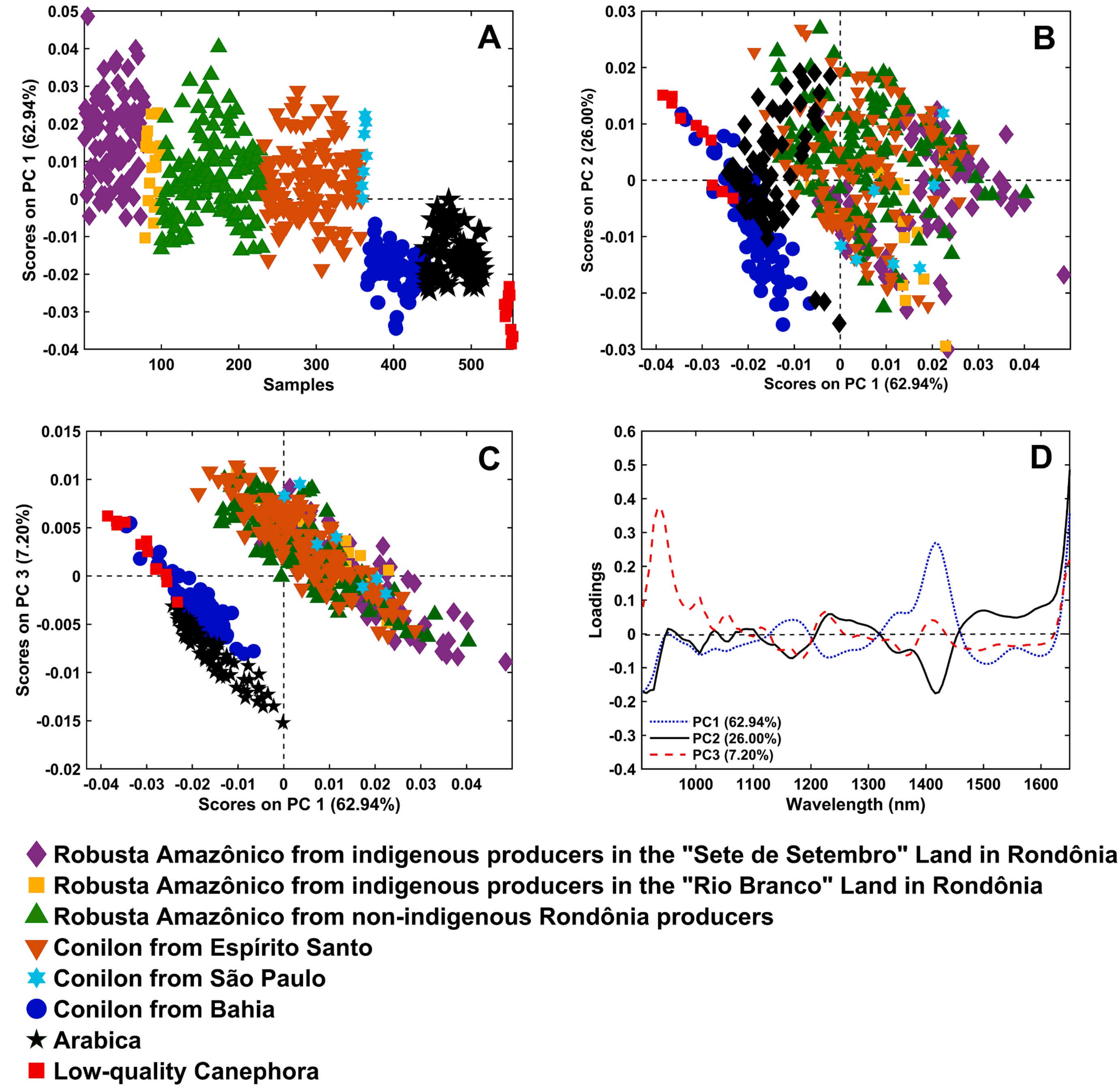

Fig 4. PCA scores (A-C) and respective loadings (D) in a model with all coffee samples.

possible to say that the portable NIR performed similarly to the benchtop NIR in this case. The main advantage of portable NIR in this study is that it is cheaper to apply it as laboratory equipment, even though its main objective is field analysis. For industrial-scale analysis with a larger number of samples, it can further simplify the procedure and increase the flow of analysis.

3.6. *Authentication of GI Canephora coffees*

GI Canephora coffees were also authenticated using DD-SIMCA models (Table 2c), where 3 PCs achieved a 100% correct classification in both the training and test sets. Therefore, DD-SIMCA provided more a method to certify Canephora from specific GI in Brazil. These results are consistent with those obtained by the binary PLS-DA models.

3.7. *Chemical interpretation of the PLS-DA with VIP scores*

An advantage of PLS-DA under DD-SIMCA is that through VIP (Variable Importance in Projection) scores, it is possible to identify significant variables in the modeling and recognize useful complementary information for chemical interpretation (Foschi et al., 2021). Fig. 5 shows VIP scores. The dashed line corresponds to a cut-off to define significant spectral variables, those with a VIP index greater than one. Multi-class PLS-DA (Fig. 5A-E) exhibited different patterns of NIR-VIP-bands to discriminate the five coffee classes. Robusta Amazônico from indigenous (Fig. 5A) and non-indigenous producers (Fig. 5B) showed slightly similar NIR-VIP-bands. Conilon from Espírito Santo showed the most distinct NIR-VIP band pattern (Fig. 5C). Conilon from Bahia (Fig. 5D) and Arabica coffees (Fig. 5E) also showed a similar NIR-VIP band profile. However, a prominent band around 900 nm for Arabica

Table 2
Results for PLS-DA (a and b) and DD-SIMCA (c) models to classify and authenticate coffee samples using portable microNIR.

(a) Multi-class PLS-DA	Robusta Amazônico from the indigenous	Robusta Amazônico from the non-indigenous	Conilon from Espírito Santo	Conilon from Bahia	Arabica coffees
LVs	19				
Variance captured in Y block	72.06				
Sensitivity CAL	95.9	90.9	96.8	100.0	100.0
Specificity CAL	94.1	90.7	95.1	100.0	100.0
Sensitivity CV	89.0	79.8	92.6	100.0	100.0
Specificity CV	91.9	87.9	93.0	100.0	100.0
Sensitivity PRED	84.0	97.0	96.8	100.0	100.0
Specificity PRED	95.0	92.4	91.5	100.0	100.0

(b) GI classification with PLS-DA	Robustas Amazônicos	Canephora without GI	Conilon from Espírito Santo	Canephora without GI	Robustas Amazônicos	Conilon from Espírito Santo
LVs	2		2		19	
Variance captured in Y block	81.1		92.9		70.9	
Sensitivity CAL	100.0	100.0	100.0	100.0	96.0	96.8
Specificity CAL	100.0	100.0	100.0	100.0	96.8	96.0
Sensitivity CV	100.0	100.0	100.0	100.0	91.4	92.6
Specificity CV	100.0	100.0	100.0	100.0	92.6	91.4
Sensitivity PRED	100.0	100.0	100.0	100.0	93.1	93.5
Specificity PRED	100.0	100.0	100.0	100.0	93.5	93.1

(c) GI authentication with DD-SIMCA	Robustas Amazônicos	Canephora without GI	Conilon from Espírito Santo	Canephora without GI
PCs	3		3	
Variance captured in the PCA	96.9		96.8	
Sensitivity TRAINING	100.0	–	100.0	–
Specificity TEST	100.0	100.0	100.0	100.0

Table 3
Comparison of PLS-DA models obtained with portable (present work) and benchtop (obtained in Baqueta et al. 2023) NIR instruments based on prediction data sets.

(a) Multi-class PLS-DA	Robustas Amazônicos from the indigenous		Robustas Amazônicos from the non– indigenous		Conilon from Espírito Santo		Conilon from Bahia		Arabica coffees	
	Portable NIR	Benchtop NIR	Portable NIR	Benchtop NIR	Portable NIR	Benchtop NIR	Portable NIR	Benchtop NIR	Portable NIR	Benchtop NIR
Sensitivity PRED	84.0	100.0	97.0	93.9	96.8	96.8	100.0	100.0	100.0	100.0
Specificity PRED	95.0	97.0	92.4	90.2	91.5	98.9	100.0	100.0	100.0	100.0

(b) GI classification with PLS-DA	Robustas Amazônicos		Canephora without GI		Conilon from Espírito Santo		Canephora without GI	
	Portable NIR	Benchtop NIR	Portable NIR	Benchtop NIR	Portable NIR	Benchtop NIR	Portable NIR	Benchtop NIR
Sensitivity PRED	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Specificity PRED	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

coffees (Fig. 5E) differed from its NIR-VIP bands.

In binary PLS-DA models, similar NIR-VIP-bands were significant in discriminating GI Canephora coffees from those without GI (Fig. 5F-G). There was a slight difference that comes more from the magnitude of the VIP index than from its profile. For Robusta Amazônico and Conilon from Espírito Santo, a series of NIR-VIP-bands in the spectral microNIR region were significant to model their differences (Fig. 5H).

It is not a simple task to assign the absorption bands in NIR spectroscopy because of the overlapping wide bands and this becomes even more challenging when applied to the coffee matrix, as in this study. The coffee beans before roasting contain several chemical compounds, which react and interact amongst during the roasting, resulting in hundreds of products (Ribeiro et al., 2011). However, caffeine, trigonelline, sugars, lipids and chlorogenic acids are often reported to be of importance for NIR coffee spectra. Fig. 6 illustrates the region of absorbance of the main compounds found in coffee analyzed by the NIR spectroscopy in the range of the portable equipment (around 900–1700 nm) based on previous literature (Barbin et al., 2014; Ribeiro et al., 2011).

It is difficult to make specific assignments taking into account even the NIR-VIP-bands with variables selected and consequently reduced than the whole spectra, because many compounds present in coffee share the same absorption region. Therefore, other types of analysis could be tested for interpreting the absorption regions in future studies. One of them is the quantum chemical spectra simulation of NIR spectra, which could reveal more specific information about the wavelength regions where the best-correlated spectral information resides and unveils the interactions of the target analyte with the surrounding matrix, as previously demonstrated by Beć et al. (Beć et al., 2022). Although it should be noted that this is only a feasibility study, the results obtained

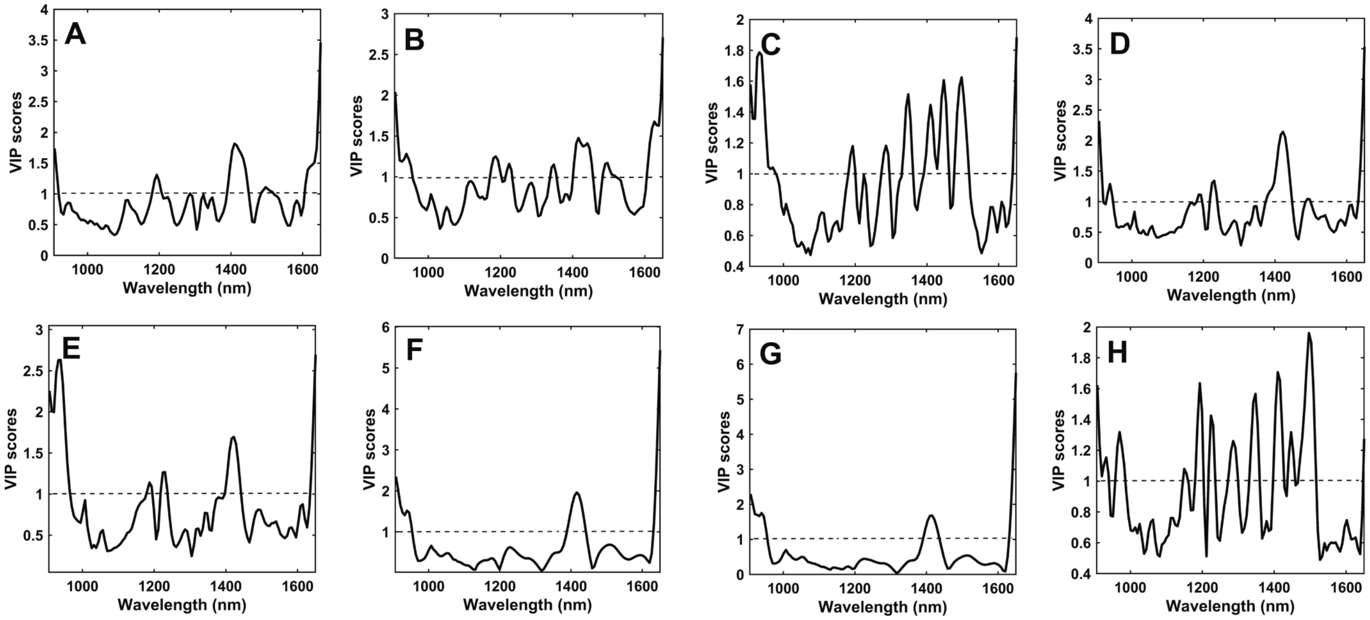

Fig. 5. VIP analysis and chemical interpretation of the PLS-DA models. Multi-class PLS-DA (A – class 1 – Robusta Amazônico from indigenous producers), (B – class 2 – Robusta Amazônico from non-indigenous Rondônia producers), (C – class 3 – Conilon from Espírito Santo), (D – class 4 – Conilon from Bahia), (E – class 5 – specialty Arabica coffees); Binary PLS-DA models to differentiate all Robusta Amazônico together of both producers versus Canephora without GI (E), Conilon from Espírito Santo versus Canephora without GI (F) and all Robusta Amazônico together of both producers versus Conilon from Espírito Santo (H).

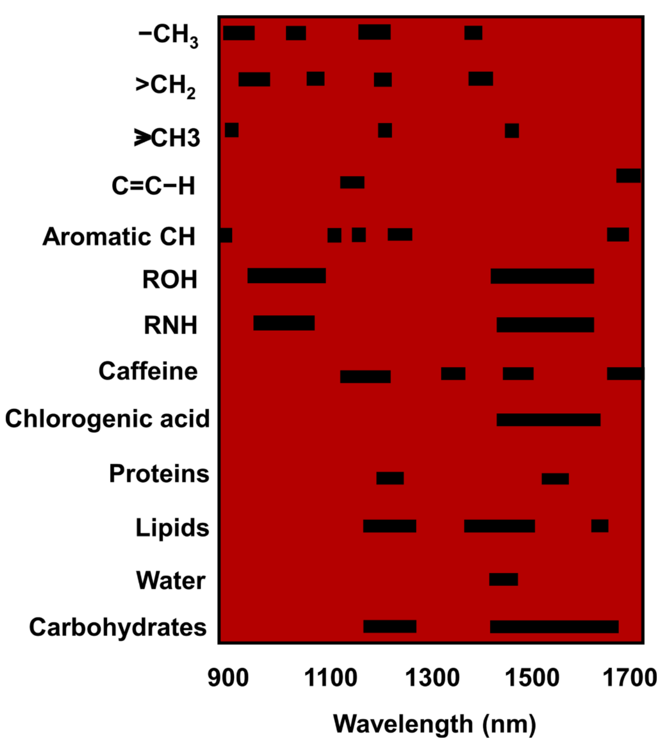

Fig. 6. Graphical representation of the wavelength range of the portable NIR in which have been reported the greatest absorbance of main components found in coffee (adapted from Barbin et al., 2014).

encourage the realization of a future work dedicated to the development of a practical guide for interpretive NIR spectra in coffee, which could benefit from this approach to realize the assignment of absorption bands more specifically. However, the applicability of the proposed methodology requires further research.

Previous studies have shown that caffeine, trigonelline, sugars, lipids, and chlorogenic acids exhibit several vibration bands in the range of the portable NIR from ∼ 906 to ∼ 1650 nm (Barbin et al., 2014; Ribeiro et al., 2011). In all cases, VIP indexes were often higher for variables in the spectral region around 1400 nm, an area strongly associated with the presence of carbohydrates or the absorption of chlorogenic acid and lipids (Baqueta, Coqueiro, Março, Mandrone, et al., 2021; Barbin et al., 2014; Ribeiro et al., 2011). Variables around 1200 nm were also highlighted as relevant predictors for all PLS-DA models, except for binary versions of GI versus non-GI Canephora, where they were insignificant (Fig. 5F and G). These spectral intervals are associated with caffeine, proteins, lipids, and carbohydrates in coffee. Since portable NIR does not cover the range of combination bands, the absorption bands of specific metabolites and other chemical assignments in coffee beans are linked to the third, second, and first overtones (Barbin et al., 2014; Ribeiro et al., 2011).

4. Conclusion

This paper has brought a scientific contribution with a new portable spectroscopy method for the direct chemical analysis of new roasted Brazilian Canephora coffees. The subject is relevant in coffee analytical chemistry because the use of portable spectrophotometer is particularly attractive for the industry. The new contribution was to apply and validate the use of portable spectroscopy in the certification of new Canephora coffees of high added value from specific origins registered with geographical indication. The study was meticulously done by using more than 500 real samples to exhaustively explore the use chemometric models with different authentication/classification purposes. Having a very large dataset was a plus of this study. The results indicated that the discrimination among the three Canephora coffees with geographical indication (Robusta Amazônico from indigenous and non-indigenous producers, and Conilon from Espírito Santo) as well as the discrimination between indigenous and non-indigenous producers of Robusta Amazônico was not a straightforward task.

Also, it was shown that portable NIR can provide similar classification results to benchtop NIR when used for the same discrimination purpose and same samples, using a cheaper spectrometer. The models could be applied to the final product according to the discrimination problem. A suggested strategy would be to use the multi-class PLS-DA model for a preliminary analysis of the sample and then certify that it is a GI Canephora using the binary PLS-DA models and DD-SIMCA, which is more appropriate for authentication. The portable spectroscopy is a trend, and this method brings a high possibility to be implemented in small cooperatives, industries, or control agencies in the future to dispose of a cheaper coffee certification method.

CRediT authorship contribution statement

Michel Rocha Baqueta: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Federico Marini:** Writing – review & editing. **Rodrigo Barros Rocha:** Methodology. **Patrícia Valderrama:** Conceptualization, Supervision, Formal analysis, Investigation, Methodology, Data curation, Writing – review & editing. **Juliana Azevedo Lima Pallone:** Conceptualization, Supervision, Funding acquisition, Investigation, Methodology, Data curation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the São Paulo Research Foundation (FAPESP) (grant #2019/21062-0 and grant #2022/04068-8; process #2022/03268-3), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (process 306606/2020-8, 402441/2022-2, and process 310982/2020-0), and Agência Brasileira de Desenvolvimento Industrial (process 23200.19/0070-2-01 and process 30.20.90.027.00.00). The authors would like to thank the EMBRAPA Rondônia, cooperatives and coffee producers for their invaluable help with the acquisition of the samples for the study.

References

Alves, E. A., de Souza, C. A., Rocha, R. B., Pereira, L. L., de Lima, P. P., & Lourenço, J. L. R. (2020). Efeito da fermentação sobre qualidade da bebida do café robusta (*Coffea canephora*) cultivado na Amazônia ocidenta. *Revista Ifes Ciência, 6*(3), 159–170. https://doi.org/10.36524/ric.v6i3.875

Baqueta, M. R., Alves, E. A., Valderrama, P., & Pallone, J. A. L. (2023). Brazilian Canephora coffee evaluation using NIR spectroscopy and discriminant chemometric techniques. *Journal of Food Composition and Analysis, 116*. https://doi.org/10.1016/j.jfca.2022.105065

Baqueta, M. R., Coqueiro, A., Março, P. H., Mandrone, M., Poli, F., & Valderrama, P. (2021). Integrated 1H NMR fingerprint with NIR spectroscopy, sensory properties, and quality parameters in a multi-block data analysis using ComDim to evaluate coffee blends. *Food Chemistry, 355*, Article 129618. https://doi.org/10.1016/j.foodchem.2021.129618

Baqueta, M. R., Coqueiro, A., Março, P. H., & Valderrama, P. (2021). Multivariate classification for the direct determination of cup profile in coffee blends via handheld near-infrared spectroscopy. *Talanta, 222*, Article 121526. https://doi.org/10.1016/J.TALANTA.2020.121526

Barbin, D. F., Felicio, A. L. de S. M., Sun, D.-W., Nixdorf, S. L., & Hirooka, E. Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food Research International*, *61*, 23–32. 10.1016/J.FOODRES.2014.01.005.

Beć, K. B., Grabska, J., & Huck, C. W. (2022). Miniaturized NIR Spectroscopy in Food Analysis and Quality Control: Promises, Challenges, and Perspectives. *Foods, 11*(10). https://doi.org/10.3390/foods11101465

Brazil. (2021a). *EMBRAPA. Embrapa launches documentary on the production of fine Robusta coffee in Amazonian terroir.*

Brazil. (2021b). *INPI. Instituto Nacional da Propriedade Industrial. Espírito Santo é reconhecido como indicação de procedência para o café conilon.* https://www.gov.br/inpi/pt-br/central-de-conteudo/noticias/espirito-santo-e-reconhecido-como-indicacao-de-procedencia-para-o-cafe-conilon.

Brazil. (2021c). *INPI. Instituto Nacional da Propriedade Industrial. Indicações Geográficas.* https://www.gov.br/inpi/pt-br/servicos/indicacoes-geograficas.

Brazil. (2021d). *INPI. Instituto Nacional da Propriedade Industrial. Matas de Rondônia é a mais nova Denominação de Origem para café.* https://www.gov.br/inpi/pt-br/central-de-conteudo/noticias/matas-de-rondonia-e-a-mais-nova-denominacao-de-origem-para-cafe.

Caporaso, N., Whitworth, M. B., Grebby, S., & Fisk, I. D. (2018). Non-destructive analysis of sucrose, caffeine and trigonelline on single green coffee beans by hyperspectral imaging. *Food Research International, 106*, 193–203. https://doi.org/10.1016/J.FOODRES.2017.12.031

Conab. (2023). Acompanhamento da Safra Brasileira de café - Safra 2023 - Primeiro levantamento. *Observatório Agrícola, 10*(1), 10–11. https://www.conab.gov.br/info-agro/safras/cafe.

Correia, R. M., Andrade, R., Tosato, F., Nascimento, M. T., Pereira, L. L., Araújo, J. B. S., … Romão, W. (2020). Analysis of Robusta coffee cultivated in agroforestry systems (AFS) by ESI-FT-ICR MS and portable NIR associated with sensory analysis. *Journal of Food Composition and Analysis, 94*, Article 103637. https://doi.org/10.1016/J.JFCA.2020.103637

Correia, R. M., Tosato, F., Domingos, E., Rodrigues, R. R. T., Aquino, L. F. M., Filgueiras, P. R., … Romão, W. (2018). Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta, 176*, 59–68. https://doi.org/10.1016/J.TALANTA.2017.08.009

dos Santos, V. J., Baqueta, M. R., Março, P. H., Valderrama, P., & Visentainer, J. V. (2021). Human Milk Lactation Phases Evaluation Through Handheld Near-Infrared Spectroscopy and Multivariate Classification. *Food Analytical Methods, 14*(5), 873–882. https://doi.org/10.1007/s12161-020-01924-y

Filete, C. A., Moreira, T. R., dos Santos, A. R., dos Santos Gomes, W., Guarçoni, R. C., Moreli, A. P., … Pereira, L. L. (2022). The New Standpoints for the Terroir of Coffea canephora from Southwestern Brazil: Edaphic and Sensorial Perspective. *Agronomy, 12*(8). https://doi.org/10.3390/agronomy12081931

Foschi, M., Biancolillo, A., Vellozzi, S., Marini, F., D'Archivio, A. A., & Boqué, R. (2021). Spectroscopic fingerprinting and chemometrics for the discrimination of Italian Emmer landraces. *Chemometrics and Intelligent Laboratory Systems, 215*, Article 104348. https://doi.org/10.1016/J.CHEMOLAB.2021.104348

Francisco, J. S., Dias, R. C. E., Alves, E. A., Rocha, R. B., Dalazen, J. R., Mori, A. L. B., & Benassi, M. de T. (2021). Natural intervarietal hybrids of coffea canephora have a high content of diterpenes. *Beverages*, *7*(4), 1–9. 10.3390/beverages7040077.

Geladi, P., MacDougall, D., & Martens, H. (1985). Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy, 39*(3), 491–500. https://doi.org/10.1366/0003702854248656

Giraudo, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E., & Geobaldo, F. (2019). Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. *Food Control, 99*, 137–145. https://doi.org/10.1016/J.FOODCONT.2018.12.033

Gomes, W. dos S., Pereira, L. L., Filete, C. A., Moreira, T. R., Guarçoni, R. C., Catarina da Silva Oliveira, E., Moreli, A. P., Guimarães, C. V., Simmer, M. M. B., Júnior, V. L., Romão, W., de Castro, E. V. R., & Partelli, F. L. (2022). Changes in the Chemical and Sensory Profile of Coffea canephora var. Conilon Promoted by Carbonic Maceration. *Agronomy*, *12*(10). 10.3390/agronomy12102265.

ICO. (2021). *International Coffee Organization. Coffee production report.* http://www.ico.org/prices/po-production.pdf.

ICO. (2022). *International Coffee Organization. Botanic aspects.* https://www.ico.org/pt/botanical_p.asp.

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*(1), 137–148.

Ky, C. L., Louarn, J., Dussert, S., Guyot, B., Hamon, S., & Noirot, M. (2001). Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild Coffea arabica L. and C. canephora P. accessions. *Food Chemistry*, *75*(2), 223–230. 10.1016/S0308-8146(01)00204-7.

Lemos, M. F., Perez, C., da Cunha, P. H. P., Filgueiras, P. R., Pereira, L. L., Almeida da Fonseca, A. F., … Scherer, R. (2020). Chemical and sensory profile of new genotypes of Brazilian Coffea canephora. *Food Chemistry, 310*(July 2019). https://doi.org/10.1016/j.foodchem.2019.125850

Liu, C., Yang, N., Yang, Q., Ayed, C., Linforth, R., & Fisk, I. D. (2019). Enhancing Robusta coffee aroma by modifying flavour precursors in the green coffee bean. *Food Chemistry, 281*(November 2018), 8–17. https://doi.org/10.1016/j.foodchem.2018.12.080

López, M. I., Callao, M. P., & Ruisánchez, I. (2015). A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. *Analytica Chimica Acta, 891*, 62–72. https://doi.org/10.1016/j.aca.2015.06.032

Mishra, P., Marini, F., Brouwer, B., Roger, J. M., Biancolillo, A., Woltering, E., & van Echtelt, E. H. (2021). Sequential fusion of information from two portable spectrometers for improved prediction of moisture and soluble solids content in pear fruit. *Talanta, 223*, Article 121733. https://doi.org/10.1016/J.TALANTA.2020.121733

Monteiro, P. I., Santos, J. S., Alvarenga Brizola, V. R., Pasini Deolindo, C. T., Koot, A., Boerrigter-Eenling, R., … Granato, D. (2018). Comparison between proton transfer reaction mass spectrometry and near infrared spectroscopy for the authentication of Brazilian coffee: A preliminary chemometric study. *Food Control, 91*, 276–283. https://doi.org/10.1016/J.FOODCONT.2018.04.009

Oliveira, E. C. da S., Guarçoni, R. C., de Castro, E. V. R., de Castro, M. G., & Pereira, L. L. (2020). Chemical and sensory perception of robusta coffees under wet processing. *Coffee Science*, *15*(1), 1–8. 10.25186/.v15i.1672.

Pomerantsev, A. L., & Rodionova, O. Y. (2018). Multiclass partial least squares discriminant analysis: Taking the right way—A critical tutorial. *Journal of Chemometrics, 32*(8). https://doi.org/10.1002/CEM.3030

Ribeiro, J. S., Ferreira, M. M. C., & Salva, T. J. G. (2011). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta, 83*(5), 1352–1358. https://doi.org/10.1016/J.TALANTA.2010.11.001

Rodionova, O. Y., & Pomerantsev, A. L. (2020). Chemometric tools for food fraud detection: The role of target class in non-targeted analysis. *Food Chemistry, 317*, Article 126448. https://doi.org/10.1016/J.FOODCHEM.2020.126448

Rodionova, O. Y., Oliveri, P., & Pomerantsev, A. L. (2016). Rigorous and compliant approaches to one-class classification. *Chemometrics and Intelligent Laboratory Systems, 159*, 89–96. https://doi.org/10.1016/J.CHEMOLAB.2016.10.002

Rodionova, O. Y., Titova, A. V., & Pomerantsev, A. L. (2016). Discriminant analysis is an inappropriate method of authentication. *TrAC Trends in Analytical Chemistry, 78*, 17–22. https://doi.org/10.1016/J.TRAC.2016.01.010

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry, 36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

Souza, C. A. de, Alves, E. A., Rocha, R. B., Espindula, M. C., & Teixeira, A. L. (2021). Perfis sensoriais dos cafeeiros cultivados na Amazônia Ocidental. In Fábio Luiz Partelli; Lucas Louzada Pereira (Ed.), *Café conilon: Conilon e Robusta no Brasil e no Mundo* (1st ed., pp. 187–198). Universidade Federal do Espírito Santo.

Souza, J. C., Pasquini, C., & Hespanhol, M. C. (2022). Feasibility of compact near-infrared spectrophotometers and multivariate data analysis to assess roasted ground coffee traits. *Food Control, 138*(April), Article 109041. https://doi.org/10.1016/j.foodcont.2022.109041

Zani Agnoletti, B., dos Santos Gomes, W., Falquetto de Oliveira, G., Henrique da Cunha, P., Helena Cassago Nascimento, M., Cunha Neto, Á., Louzada Pereira, L., Vinicius Ribeiro de Castro, E., Catarina da Silva Oliveira, E., & Roberto Filgueiras, P. (2022). Effect of fermentation on the quality of conilon coffee (Coffea canephora): Chemical and sensory aspects. *Microchemical Journal*, *182*, 107966. 10.1016/J.MICROC.2022.107966.

Zontov, Y. V., Rodionova, O. Y., Kucheryavskiy, S. V., & Pomerantsev, A. L. (2017). DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach. *Chemometrics and Intelligent Laboratory Systems, 167*, 23–28. https://doi.org/10.1016/J.CHEMOLAB.2017.05.010