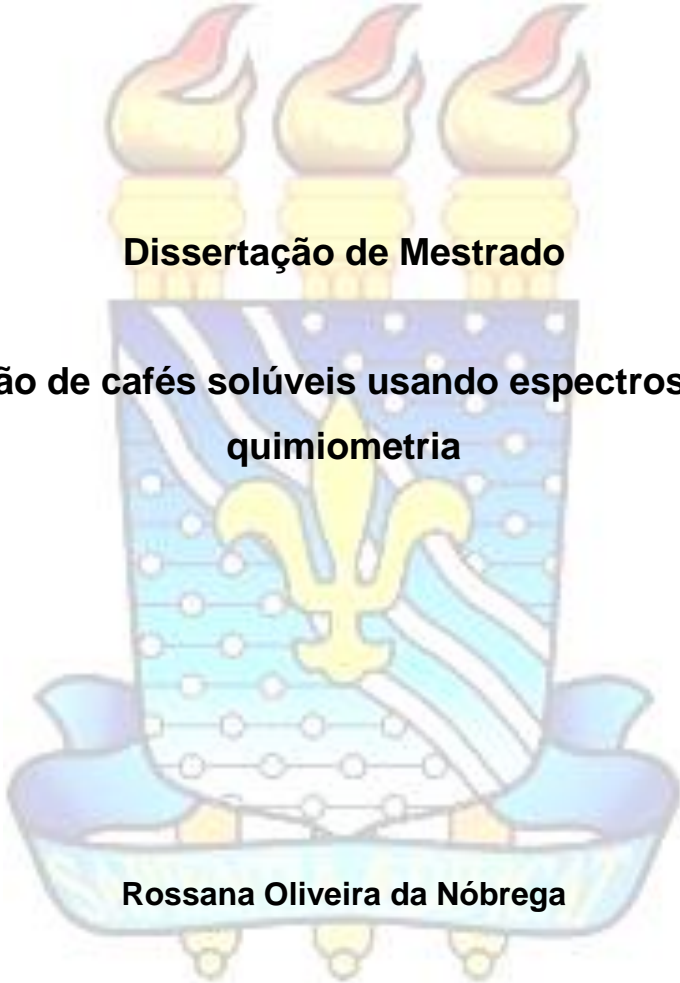




**Universidade Federal da Paraíba**  
**Centro de Ciências Exatas e da Natureza**  
**Departamento de Química**  
**Programa de Pós-Graduação em Química**



**Dissertação de Mestrado**

**Classificação de cafés solúveis usando espectroscopia NIR e  
quimiometria**

**Rossana Oliveira da Nóbrega**

**João Pessoa – PB – Brasil**  
**Julho de 2021**



**Universidade Federal da Paraíba**  
**Centro de Ciências Exatas e da Natureza**  
**Departamento de Química**  
**Programa de Pós-Graduação em Química**

## **Dissertação de Mestrado**

### **Classificação de cafés solúveis usando espectroscopia NIR e quimiometria**

**Rossana Oliveira da Nóbrega\***

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal da Paraíba como requisito para obtenção do título de Mestre em Química, área de concentração em Química Analítica.

**Orientador: Prof. Dr. Mário Cesar Ugulino de Araújo**

**Coorientador: Dr. David Douglas de Sousa Fernandes**

**\*Bolsista CAPES**

**João Pessoa – PB – Brasil**

**Julho de 2021**

**Catálogo na publicação**  
**Seção de Catálogo e Classificação**

N754c Nóbrega, Rossana Oliveira da.

Classificação de cafés solúveis usando espectroscopia NIR e quimiometria / Rossana Oliveira da Nóbrega. - João Pessoa, 2021.

77 f. : il.

Orientação: Mário César Ugulino de Araújo.

Coorientação: David Douglas de Sousa Fernandes.

Dissertação (Mestrado) - UFPB/CCEN.

1. Café. 2. Espectroscopia NIR. 3. DD-SIMCA. 4. PLS-DA. 5. iSPA-PLS-DA. I. Araújo, Mário César Ugulino de. II. Fernandes, David Douglas de Sousa. III. Título.

UFPB/BC

CDU 663.93(043)

# Classificação de cafés solúveis usando espectroscopia NIR e quimiometria.

Dissertação de Mestrado apresentada pela aluna Rossana Oliveira da Nóbrega e aprovada pela banca examinadora em 23 de julho de 2021.



Prof. Dr. Mário César Ugulino de Araújo  
Matrícula SIAPE 334937

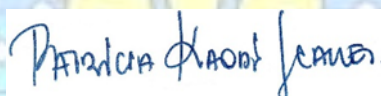
---

Prof. Dr. Mário César Ugulino de Araújo  
Orientador/Presidente  
(Departamento de Química – CCEN/UFPB)



---

Prof. Dr. David Douglas de Sousa Fernandes  
(2º. Orientador)  
(Departamento de Química – CCEN/UFPB)



---

Profa. Dra. Patricia Kaori Soares  
(Examinadora)  
Instituto de Química da UFRN



---

Prof. Dr. Márcio José Coelho de Pontes  
(Examinador)  
(Departamento de Química – CCEN/UFPB)

Assinaturas da Banca realizadas em modo Webconferência em 23/07/2021, digitalizadas e certificadas pelo Prof. Mário



César Ugulino de Araújo (SIAPE 1348826) em 23/07/2021 \_\_\_\_\_

## Agradecimentos

A Deus, meu eterno agradecimento pela vida, por cuidar tão bem de mim e ser meu sustento durante esta jornada.

À minha Mãe, Joana Darc, por todo amor, cuidado, que com muito incentivo e esforço me permitiu concluir mais uma etapa nesta carreira acadêmica.

À minha irmã Rafaela por estar sempre comigo, me apoiando e acompanhado cada conquista.

A Ronaldo, por todo companheirismo, amor e incentivo em todos os momentos vividos durante a graduação e agora no mestrado.

Ao meu professor e orientador, Dr. Mário César Ugulino de Araújo por todo comprometimento, orientação, confiança e ensinamentos durante o mestrado.

Ao meu coorientador, Dr. David Douglas, por sua eterna paciência, disponibilidade e pelo conhecimento compartilhado.

A Banca Examinadora pelo aceite do convite e pela contribuição nesse estudo.

A toda minha família e amigos que torceram e acreditaram em mim.

Aos professores da Pós-Graduação de Química da Universidade Federal da Paraíba, por compartilharem seus conhecimentos na minha formação acadêmica

Aos colegas do LAQA, pela ajuda, pelos momentos de descontração e troca de conhecimentos.

À Universidade Federal da Paraíba.

A CAPES pela bolsa concedida.

# Sumário

<b>RESUMO</b> .....	<b>VIII</b>
<b>ABSTRACT</b> .....	<b>IX</b>
<b>LISTA DE FIGURAS</b> .....	<b>X</b>
<b>LISTA DE TABELAS</b> .....	<b>XII</b>
<b>LISTAS DE ABREVIATURAS E SIGLAS</b> .....	<b>XIII</b>
<b>1 INTRODUÇÃO</b> .....	<b>15</b>
<b>2 OBJETIVOS</b> .....	<b>18</b>
<b>2.1 OBJETIVO GERAL</b> .....	<b>18</b>
<b>2.2 OBJETIVOS ESPECÍFICOS</b> .....	<b>18</b>
<b>3 FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>20</b>
<b>3.1 CAFÉ SOLÚVEL</b> .....	<b>20</b>
3.1.1 ETAPAS DE PRODUÇÃO DE CAFÉ SOLÚVEL.....	22
3.1.2 PROCESSO DE DESCAFEINAÇÃO.....	25
<b>3.2 METODOLOGIAS ANALÍTICAS PARA AUTENTICIDADE DO CAFÉ</b> .....	<b>27</b>
<b>3.3 ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO</b> .....	<b>27</b>
<b>3.4 QUIMIOMETRIA</b> .....	<b>30</b>
3.4.1 PRÉ-PROCESSAMENTO DOS DADOS.....	31
3.4.2 TÉCNICAS DE RECONHECIMENTO DE PADRÕES.....	32
3.4.3 ANÁLISE POR COMPONENTES PRINCIPAIS.....	33
3.4.4 MODELAGEM INDEPENDENTE E FLEXÍVEL POR ANALOGIA DE CLASSES DIRECIONADA PELOS DADOS (DD-SIMCA).....	34
3.4.5 ANÁLISE DISCRIMINANTE POR MÍNIMOS QUADRADOS PARCIAIS (PLS-DA).....	36
3.4.6 SELEÇÃO DE INTERVALOS DE VARIÁVEIS.....	37
3.4.6.1 ALGORITMO DAS PROJEÇÕES SUCESSIVAS PARA SELEÇÃO DE INTERVALOS EM ANÁLISE DISCRIMINANTE POR MÍNIMOS QUADRADOS PARCIAIS (ISPA-PLS-DA).....	39
3.4.7 VALIDAÇÃO DE MODELOS DE CLASSIFICAÇÃO.....	39
<b>4 METODOLOGIA</b> .....	<b>43</b>
<b>4.1 AQUISIÇÃO DE AMOSTRAS</b> .....	<b>43</b>
<b>4.2 AQUISIÇÃO DOS ESPECTROS NIR</b> .....	<b>43</b>
<b>4.3 PROCEDIMENTO QUIMIOMÉTRICO E SOFTWARE</b> .....	<b>44</b>
<b>5 RESULTADOS E DISCUSSÃO</b> .....	<b>48</b>
<b>5.1 CLASSIFICAÇÃO DE CAFÉS SOLÚVEIS DESCAFEINADO USANDO ESPECTROSCOPIA NIR E DD-SIMCA</b>	<b>48</b>

5.1.1 ESPECTROS NIR BANCADA E PORTÁTIL.....	48
5.1.1.2 PRÉ-PROCESSAMENTOS .....	49
5.1.1.3 ANÁLISE POR COMPONENTES PRINCIPAIS .....	51
5.1.1.4 CLASSIFICAÇÃO .....	52
<b>5.2 CLASSIFICAÇÃO DE CAFÉS SOLÚVEIS TRADICIONAL E EXTRAFORTE USANDO ESPECTROSCOPIA NIR ASSOCIADA A PLS-DA E ISPA-PLS-DA.....</b>	<b>57</b>
5.2.1 ESPECTROSCOPIA NIR BANCADA E PORTÁTIL.....	57
5.2.1.1 ANÁLISE POR COMPONENTES PRINCIPAIS .....	58
5.2.1.2 CLASSIFICAÇÃO .....	59
<b>6 CONCLUSÕES E PERSPECTIVAS.....</b>	<b>66</b>
<b>6.1 CONCLUSÕES .....</b>	<b>66</b>
<b>6.2 PERSPECTIVAS.....</b>	<b>66</b>
<b>REFERÊNCIAS.....</b>	<b>68</b>
<b>ANEXO 1.....</b>	<b>76</b>
<b>ANEXO 2.....</b>	<b>77</b>



## RESUMO

O café solúvel é uma bebida obtida a partir da desidratação do extrato de café torrado, cuja qualidade é caracterizada por suas propriedades organolépticas, que depende de diversos fatores, como a variabilidade do grão, extração da cafeína, moagem, processo de torra, entre outros. Contudo, a qualidade desse produto deve ser monitorada para garantir a segurança alimentar e aceitação por parte dos consumidores e órgãos reguladores. Assim, este trabalho objetivou desenvolver metodologias analíticas verdes baseadas em Espectroscopia do Infravermelho Próximo (NIR: *near infrared spectroscopy*) de bancada e portátil e quimiometria para avaliação da conformidade dos cafés solúveis comerciais, a respeito do tipo e grau de torra. A primeira abordagem do estudo consiste na classificação das amostras de cafés solúveis descafeinado em relação ao café solúvel regular (com cafeína), utilizando a Modelagem Independente e Flexível por Analogia de Classes direcionada pelos dados (DD-SIMCA: *Data-Driven – Soft Independent Modelling of Class Analogy*), em ambos os equipamentos NIR de bancada e NIR portátil. Nessa abordagem, os resultados obtidos foram 100% de sensibilidade, especificidade e acurácia. Na segunda abordagem, os cafés solúveis regulares foram classificados a respeito do grau de torra, tradicional e extraforte, utilizando Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA: *Partial Least Squares Discriminant Analysis*) e o Algoritmo de Projeções Sucessivas para seleção de intervalos em PLS-DA (*iSPA-PLS-DA: Successive Projections Algorithm for Interval Selection in Partial Least-Squares*). Para os espectros do NIR de bancada, obteve-se o melhor resultado com o método *iSPA-PLS-DA*, ao usar o pré-processamento média móvel com correção de espalhamento multiplicativo (MM+MSC), alcançando 96,7% de taxa de acurácia na discriminação das amostras em suas respectivas classes. No caso do NIR portátil, o melhor desempenho de classificação foi observado para *iSPA-PLS-DA* com o pré-processamento média móvel e correção de linha de base *offset* (MM+BO), com 98% de acurácia. Portanto, este estudo mostrou o potencial da espectroscopia NIR juntamente com as ferramentas quimiométricas de classificação para análise rápida, não destrutiva e direta do café solúvel, e que podem ser úteis para avaliação dos parâmetros de qualidade durante o processo industrial, bem como do produto acabado.

**Palavras-chave:** Café, espectroscopia NIR, DD-SIMCA, PLS-DA, *iSPA-PLS-DA*.

## ABSTRACT

Soluble coffee is a drink obtained from the dehydration of roasted coffee extract, the quality of which is characterized by its organoleptic properties, which depend on several factors, such as grain variability, caffeine extraction, grinding, roasting process, among others. However, the quality of this product must be monitored to ensure food safety and acceptance by consumers and regulatory bodies. Thus, this work aimed to develop green analytical methodologies based on bench and portable Near Infrared Spectroscopy (NIR: near infrared spectroscopy) and chemometric to assess the conformity of commercial soluble coffees, regarding the type and degree of roast. The first approach of the study is to classify the samples of soluble decaffeinated coffees in relation to regular soluble coffee (with caffeine), using Data-Driven Independent and Flexible Class Analogy Modeling (DD-SIMCA: Data-Driven – Soft Independent Modeling of Class Analogy) on both bench-top and portable NIR equipment. In this approach, the results obtained were 100% sensitivity, specificity, and accuracy. In the second approach, regular soluble coffees were classified with respect to roasting, traditional and extra-strong, using Partial Least Squares Discriminant Analysis (PLS-DA) and the Successive Projections Algorithm for selection of intervals in PLS-DA (iSPA-PLS-DA: Successive Projections Algorithm for Interval Selection in Partial Least-Squares). For bench-top NIR spectra, the best result was obtained with the iSPA-PLS-DA method, when using the moving average pre-processing with multiplicative scatter correction (MM+MSC), reaching 96.7% of accuracy rate in the discrimination of samples in their respective classes. In the case of portable NIR, the best sorting performance was observed for iSPA-PLS-DA with moving average pre-processing and baseline offset correction (MM+BO), with 98% accuracy. Therefore, this study showed the potential of NIR spectroscopy together with chemometric classification tools for rapid, non-destructive, and direct analysis of soluble coffee, which can be useful for evaluating quality parameters during the industrial process, as well as the finished product.

**Keywords:** Coffee, NIR spectroscopy, DD-SIMCA, PLS-DA, iSPA-PLS-DA.

## LISTA DE FIGURAS

<b>Figura 1.</b> Estruturas dos compostos químicos presentes no café (a) Cafeína (b) trigonelina (c) ácidos clorogênicos. ....	21
<b>Figura 2.</b> Etapas de produção de café solúvel. Fonte: Adaptada de ABICS (2020). 22	
<b>Figura 3.</b> Representação da organização dos dados para construção do modelo PLS-DA. Fonte: Adaptada de Santana <i>et al.</i> (2020).....	37
<b>Figura 4.</b> Representação gráfica dos parâmetros de avaliação do modelo de classificação.....	41
<b>Figura 5.</b> Representação gráfica dos parâmetros de avaliação do modelo de classificação para o PLS-DA e iSPA-PLS-DA.....	41
<b>Figura 6.</b> Diagrama esquemático do processo de pesagem, obtenção dos espectros NIR e ferramentas quimiométricas.....	44
<b>Figura 7.</b> Representação da classificação de cafés solúveis em dois níveis.....	44
<b>Figura 8.</b> Conjuntos de amostras de treinamento e teste em cada classe para os dois modelos de classificação.....	45
<b>Figura 9.</b> Espectros NIR (a,c) originais (b,d) cortados para as 103 amostras de café solúveis regulares (verde) e descafeínados (azul).....	48
<b>Figura 10.</b> Espectros (a) originais e pré-processados das classes regular (verde) e descafeinado (azul) com: (a) Suavização Média Móvel (MM) com janela de 9 pontos e MSC, (c) MM e BO, (d) MM e SNV, (e) MM+LBC+BO, (f) derivação Savitzky-Golay (dSG) com primeira derivada, janela de 23 pontos e polinômio de segunda ordem (●: Regular, ●: Descafeinado).....	50
<b>Figura 11.</b> Espectros (a) originais e pré-processados das classes regular (verde) e descafeinado (azul) com: (b) Suavização Média Móvel (MM) com janela de 7 pontos e MSC, (c) MM e BO, (d) MM e SNV, (e) MM+LBC+BO, (f) derivação Savitzky-Golay (dX_7) com primeira derivada, janela de 7 pontos e polinômio de segunda ordem (●: Regular, ●: Descafeinado).....	51
<b>Figura 12.</b> Gráfico dos escores de PC1 versus PC2 para as 103 amostras de cafés solúveis utilizando os espectros pré-processados: (a) Suavização Média Móvel (MM) e MSC, (b) MM e BO, (c) MM e SNV, (d) MM+LBC+BO, (e) derivação Savitzky-Golay (dSG) com primeira derivada e polinômio de segunda ordem (●: Regular, ●: Descafeinado).....	52

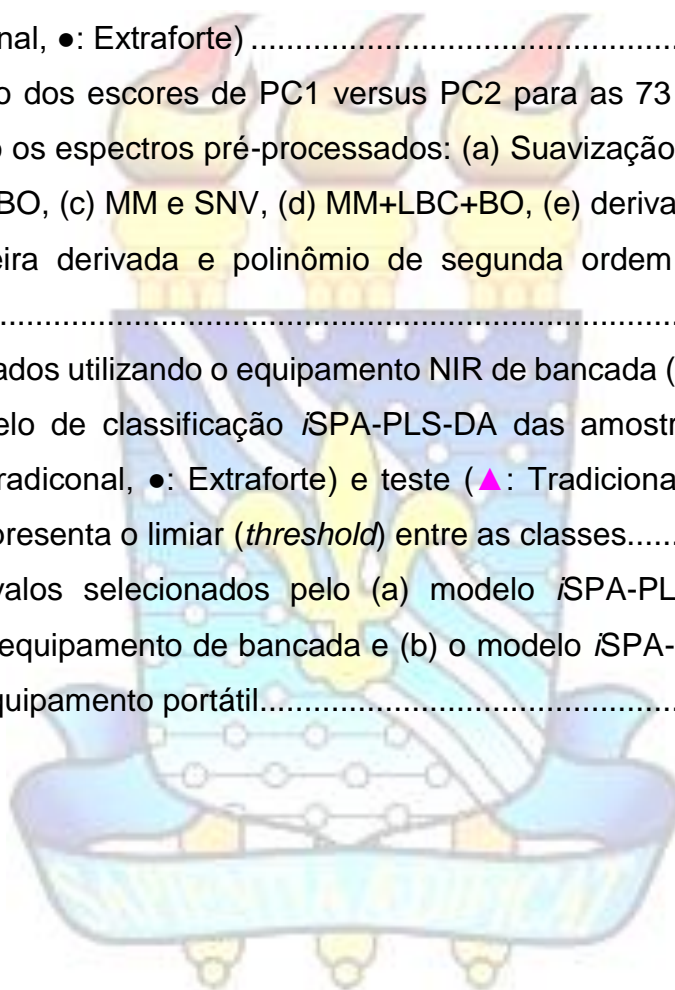
**Figura 13.** Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) para o algoritmo DD-SIMCA desenvolvidos no NIR de bancada (a, b) e portátil (c, d) (●) descafeinado (●) regular. As linhas verdes representam a borda entre as amostras regulares e extremas; a linha vermelha é a fronteira da área de aceitação .....55

**Figura 14.** Espectros pré-processados das classes tradicional e extraforte com: (a) Suavização Média Móvel (MM) e MSC, (b) MM e BO, (c) MM e SNV, (d) MM+LBC+BO, (e) derivação Savitzky-Golay (dSG) com primeira derivada e polinômio de segunda ordem (●: Tradicional, ●: Extraforte) .....57

**Figura 15.** Gráfico dos escores de PC1 versus PC2 para as 73 amostras de cafés solúveis utilizando os espectros pré-processados: (a) Suavização Média Móvel (MM) e MSC, (b) MM e BO, (c) MM e SNV, (d) MM+LBC+BO, (e) derivação Savitzky-Golay (dSG) com primeira derivada e polinômio de segunda ordem (●: Tradicional, ●: Extraforte).....58

**Figura 16.** Resultados utilizando o equipamento NIR de bancada (a) e portátil (b) para previsão do modelo de classificação *i*SPA-PLS-DA das amostras do conjunto de treinamento (●: Tradicional, ●: Extraforte) e teste (▲: Tradicional, ▲: Extraforte). A linha vermelha representa o limiar (*threshold*) entre as classes.....62

**Figura 17.** Intervalos selecionados pelo (a) modelo *i*SPA-PLS-DA ( $w=20$ ) com MM+MSC para o equipamento de bancada e (b) o modelo *i*SPA-PLS-DA ( $w=5$ ) com MM+BO para o equipamento portátil.....63



## LISTA DE TABELAS

<b>Tabela 1</b> – Teores dos compostos químicos do café verde e torrado (g/100g) para as espécies <i>Coffea arabica</i> e <i>canephora</i> (Robusta) .....	21
<b>Tabela 2</b> – Resultados dos parâmetros de desempenho sensibilidade, especificidade e acurácia para DD-SIMCA nos conjuntos de treinamento e teste .....	53
<b>Tabela 3</b> – Matriz de confusão obtido pelo algoritmo PLS-DA para os espectros NIR de bancada das 73 amostras de cafés solúveis comerciais.....	60
<b>Tabela 4</b> – Matriz de confusão obtido pelo algoritmo PLS-DA para os espectros NIR portátil das 73 amostras de cafés solúveis comerciais.....	60
<b>Tabela 5</b> – Resultados da classificação para o grau de torra em cafés solúveis usando espectros NIR de bancada pré-processados e algoritmo de classificação iSPA-PLS-DA .....	64



## LISTAS DE ABREVIATURAS E SIGLAS

ABIC	Associação Brasileira da Indústria de Café
ABICS	Associação Brasileira da Indústria de Café Solúvel
AE	Acetato de Etila
ANVISA	Agência Nacional de Vigilância Sanitária
BO	Correção de linha de base <i>offset</i> , do inglês <i>baseline offset correction</i> .
DCM	Diclorometano
DD-SIMCA	Modelagem Independente e Flexível por Analogia de Classes direcionada pelos dados, do inglês <i>Data Driven-Soft Independent Modeling of Class Analogy</i>
DRIFTS	Espectroscopia de refletância difusa no infravermelho médio por transformada de Fourier, do inglês <i>Diffuse Reflectance Infrared Fourier Transform Spectroscopy</i> )
EF	Extraforte
iSPA-PLS-DA	Algoritmo das Projeções Sucessivas para Seleção de Intervalos em Análise Discriminante por Mínimos Quadrados Parciais, do inglês <i>Successive Projections Algorithm for Interval Selection in Partial Least-Squares</i>
KS	Algoritmo Kennard-Stone
LDA	Análise discriminante linear, do inglês <i>Linear Discriminant Analysis</i>
LBC	Correção de Linha de Base Linear, do inglês <i>linear baseline correction</i>
MM	Média Móvel, do inglês <i>moving mean</i>
MSC	Correção de Espalhamento Multiplicativo, do inglês <i>multiplicative scatter correction</i>
NIR	Infravermelho próximo, do inglês <i>Near Infrared</i>
OCP-PLS	Mínimos quadrados parciais de uma classe, do inglês <i>One-Class Partial Least Squares</i>
PCA	Análise por Componentes Principais, do inglês <i>Principal Component Analysis</i>
PLS-DA	Análise Discriminante por Mínimos Quadrados Parciais, do inglês <i>Partial Least Squares Discriminant Analysis</i>
SCAA	Associação de Cafés Especiais da América, do inglês <i>Specialty Coffee Association of America</i>
SIMCA	Modelagem Independente e Flexível por Analogia de Classe, do inglês <i>Soft Independent Modeling of Class Analogy</i>
SNV	Varição Normal Padrão, do inglês <i>Standard Normal Variate</i>
TD	Tradicional

*Capítulo 1*  
**INTRODUÇÃO**

---

## 1 INTRODUÇÃO

O Brasil é líder mundial de produção e exportação de café solúvel. Em 2020, registrou um crescimento de 2,4% nas exportações em comparação ao ano de 2019. No total, 4,1 milhões de sacas de café solúvel com 60 kg foram exportadas para 102 países, tendo principal destino o Estados Unidos como primeiro e maior consumidor mundial (ABICS, 2021).

O café solúvel é uma bebida proveniente da desidratação do extrato de café exclusivamente torrado a partir dos grãos das espécies *Coffea arabica* e *Coffea canéfora*, conhecida como Conilon ou Robusta (GHIRIŞAN; MICLĂUŞ, 2017; ABICS, 2020). As formas de comercialização de café solúvel diferem em função do método de secagem, sendo: *Spray dried*, aglomerado, *Freeze dried* ou liofilizado. Além disso, pode ser categorizado de acordo com a “ausência” de cafeína ou descafeinado e com a presença de cafeína de diferentes graus de torra, sendo os mais comuns os cafés tradicionais e extraforte (SOUSA *et al.*, 2010; FRANCA, 2016).

O café solúvel descafeinado é geralmente usado por pessoas que querem evitar ou limitar absorção desse alcaloide pelo organismo. Estudos demonstram que a absorção de cafeína em excesso pode ocasionar efeitos indesejáveis, como taquicardia, ansiedade, tremores, distúrbios gastrointestinais e insônia (FRANCA, 2016). Mas, mesmo em pequenas dosagens também pode provocar reações adversas em pessoas que apresentam alta sensibilidade (MEINHART *et al.*, 2010). Diante desses efeitos indesejáveis à saúde, torna-se importante avaliar a conformidade do café descafeinado para segurança alimentar do consumidor. Segundo a Agência Nacional de Vigilância Sanitária (ANVISA, 2005), o teor de cafeína em café solúvel descafeinado não deve ser superior a 0,3% m/m.

Um outro aspecto que influencia na qualidade do café é o processo de torra, pois esse procedimento é responsável por conceber características finais ao café. O nível de torra (clara, média ou escura) evidencia as propriedades organolépticas, mas oculta impurezas e grãos de baixa qualidade, como também a presença de possíveis adulterantes (cascas, paus, milho, entre outros) que causam impacto na qualidade da bebida do café (CORREIA *et al.*, 2018; ESTEKI; SHAHSAVARI; SIMAL-GANDARA, 2018).

A avaliação dos parâmetros estabelecidos para a obtenção da qualidade do café solúvel pelos órgãos de fiscalização e comercialização tem sido realizada com base na identificação e quantificação das espécies químicas ou da determinação dos parâmetros físico-químicos e sensoriais (ANVISA, 2005; INSTITUTO ADOLFO LUTZ, 2008). Para avaliação do nível de torra é utilizada a análise de cor por meio do Sistema Agron/SCAA como critério para padronizar o produto final e interromper o processo de torra (AGTRON, 2010; ABIC, 2020). Entretanto, a maioria destes métodos adotados pelas legislações possuem algumas desvantagens, tais como: elevado custo, lentidão de análise, utilização de solventes e subjetividade da análise visual e sensorial.

Nesse contexto, a Espectroscopia de Infravermelho Próximo (do inglês, *Near Infrared Spectroscopy* - NIRS) emergiu como uma alternativa promissora e confiável para análise da qualidade dos produtos alimentícios e tem sido amplamente utilizada para obter informações rápidas sobre a composição química e propriedades relacionadas ao café (BARBIN *et al.*, 2014). Além disso, esta técnica apresenta vantagem por não ser destrutiva, não necessitar do uso de reagentes químicos e evitar muitas etapas de pré-tratamentos da amostra, gerando baixa quantidade de resíduos, em concordância com os princípios da Química Verde.

Alguns estudos relataram o potencial da espectroscopia NIR junto as técnicas quimiométricas na capacidade de classificação/autenticação das amostras de cafés torrados e moídos de diferentes qualidades e variedades (ESTEBAN-DÍEZ, 2007; SANTOS *et al.*, 2012; BAQUETA *et al.*, 2019). No entanto, não há relatos sobre aplicação da espectroscopia NIR e nem a utilização de dispositivos portáteis para monitoramento da qualidade de café solúveis em relação ao tipo (regulares e descafeínados) com diferentes lotes e variedades de torrefação, moagem e secagem. Desta forma, aponta-se como alternativa viável o desenvolvimento de metodologias analíticas verdes baseadas em espectroscopia NIR e ferramentas quimiométricas para classificação de amostras de cafés solúveis em relação a presença de cafeína (regulares e descafeínados) e ao seu grau de torra (tradicional e extraforte) sem qualquer preparação de amostra, em contraposição aos seus respectivos métodos oficiais.

*Capítulo 2*  
**OBJETIVOS**

---

## 2 OBJETIVOS

### 2.1 Objetivo Geral

Desenvolver metodologias analíticas rápidas, de baixo custo, não destrutivas e não invasivas que sigam os Princípios da Química Verde (*Principles of green chemistry*) utilizando espectroscopia NIR e ferramentas quimiométricas de classificação multivariada para avaliação dos cafés solúveis comerciais em relação a presença de cafeína (regulares e descafeinado) e ao seu grau de torra.

### 2.2 Objetivos Específicos

- Empregar a espectroscopia NIR, em instrumentos de bancada e portáteis de baixo custo em conjunto ao DD-SIMCA para classificar cafés solúveis em virtude da presença de cafeína;
- Utilizar espectroscopia NIR em instrumentos de bancada e portáteis de baixo custo para discriminar cafés solúveis de acordo com o grau de torrefação (tradicional e extraforte), empregando modelos PLS-DA com e sem seleção de variáveis (SPA-PLS-DA);

*Capítulo 3*  
**FUNDAMENTAÇÃO  
TEÓRICA**

---

## 3 FUNDAMENTAÇÃO TEÓRICA

### 3.1 Café Solúvel

O café solúvel foi inventado pelo químico japonês Satori Kato no ano de 1901, em Chicago (ABICS, 2021). Em 1937, a empresa Nestle desenvolveu um pó de café liofilizado para aumentar o consumo de café e solucionar o excesso de oferta de matéria prima no mercado brasileiro. Somente em 1960, a indústria de café solúvel foi implantada no Brasil. A partir disso, a industrialização motivou a exportação do produto e gerou grandes oportunidades. Desde então, o Brasil tem permanecido na liderança mundial de produção e exportação do setor (ABICS, 2020).

O café solúvel é uma bebida obtida a partir da desidratação do extrato aquoso dos grãos torrados das variedades arábica e robusta. No entanto, essas duas espécies são as mais importantes economicamente e, portanto, as mais cultivadas no mundo, sendo diferenciadas não apenas em relação as suas propriedades botânicas, físicas, organolépticas e química, mas também em termos de valor comercial (BERTONE *et al.*, 2016; GHIRIŞAN; MICLĂUŞ, 2017; ABICS, 2020).

A espécie arábica é originada da África Oriental, cultivada em regiões com altitudes acima de 800 metros e com condições climáticas favoráveis. Em função disso, produzem bebidas de qualidade superior, possuindo sabor menos amargo, de aroma intenso e menor teor de cafeína, sendo utilizadas em fabricações de cafés especiais e gourmet com preços mais elevados aos do robusta (ASSIS; OLIVEIRA; SENA, 2017).

A espécie *canephora* (robusta) é a mais popular para produção da bebida devido a sua atratividade econômica e o alto teor de sólidos solúveis na sua composição, apresentando maior rendimento industrial. Portanto, é muito utilizado em misturas de grãos (blends) e na indústria de café solúvel, a fim de manter uniformidade no produto, obtendo uma bebida de acordo com as preferências ou hábitos do consumidor e de menor custo (FERNANDES *et al.*, 2003; MALAQUIAS; CELESTINO; XAVIER, 2018).

Na **Tabela 1**, é apresentada a composição química dos grãos das variedades arábica e robusta antes e depois da torrefação. Em comparação com o grão verde robusta, o grão verde arábica têm alto valor de sacarose, lipídio e trigonelina, mas

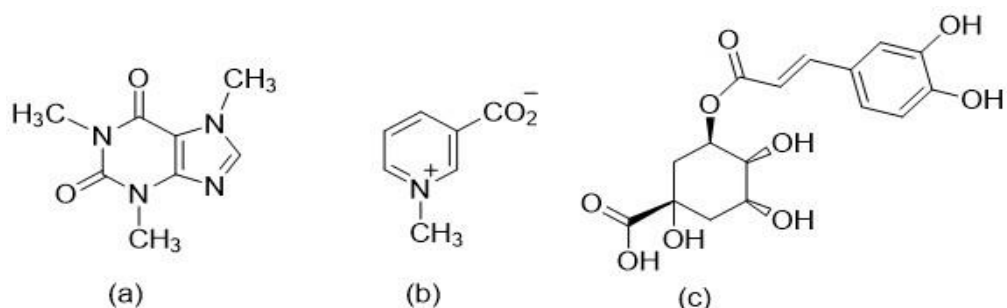
possui menor teor de cafeína e ácido clorogênico. A torrefação influencia bastante na composição química dos grãos, resultando na diminuição dos teores de sacarose, proteína, aminoácidos, ácidos clorogênicos e água. Simultaneamente, com a formação de melanoidinas que é responsável pelo sabor e mudança de cor do café torrado (WANG; LIM, 2015).

**Tabela 1** – Teores dos compostos químicos do café verde e torrado (g/100g) para as espécies *Coffea arabica* e *canephora* (Robusta)

Compostos químicos	<i>Coffea arabica</i>		<i>Coffea canephora</i>	
	Verde	Torrado	Verde	Torrado
Sacarose	6,2 – 8,4	0,0	0,9 – 4,8	0,0
Lipídios	15,0 – 18,0	17,0	8,0 – 12,0	11,0
Proteínas	8,5 – 12,0	7,5	8,5 – 12,0	7,5
Aminoácidos	0,2 – 0,8	0,0	0,2 – 0,8	0,0
Ácidos clorogênicos	6,7 – 9,2	2,5	7,1 – 12,1	3,8
Cafeína	0,8 – 1,4	1,3	1,7 – 2,4	2,4
Trigonelina	0,6 – 1,2	1,0	0,3 – 0,9	0,7
Água	8,0 – 12,0	0,0 – 5,0	8,0 – 12,0	0,0 – 5,0
Melanoidinas	0,0	23,0	0,0	23,0

Fonte: Adaptada de Wang e Lim (2015).

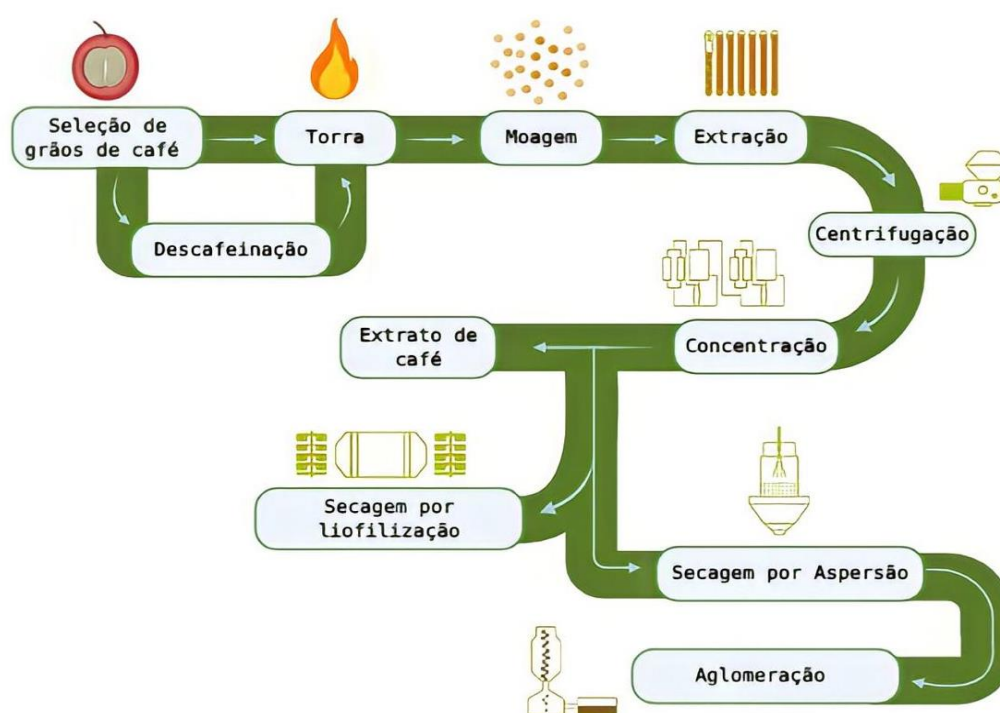
Entre esses compostos presentes no café, destacam-se a cafeína, trigonelina e ácidos clorogênicos, apresentados na **Figura 1**. No qual contribuem significativamente para as características da bebida, como sabor e aroma.



**Figura 1.** Estruturas dos compostos químicos presentes no café (a) Cafeína (b) trigonelina (c) ácidos clorogênicos.

### 3.1.1 Etapas de produção de café solúvel

O processamento industrial do café solúvel envolve uma série de etapas como mostra a **Figura 2** (ABICS, 2020). Os grãos de café das variedades arábica e robusta são selecionados de acordo com as características sensoriais desejadas (aroma e sabor). Entretanto, os cafés solúveis apresentam maior proporção de grãos da variedade robusta fazendo parte de 85% da matéria prima consumida, representando em média até 30% do extrato de sólidos solúveis, o que torna o Robusta insubstituível economicamente na fabricação de café solúvel (ABICS, 2018).



**Figura 2.** Etapas de produção de café solúvel. Fonte: Adaptada de ABICS (2020) (Criada com BioRender.com)

Os grãos de café cru são separados das impurezas e enviados aos torrefadores. A torrefação é uma das etapas mais complexas e importantes do processo de fabricação do café. Essa etapa está relacionada com as propriedades organolépticas (sabor, aroma e cor) devido as mudanças físicas, químicas e sensoriais durante o processo de torra, impactando positivamente na qualidade do café até certo ponto, pois uma torra acentuada compromete o produto (WEI; TANOKURA, 2015; POISSON *et al.*, 2017). Essa qualidade também depende inicialmente da composição química que está intimamente associada à espécie do grão, método de colheita, condições de secagem e armazenamento (CAPORASO *et al.*, 2018).

A torra consiste em um processo térmico a seco em função de duas variáveis, tempo e temperatura, variando de acordo com as características desejadas para o produto e, para esse efeito, a cor é um critério bastante utilizado para avaliar o nível de torra (SCHENKER; ROTHGEB, 2017). Este parâmetro é medido pela variável Agtron, por meio de discos Agtron da *Specialty Coffee Association of América* – SCAA, com escala numérica de #25 a #95, que representa torra mais escura e a mais clara, respectivamente.

Segundo a ABIC, para os cafés comercializados no Brasil, são obtidas torras moderadamente claras (Agtron 75) a moderadamente escuras (Agtron 45), essa mudança de coloração se deve a formação das melanoidinas causadas pela reação de Maillard, a pirólise dos compostos orgânicos e caramelização dos açúcares (CID; PEÑA, 2016).

É importante ressaltar que a preferência da torra (clara, média e escura) depende do local e dos consumidores. No Brasil, os consumidores apreciam o café com torra média a moderadamente escura, porém a torra mais intensa (torra escura) pode intencionalmente mascarar grãos de baixa qualidade, resultando em uma bebida de sabor amargo e forte (CORREIA *et al.*, 2018; MALAQUIAS; CELESTINO; XAVIER, 2018). Portanto, no comércio brasileiro, os cafés com processo de torra mais intenso são comumente rotulados como extraforte (RIBEIRO, 2009; KALSCHNE *et al.*, 2021).

O processo térmico ocorre em três estágios, que são: a secagem, pirólise e o resfriamento. A princípio, as primeiras mudanças observadas ocorrem acima de 50°C quando a desnaturação da proteína, e à medida que a temperatura aumenta até atingir 100°C, ocorre a evaporação da água livre obtendo grãos mais pigmentados devido as reações de Maillard e degradação de Strecker, resultando em diversos compostos, dentre eles as melanoidinas (CID; PEÑA, 2016).

Acerca de 150°C, uma quantidade de compostos voláteis como vapor de água, dióxido de carbono e monóxido de carbono são liberados decorrentes das reações de degradação térmica. Em temperaturas entre 180°C e 200°C é marcado pelo segundo estágio, pirólise, com a ruptura do endosperma, fissuração dos grãos e a formação de produtos aromáticos responsáveis pelas propriedades organolépticas que são características do café. Na terceira fase, a interrupção do processo térmico, de modo a evitar a carbonização do grão, então, são retirados do torrador e, em seguida, são resfriados através de uma corrente de ar fria (CID; PEÑA, 2016).

Após a torrefação, os grãos torrados são moídos em partículas menores, através de um esforço mecânico para aumentar a superfície de contato e permitir a extração dos compostos solúveis de maneira mais rápida e apropriada para preparo da bebida de café (BLITTERSDORFF; KLATT, 2017). O nível de moagem e a dimensão das partículas influenciam diretamente na extração e na qualidade da bebida, uma moagem de partículas muito fina pode produzir uma bebida amarga por causa da extração excessiva, enquanto uma moagem de partículas maiores pode prejudicar a eficiência da extração, originando uma bebida insípida (SUNARHARUM; WILLIAMS; SMYTH, 2014).

Os grãos torrados e moídos são submetidos a extração aquosa sob pressão e alta temperatura para remover substâncias solúveis do grão por infusão. Este processo é realizado por percoladores de aço inoxidável, por meio de tubos que possibilitam a passagem da água de uma coluna para a outra sucessivamente, o que promove o enriquecimento de sólidos solúveis em relação a matéria-prima. Antes de ir para a etapa de concentração, o extrato é centrifugado para retirada de partículas insolúveis (ABICS, 2020).

Neste estágio, parte da água contida no extrato de café é removida através da utilização de evaporadores ou crio-concentradores, podendo alcançar até 60% de sólidos solúveis para facilitar a secagem para obtenção do pó solúvel (ABICS, 2020). O extrato de café concentrado pode ser vendido para indústrias que fabricam sobremesas, bebidas geladas ou refrigerantes a base de café, como também para a produção de café solúvel que pode ser realizada por dois principais processos de secagem: liofilização (*freeze-drying*) e aspersão (*spray-drying*).

Na secagem por aspersão, o extrato de café concentrado é pulverizado no topo de uma torre de secagem com fluxo de ar quente na forma de pequenas esferas, e a água contida nessas gotículas do extrato são evaporadas durante sua queda até a base da torre, resultando em um pó como produto final. Como etapa adicional do processo, o café solúvel em pó pode ser transformado em café solúvel aglomerado, através do contato do vapor d'água com o pó, de modo que sejam umedecidos o suficiente para aderirem umas às outras e se agregarem produzindo partículas granuladas. No entanto, a etapa de aglomeração tem como objetivo promover a solubilidade do café em pó na água (GHIRIŞAN; MICLĂUŞ, 2017).

Por outro lado, a liofilização ocorre em temperaturas inferiores a  $-40^{\circ}\text{C}$ , temperatura em que o extrato de café é concentrado e congelado na forma de cristais,

e a seguir, são triturados em moinhos especiais e depositados em bandejas para serem conduzidos em uma câmara de vácuo, em que acontece a desidratação, fazendo com que a água seja removida passando do estado sólido diretamente para o estado gasoso por sublimação.

Contudo, o café liofilizado preserva ao máximo as características e os compostos aromáticos, resultando em uma bebida encorpada com qualidade superior, embora seja um processo de alto custo operacional (ANANDHARAMAKRISHNAN, 2019; ABICS, 2020).

### 3.1.2 Processo de descafeinação

A cafeína (1,3,7-trimetilxantina) e seus efeitos são intrínsecos ao café, sejam positivos ou negativos, dependendo da condição individual e das quantidades ingeridas pelo consumidor. Embora a cafeína seja uma substância psicoativa responsável por efeitos positivos, como aumento de concentração e bom desempenho nas atividades, o consumo excessivo pode causar problemas de saúde, incluindo azia, insônia, taquicardia e ansiedade. Por outro lado, mesmo em pequenas quantidades, pode causar reações adversas em pessoas que apresentam alta sensibilidade a esse alcaloide (PIETSCH, 2017).

Diante desses efeitos indesejáveis à saúde, consumidores sensíveis optam por consumir café descafeinado. Alguns trabalhos relatam que mesmo com baixo teor de cafeína, como encontradas no café descafeinado, podem apresentar propriedades psicoativa (MEINHART *et al.*, 2010; HASKELL *et al.*, 2008). Por isso, a avaliação da conformidade de cafeína declarado no rótulo do produto tem se tornado o parâmetro importante para segurança do consumidor. A Agência Nacional de Vigilância Sanitária (ANVISA), através da Portaria Nº 277, de 22 de setembro de 2005, estabelece o teor de cafeína máximo de 0,3% para café solúvel descafeinado.

É importante destacar que a fabricação de café solúvel descafeinado é realizada a partir do processo de descafeinação dos grãos de café cru através de solventes orgânicos, dióxido de carbono supercrítico e/ou imersão em água, antes da torra. Em seguida, são enviados para as torrefadoras e passa por todas as etapas citadas anteriormente (**Figura 2**), chegando até 97% da remoção da cafeína (FRANCA, 2016; ZABOT, 2020).

O processo de descafeinação por solventes orgânicos pode ocorrer através de uma grande variedade de solventes, mas, geralmente, apenas dois são adequados: Diclorometano (DCM) e o Acetato de Etila (AE). Primeiro, os grãos são contactados com vapor e água para aumentar o teor de umidade de aproximadamente de 10 para 25%, ou mesmo 40%. Após isto, na coluna de extração, o solvente circula em volta dos grãos embebidos em água para eliminar a cafeína, e essa etapa é realizada múltiplas vezes, por um período de até 10h. O solvente orgânico é drenado e evaporado dos grãos verdes e, posteriormente, é vaporizado na torra (PIETSCH, 2017; FRANCA, 2016).

A razão pelo uso da água como agente de extração ao invés de solventes orgânicos é devido ao processo ser menos agressivo a saúde e também em consequência da qualidade da bebida que é afetada pela perda dos compostos aromáticos do café. Neste processo, o aumento da solubilidade da cafeína na água ocorre em função da temperatura e, para isso, a extração tem sido realizada com água quente à pressão atmosférica. Os grãos verdes são submersos em água concentrada com componentes hidrossolúveis do café para manter a qualidade do sabor e aroma e, em seguida, o extrato saturado com cafeína percorre continuamente pelo filtro de carvão ativado que retém apenas a cafeína, esse processo acontece de forma cíclica durante 8 a 10 horas (FRANCA, 2016).

No método de dióxido de carbono supercrítico é utilizado  $\text{CO}_2$ , substância de alta pureza, para remover seletivamente a cafeína, sem que haja mudanças bruscas na composição dos grãos. Os grãos são embebidos em água na presença de  $\text{CO}_2$  supercrítico, em temperaturas e pressões superiores as condições críticas (ponto crítico). Conforme o  $\text{CO}_2$  supercrítico circula, a cafeína é extraída do grão e retida no carvão ativado. Os grãos são separados do carvão por meio de peneiras vibratórias (ZABOT, 2020; FRANCA, 2016).

Em ambos os métodos de descafeinação, a etapa final é marcada pela secagem dos grãos de café descafeinado com o objetivo de recuperar o teor de umidade inicial. Estes métodos podem influenciar nos componentes do aroma do café. Mas, desconsiderando esse aspecto, a indústria escolhe seu próprio processo de descafeinação, tecnologias de processo e métodos de degustação para propor seu próprio sabor da marca (PIETSCH, 2017).

### 3.2 Metodologias analíticas para autenticidade do café

As técnicas utilizadas no controle de qualidade e autenticidade de café ocorrem por meio de métodos convencionais, tais como físico-químicos (ANVISA, 2005; INSTITUTO ADOLFO LUTZ, 2008), sensoriais (ESTEBAN-DÍEZ; GONZÁLEZ-SÁIZ; PIZARRO, 2004), cromatográficos (ROSTAGNO *et al.*, 2011; YASHIN, 2017), e para avaliação e monitoramento do nível de torra é realizado a análise de cor através de discos Agrtron/SCAA de diferentes tonalidades de marrom (CID; PEÑA, 2016). Essas metodologias adotadas por agências nacionais e internacionais baseiam-se no uso de instrumentos sofisticados e de alto custo, na maioria das vezes requer pré-tratamento de amostras, elevado consumo de tempo e subjetividade da análise sensorial e visual.

No entanto, há relatos na literatura sobre o desenvolvimento de métodos rápidos e de baixo custo, os quais são capazes de classificar o produto em termos de autenticidade, composição química, origem geográfica, adulterações, entre outros. Esses métodos incluem espectroscopia de absorção molecular do ultravioleta e visível (UV-Vis) (SOUTO *et al.*, 2015a), espectroscopia NIR (SANTOS *et al.*, 2012; MONTEIRO *et al.*, 2018), imagens digitais (SOUTO *et al.*, 2015b) e voltametria (MORAIS *et al.*, 2019).

A utilização desses métodos alternativos na indústria cafeeira pode auxiliar no controle de qualidade do produto final, além de ajudar na otimização do processo de produção dos cafés, gerando uma economia significativa para a indústria (BAQUETÁ *et al.*, 2021).

Portanto, a Espectroscopia de Infravermelho Próximo de bancada e portátil associadas a ferramentas quimiométricas podem ser uma alternativa vantajosa para análises diretas, rápidas e de baixo custo para classificação de cafés solúveis comerciais.

### 3.3 Espectroscopia no Infravermelho Próximo

A Espectroscopia de Infravermelho Próximo (NIRS) é uma técnica analítica baseada na espectroscopia vibracional proveniente da interação da radiação eletromagnética com os constituintes da amostra na faixa de comprimento de onda entre 750 e 2500 nm ( $13300$  a  $4000$   $\text{cm}^{-1}$ ), e mede a energia absorvida pelas moléculas, de forma que os grupos funcionais presentes nessas moléculas contêm

informações analíticas quantitativas e qualitativas a respeito da amostra (ZAREEF *et al.*, 2020).

Os principais fenômenos responsáveis pelo surgimento das bandas de absorção nos espectros NIR são os sobretoms e combinações de modos vibracionais, entretanto, para que ocorra a absorção da radiação NIR, esses fenômenos devem ser acompanhados por uma variação do momento de dipolo. A intensidade de uma determinada banda de absorção depende da mudança de dipolo durante a transição vibracional na região do infravermelho e do seu grau de anarmonicidade associado a grupos funcionais envolvendo o átomo de hidrogênio (o mais leve) e outro elemento (mais pesado), como carbono, enxofre, nitrogênio e oxigênio, o que corresponde às respectivas ligações, C-H, S-H, N-H, O-H (PASQUINI, 2003; PASQUINI, 2018).

Os sistemas moleculares vibracionais assemelham-se melhor ao modelo de um oscilador anarmônico, em que os níveis de energia vibracional não são igualmente espaçados. Esse modelo pode prevê a ocorrência de transições entre os estados vibracionais com  $\Delta v > \pm 2$ , e conforme o número quântico vibracional ( $v$ ) aumenta, a probabilidade de transições diminui. Essas transições são designadas de transições de sobretom, seus valores de frequência correspondem a duas ou três vezes a frequência vibracional fundamental, que são conhecidos como primeiro e segundo sobretom, respectivamente (TIBOLA *et al.*, 2018).

Bandas de combinação ocorrem quando a interação dos modos vibracionais com mudanças simultâneas de energia, cuja frequência será igual a soma das frequências, capazes de excitar cada um dos modos vibracionais, localizada na faixa característica do infravermelho próximo de 1900 a 2500 nm. Devido à complexidade, as bandas de absorção no NIR possuem características típicas, como largas, sobrepostas e de intensidade fraca, cerca de 10 a 1000 vezes inferiores que suas bandas fundamentais na região do infravermelho médio, dificultando a interpretação dos dados espectrais.

Entretanto, a utilização da técnica NIR requer a combinação com ferramentas quimiométricas para o tratamento dos espectros, proporcionando métodos diretos, não destrutivos e de baixo custo, sem necessidade do consumo de reagentes químicos e pré-tratamento da amostra, estando de acordo com os princípios da química verde (TIBOLA *et al.*, 2018).

Nas últimas décadas, a utilização das análises de espectroscopia NIR para avaliação da composição química e propriedades sensoriais relacionadas a qualidade

do café tem se mostrado técnicas confiáveis e promissoras (BARBIN *et al.*, 2014; BAQUETA; COQUEIRO; VALDERRAMA, 2019). No entanto, poucos estudos têm enfatizado o uso da espectroscopia NIR para classificação de cafés solúveis, de acordo com o grau de torra e o teor de cafeína, e apenas alguns desses estudos são baseados em amostras de café torrado e moído, como apresentados a seguir.

A espectroscopia NIR foi explorada como alternativa aos métodos convencionais para detecção de componentes adulterantes em café instantâneo. Amostras puras e adulteradas com glicose, amido ou chicória foram discriminadas pela técnica de análise discriminante linear (LDA, do inglês: *Linear Discriminant Analysis*), precedida pela análise de componentes principais (PCA, do inglês: *Principal Component Analysis*) para redução da dimensionalidade. Quatro Componentes Principais (PCs, do inglês: *Principal Component*) foram suficientes para uma boa discriminação, alcançando uma taxa de classificação de 98% (BRIANDET; KEMSLEY; WILSON, 1996).

Ribeiro (2010), em estudo com 29 amostras comerciais de café torrado e moído, utilizando espectroscopia de refletância difusa no infravermelho médio (DRIFTS, do inglês: *Diffuse Reflectance Infrared Fourier Transform Spectroscopy*) juntamente com PCA e Regressão por Mínimos Quadrados Parciais para Análise Discriminante (PLS-DA, do inglês: *Partial Least Squares for Discriminant Analysis*), obteve uma clara discriminação entre as classes de café descafeinado e de café tradicional e torra escura.

Santos *et al.* (2012) propuseram o uso de espectroscopia NIR para classificar três variedades de café torrado e moído, sendo *gourmet*, tradicional e descafeinado provenientes de supermercados e indústrias de torrefação das cidades brasileiras. Os autores empregaram dois métodos de classificação: Modelagem Independente e Flexível por Analogia de Classe (SIMCA, do inglês: *Soft Independent Modeling of Class Analogy*) e PLS-DA que obtiveram 90,8% e 96,3% de classificação correta, respectivamente. Concluindo que essa técnica pode eventualmente substituir os seus respectivos métodos oficiais.

A espectroscopia portátil na região do infravermelho próximo (microNIR) foi avaliada quanto a seu potencial no controle de qualidade do café arábica. A técnica MicroNIR combinada com PCA e Regressão por Mínimos Quadrados Parciais (RPLS, do inglês: *Partial Least Squares*) mostrou-se eficaz na identificação de adulterantes

(milho e cascas/palitos) e na discriminação dos níveis de torra (claro, médio e escuro) entre os blends de café arábica/robusta (CORREIA *et al.*, 2018).

Baqueta *et al.* (2020) investigaram o potencial da espectroscopia NIR portátil associado ao PLS-DA, para avaliar os perfis de xícara em *blends* de café torrado e moído de diferentes variedades e graus de torra. As amostras foram classificadas como *blend* duro/riado torra média (classe 1), duro/riado torra escura (classe 2) e riado/rio torra média escura (classe 3). A metodologia proposta obteve 100% de classificação correta para as amostras da classe 1, enquanto para ambas as classes 2 e 3 foi superior a 90% nos conjuntos de previsão (BAQUETA, 2021).

### 3.4 Quimiometria

Em 1971, o químico orgânico Svante Wold criou o termo “quimiometria”. Mais à frente, em 1972, ele publicou o primeiro artigo que mencionava a palavra “quimiometria”. Dois anos depois, em parceria com Bruce Kowalski, fundaram a *International Chemometric Society*, cuja função seria fortalecer o elo entre as áreas de química, matemática e estatística. Desde então, a quimiometria vem se desenvolvendo e tem sido aplicada em diversas áreas, o que facilita a extração de informações relevantes de uma série de dados químicos para fins qualitativos ou quantitativos (CORNEJO-BÁEZ *et al.*, 2020; FERREIRA, 2016).

Atualmente, houve um crescimento de publicações científicas voltadas para o desenvolvimento de métodos de classificação multivariada, com foco na autenticação de alimentos através de técnicas instrumentais que produzem grandes quantidades de informações (variáveis) para um grande grupo de amostras (objetos) em um curto período de tempo (JIMÉNEZ-CARVELO; CUADROS-RODRÍGUEZ, 2019). Isto leva a existência de uma matriz de dados multivariados, que requer o emprego das ferramentas estatísticas e matemáticas para extrair o máximo de informação útil para alcançar resultados e, conseqüentemente, tomadas de decisões destes dados (FERREIRA, 2016).

Em alguns casos, antes da análise quimiométrica, os dados devem ser pré-processados e o pré-processamento ideal deve ser selecionado para que a técnica quimiométrica seja aplicada.

### 3.4.1 Pré-processamento dos dados

O pré-processamento dos dados tem a finalidade de minimizar a variância aleatória, como também as fontes de variação indesejáveis que não estão relacionadas com a propriedade de interesse. As análises de dados espectroscópicos, em especial os espectros NIR, exibem variações sistemáticas indesejáveis, que são afetadas pelas propriedades físicas da amostra, tais como: diferenças no tamanho, distribuição das partículas e variações de densidade da amostra. Como consequência, essa variação deve ser eliminada ou minimizada, pois a não correlação com as características químicas da amostra pode prejudicar a capacidade preditiva dos modelos desenvolvidos. Portanto, a utilização de pré-processamento adequado, reduz tais efeitos físicos a fim de modelar o efeito químico de forma mais eficaz (TIBOLA *et al.*, 2018; HUANG; ROMERO-TORRES; MOSHGBAR, 2010).

Os métodos de pré-processamento podem ser empregados nas amostras (linhas) e nas variáveis (colunas) da matriz de dados (FERREIRA, 2016). Os tipos de pré-processamento aplicados nas amostras do presente estudo são: Suavização por média móvel (MM, do inglês: *moving mean*), correção de linha de base linear (LBC, do inglês: *linear baseline correction*) e offset (do inglês: *baseline offset*), variação normal padrão (SNV, do inglês: *Standard Normal Variation*), correção multiplicativa de sinal (MSC, do inglês: *Multiplicative Scattering Correction*) e derivada Savitzky-Golay. E direcionada as variáveis, a centralização dos dados na média.

No entanto, existem diferentes ferramentas de pré-processamento para minimizar variações aleatórias (ruído experimental) e informações sistemáticas indesejáveis. O ruído instrumental pode ser reduzido através de técnicas de alisamento, sendo utilizado para esse fim, a suavização por média móvel. Enquanto isso, as variações sistemáticas podem ser eliminadas ou reduzidas por meio de derivadas de Savitzky-Golay, SNV, MSC e correção de linha de base.

As derivadas são ferramentas matemáticas úteis para corrigir o desvio da linha de base dos espectros, pois remove deslocamentos constantes e lineares, além de resolver sobreposição de picos, possibilitando uma melhora na resolução dos picos existentes nos sinais originais. A primeira derivada é aplicada para eliminar desvios de linha-base aditivos, já a segunda derivada elimina também efeitos multiplicativos (MANLEY; BAETEN, 2018; TIBOLA *et al.*, 2018).

O MSC é um método utilizado para corrigir problemas relacionados ao espalhamento de luz decorrente de fenômenos físicos. A normalização MSC é obtida por regressão linear para se estimar o coeficiente linear e angular entre o espectro médio e o espectro a ser corrigido. O espectro corrigido é calculado pela subtração do coeficiente linear para cada comprimento de onda original (para correção dos efeitos aditivos) e, em seguida, divide-se este pelo coeficiente angular (para correção de efeitos multiplicativos) (FERREIRA, 2016; TIBOLA *et al.*, 2018).

O SNV assemelha-se ao MSC, uma vez que ambos removem efeitos multiplicativos e interferências de espalhamento, porém, a correção por SNV se dá por meio da normalização de cada espectro pelo seu respectivo desvio padrão, não necessitando de um espectro médio de referência (OLIVERI; SIMONETTI, 2016; MANLEY; BAETEN, 2018).

Já a centralização dos dados na média tem sido muito utilizada em dados espectrais. Nesse pré-processamento, é calculado o valor médio para cada variável, e a partir disso, é subtraído para cada elemento da respectiva coluna. Cada variável passa a ter média zero, ou seja, o centro das coordenadas é movido da origem natural para a média multivariada dos dados (TIBOLA *et al.*, 2018).

### **3.4.2 Técnicas de reconhecimento de padrões**

As técnicas de reconhecimento de padrões estão divididas em dois tipos: os métodos supervisionados e os métodos não supervisionados. A análise não supervisionada busca identificar similaridade e diferença entre amostras de um conjunto total de dados, sem que haja necessidade de rotular amostras ou objetos (FERREIRA, 2016). Alguns algoritmos como a análise de componente principal (PCA, do inglês: *Principal Component Analysis*) e análise por agrupamento hierárquico (HCA, do inglês: *Hierarchical Cluster Analysis*) foram construídos para facilitar a interpretação dos resultados, inclusive na elaboração de gráficos, para extrair informações contidas no conjunto de dados multivariados.

Os métodos de classificação supervisionada usam as informações de cada amostra da classe pré-estabelecida para a construção dos modelos de classificação com o objetivo de classificar novas amostras em uma das classes conhecidas (BERRUETA; ALONSO-SALCES; HÉBERGER, 2007; TIBOLA *et al.*, 2018).

Neste trabalho, foram empregadas técnicas de reconhecimento de padrões não supervisionada, tais como: PCA e métodos supervisionados que se dividem em métodos de modelagem de classe, como Modelagem Independente e Flexível por Analogia de Classes direcionado pelos dados (DD-SIMCA, do inglês: *Data-Driven – Soft Independent Modelling of Class Analogy*) e métodos discriminatórios como Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA, do inglês: *Partial Least Squares Discriminant Analysis*) e o Algoritmo de Projeções Sucessivas para seleção de intervalos em PLS-DA (iSPA-PLS-DA, do inglês: *Successive Projections Algorithm for Interval Selection in Partial Least-Squares*).

### 3.4.3 Análise por componentes principais

A análise de componente principal é considerada a ferramenta mais usada para análise exploratória de dados, sendo comumente utilizada na primeira etapa analítica, com a finalidade de investigar a presença de padrões nos dados medidos. A PCA efetua a projeção dos dados multivariados em um espaço de menor dimensão, reduzindo assim a dimensionalidade do conjunto de dados e preservando a maior parte da variância presente nos dados originais (FERREIRA, 2016; SILVA *et al.*, 2017).

A PCA transforma as variáveis originais, possivelmente correlacionadas, em variáveis linearmente não correlacionadas, ortogonais e independentes entre si, chamada de componentes principais (PCs). Essa transformação leva em conta que as primeiras componentes principais representam a maior variância possível. Deve-se ressaltar que a primeira componente principal (PC1) explica a máxima variância dos dados originais, de modo que a segunda componente (PC2) será ortogonal à PC1 e descreve maior variância do que PC3, e as PCs sucessivas não estão correlacionadas as anteriores, de tal maneira que essas componentes são calculadas em ordem decrescente de importância (GRANATO *et al.*, 2018; SHAH; CHAUHAN; CHAUDHURY, 2020).

Matematicamente, o algoritmo PCA efetua a decomposição da matriz de dados originais ou pré-processada, matriz  $\mathbf{X}$ , em duas matrizes menores: escores ( $\mathbf{T}$ ) e pesos ( $\mathbf{P}$ ) que correspondem, respectivamente, as projeções ortogonais das amostras no eixo de cada componente principal (PC) e a contribuição de cada variável original em

relação a sua respectiva PC, mais uma matriz de resíduos (**E**), contendo a variância não explicada pelo modelo (**Equação 1**) (MILANI *et al.*, 2018).

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (\text{eq.1})$$

Graficamente, na PCA, algumas ferramentas de diagnóstico são utilizadas para investigar os conjuntos de dados, como: gráfico de resíduos que auxilia na determinação do número ótimo de PCs, gráficos de escores que revelam possível formação de agrupamentos entre as amostras e identificam amostras anômalas, e gráfico de pesos que possibilita a identificação de variáveis influentes na formação de agrupamentos no gráfico de escores, como também aponta variáveis anômalas (SOUZA; POPPI, 2012; ALLEGRETA *et al.*, 2020).

#### **3.4.4 Modelagem independente e flexível por analogia de classes direcionada pelos dados (DD-SIMCA)**

O modelo DD-SIMCA é um classificador pertencente ao método de classe única ou *one class* utilizado como uma ferramenta de reconhecimento de padrão para autenticação de alimentos, fármacos, combustíveis e entre outros. Portanto, apenas a classe “alvo” é modelada, e quaisquer amostras que não incluídas a esta classe são consideradas amostras extremas, adulteradas e/ou falsificadas, dependendo da abordagem dos dados examinados.

Este modelo foi gerado a partir da modificação do SIMCA original, diferenciando no cálculo do limite de aceitação que delimitam a classe autêntica, sem necessidade de amostras alternativas no conjunto de treinamento (RODIONOVA; TITOVA; POMERANTSEV, 2016; NEVES; POPPI, 2020).

A classe alvo é construída de acordo com as propriedades de suas amostras representantes. Essas propriedades, também chamadas de impressões digitais, que são sinais analíticos gerados por equipamentos, como: espectroscopia, cromatografia ou outras técnicas analíticas. O resultado proveniente dessas propriedades origina uma Matriz principal (**X<sub>0</sub>**), que representa um conjunto de dados das amostras autênticas, de forma que são devidamente pré-processadas e centradas na média.

O método DD-SIMCA é baseado em várias etapas. Na primeira etapa, ocorre a decomposição da matriz dos dados de treinamento da classe autêntica utilizando a

PCA (**Equação 1**) (RODIONOVA; TITOVA; POMERANTSEV, 2016). A quantidade de componentes principais (PCs) determina a complexidade do modelo que influencia diretamente na qualidade da classificação, levando em consideração que quanto maior os números de PCs, maior a variância explicada dos dados de treinamento, porém, a inclusão desnecessária de PCs pode afetar a capacidade preditiva do modelo, devido a inserção de ruído irrelevante. Portanto, a complexidade do modelo é determinada frequentemente pelo princípio da parcimônia (RODIONOVA *et al.*, 2018).

Na segunda etapa, são estimadas a distância dos escores (SD, do inglês: *score distance*),  $h_i$ , e a distância ortogonal (OD, do inglês: *orthogonal distance*),  $v_i$ , para cada amostra pertencente ao conjunto de treinamento baseado nos resultados da decomposição de PCA, conforme as **Equações 2 e 3** (ZONTOV *et al.*, 2017).

$$h_i = \mathbf{t}_i^t (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{t}_i \quad (\text{eq. 2})$$

$$v_i = \sum_{j=1}^J e_{ij}^2 \quad (\text{eq. 3})$$

A distância dos escores SD representa uma amostra dentro do espaço definido pelas componentes principais, e a distância ortogonal OD é determinada como a soma dos quadrados dos resíduos associados ao modelo PCA, que caracteriza a distância da amostra ao espaço que delimita a classe alvo. Esses parâmetros são descritos nas referências RODIONOVA; TITOVA; POMERANTSEV (2016) e ZONTOV *et al.* (2017).

A combinação destas distâncias SD e OD são descritas em termos da variável  $c$ , distância global (**Equação 4**), assumindo valores que seguem uma distribuição qui-quadrado. (RODIONOVA; TITOVA; POMERANTSEV, 2016).

$$c = N_h \frac{h}{h_0} + N_v \frac{v}{v_0} \sim \alpha \chi^2(N_h + N_v) \quad (\text{eq. 4})$$

Sendo que os parâmetros  $h_0$  e  $v_0$  são fatores de normalização das distâncias dos escores e ortogonal,  $N_h$  e  $N_v$  são os número de graus de liberdade para SD e OD, respectivamente (RODIONOVA; TITOVA; POMERANTSEV, 2016).

Desse modo, a terceira etapa define a área de aceitação ou limite para a classe autêntica de acordo com o valor do erro tipo I ( $\alpha$ ), taxa de falsa rejeição, quando as amostras da classe alvo são consideradas não pertencentes a ela. A regra da decisão

é determinada pela condição  $c \leq c_{crit}$ . Caso uma amostra possua medida de distância global menor que a distância crítica, esta é identificada como pertencente a classe alvo (NEVES; POPPI, 2020).

$$c \leq c_{crit}(\alpha) = \chi^{-2}(1 - \alpha, N_h + N_v) \quad (\text{eq. 5})$$

O modelo construído é representado por uma área de aceitação da distância ortogonal versus distância dos escores (gráfico de aceitação), determinado pelo valor  $\alpha$ . Cada amostra do conjunto de treinamento é identificada por sua posição nas novas coordenadas do gráfico de aceitação, sendo classificada como pertencente ou não a classe alvo (MÁQUINA *et al.*, 2020).

Na etapa de classificação do conjunto de teste, novas amostras (pertencentes ou não a classe alvo) são empregadas ao modelo construído. O resultado dessa classificação fornece o gráfico de aceitação correspondente. Com o emprego de uma classe alternativa, há possibilidade do cálculo do erro tipo II ( $\beta$ ), que é a taxa de falsa aceitação quando as amostras não-autênticas são classificadas na classe alvo como autêntica (NEVES; POPPI, 2020).

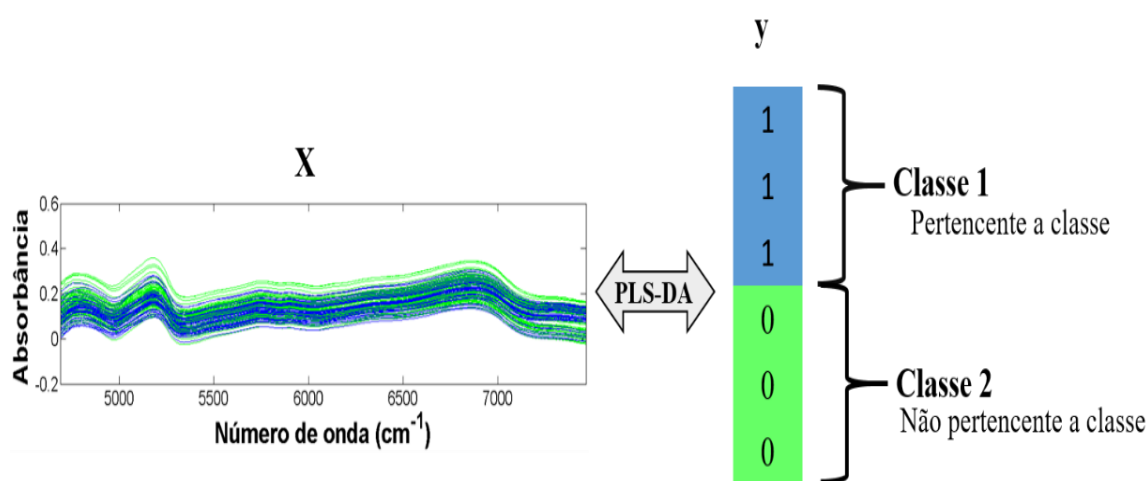
### 3.4.5 Análise discriminante por mínimos quadrados parciais (PLS-DA)

O PLS-DA é um método discriminante de dados multivariados baseado no algoritmo de regressão por mínimos quadrados parciais (PLS), que busca uma relação entre as variáveis independentes, ou respostas instrumentais ( $\mathbf{X}$ ), e as variáveis dependentes (matriz  $\mathbf{Y}$  ou vetor  $\mathbf{y}$ ) que codificam a classe da amostra, podendo assumir os valores 1 ou 0, atribuindo se a amostra pertence ou não à respectiva classe (SANTANA *et al.*, 2020).

A organização dos dados para a construção do modelo PLS-DA, em que apenas duas classes são modeladas, é ilustrado pela **Figura 3**. A matriz  $\mathbf{X}$  é representada pelos espectros, sendo que as linhas correspondem as amostras e as colunas a absorvância para cada número de onda. O vetor  $\mathbf{y}$  é construído com a mesma ordem, em que as primeiras amostras pertencem a classe 1 e as demais amostras a classe 0, ou vice-versa. Já para conjunto de dados com mais de duas classes, requer a utilização do algoritmo PLS2-DA, em que o número das

propriedades associadas as amostras são dispostas em uma matriz  $Y$ , e cada coluna corresponde a uma classe (SANTANA *et al.*, 2020; BERETON; LLOYD, 2014).

As matrizes  $X$  e  $Y$  são decompostas em um produto de outras matrizes designadas escores e pesos, levando em conta a informação na matriz  $Y$ . O PLS busca a máxima covariância entre as matrizes  $X$  e  $Y$  e, assim, gera as variáveis latentes (ZHANG *et al.*, 2020). Desse modo, a escolha do número de variáveis latentes é realizada em função da validação cruzada para as amostras de treinamento. Na predição do modelo, o limiar adotado define qual classe a amostra pertencerá (MARQUETTI, 2016).



**Figura 3.** Representação da organização dos dados para construção do modelo PLS-DA.  
Fonte: Adaptada de Santana *et al.* (2020).

Normalmente, a escolha do limiar entre as duas classes é de 0,5. Nessas condições, caso uma amostra seja prevista acima desse limite, ela será atribuída pertencente a classe (Classe 1), caso contrário, tal amostra não pertencerá a classe (Classe 2). Mas, esse valor limite também pode ser calculado de acordo com a estatística Bayesiana, uma vez que estando entre as classes é selecionado quando as duas distribuições ( $y$  previsto para cada classe) se cruzam em um determinado ponto, a fim de minimizar os erros, tanto de falso positivo quanto falso negativo (LUCA *et al.*, 2016; SANTANA *et al.*, 2020).

### 3.4.6 Seleção de intervalos de variáveis

Na análise multivariada, a etapa de seleção de variáveis é considerada importante para remoção de variáveis não informativas (como regiões ruidosas e

saturadas que não estão relacionadas às propriedades de interesse) e variáveis redundantes que possuem a mesma informação química, sendo prejudicial ao modelo.

Destaca-se que nestes casos, o procedimento de seleção de variáveis, geralmente produz melhores resultados, com a identificação de um subconjunto de dados representativos que gera os menores erros possíveis, resultando em modelos com melhor capacidade preditiva em análises discriminantes ou quantitativas (KREPPER, 2018).

No contexto de classificação multivariada, os modelos de reconhecimento de padrões supervisionados utilizam toda a informação espectral, como é feito, por exemplo, em PLS-DA (*full-spectrum* PLS-DA). No entanto, devido a redundância dos dados espectrais, tem sido muito útil o uso do PLS-DA combinado com técnicas de seleção de intervalos de variáveis, tais como: o algoritmo para seleção de intervalo único em Análise Discriminante por Mínimos Quadrados Parciais (*i*PLS-DA, do inglês: *single interval selection in Discriminant Analysis by Partial Least Squares*), em que apenas um intervalo é selecionado, e Algoritmo das Projeções Sucessivas para Seleção de Intervalos em Análise Discriminante por Mínimos Quadrados Parciais (*i*SPA-PLS-DA, do inglês: *Successive Projections Algorithm for Interval Selection in Partial Least-Squares*), quando diferentes combinações de intervalos podem ser utilizadas.

Alguns trabalhos descritos na literatura têm demonstrado o bom desempenho do *i*SPA-PLS-DA. Fernandes e colaboradores (2016) em dois estudos de caso: classificação de óleos vegetais comestíveis utilizando voltametria de onda quadrada e classificação de cepas bacterianas com base em imagens digitais obteve resultados em termos de taxa de classificação correta iguais ou superiores quando comparados com os modelos PLS-DA e *i*PLS-DA.

Pereira e colaboradores (2021) empregaram a espectroscopia NIR para autenticação de leite de cabra adulterado com leite de vaca através dos métodos Mínimos quadrados parciais de uma classe (OCP-PLS, do inglês: *One-Class Partial Least Squares*), PLS-DA e *i*SPA-PLS-DA. O *i*SPA-PLS-DA alcançou a melhor capacidade preditiva, selecionando apenas variáveis com informações químicas mais relevantes e interpretáveis.

### 3.4.6.1 Algoritmo das projeções sucessivas para seleção de intervalos em análise discriminante por mínimos quadrados parciais (*i*SPA-PLS-DA)

O modelo *i*SPA-PLS-DA foi proposto por Fernandes e colaboradores como uma nova abordagem de seleção de intervalos para modelagem de PLS-DA. De acordo com a literatura, o algoritmo *i*SPA-PLS-DA é composto por duas fases (FERNANDES *et al.*, 2016).

Na primeira fase, a matriz de treinamento é utilizada para determinar o número ótimo de fatores para o modelo PLS-DA, empregando o processo de validação cruzada. O número ótimo de fatores é obtido levando em consideração a minimização da taxa de erro. Em seguida, o conjunto total de variáveis  $k$  é dividido em  $w$  intervalos não sobrepostos. Em cada intervalo de variáveis, o vetor coluna  $\mathbf{X}_{\text{train}}$ , com a maior norma, é considerado como o elemento representativo do referido intervalo. Assim, origina uma matriz  $\mathbf{W}_{\text{train}}$  através da combinação dos vetores de coluna  $w$  e, então, utiliza-se como entrada para a sequência de projeções envolvidas no algoritmo SPA padrão, resultando em combinações de intervalos de  $m$  na matriz **SEL** (de dimensão  $(w-1) \times w$ ), definido pelo conjunto de índices  $\{\text{SEL}(1, k), \text{SEL}(2, k), \dots, \text{SEL}(m, k)\}$  (FERNANDES *et al.*, 2016).

Na segunda fase, cada combinação de intervalos na matriz **SEL** foi utilizada na construção do modelo PLS-DA, empregando novamente o processo de validação cruzada, e os modelos são calculados variando o número de fatores de 1 até o valor obtido no modelo PLS-DA anterior. Por fim, a melhor combinação de intervalos selecionada deve ter a maior taxa de classificação correta (FERNANDES *et al.*, 2016).

### 3.4.7 Validação de modelos de classificação

A performance dos modelos de classificação é determinada a partir de parâmetros de desempenho analítico, que possam avaliar a capacidade do modelo de identificar amostras pertencentes e não pertencentes à classe alvo/autêntica, sendo a sensibilidade, especificidade e acurácia os parâmetros mais comumente utilizados nestes casos. A sensibilidade é definida de quão bem o modelo é capaz de classificar as amostras da classe alvo corretamente, também pode ser definida como a taxa de verdadeiro positivos (VP) (RODIONOVA; POMERANTSEV, 2020). Para o

conjunto de treinamento e teste, a sensibilidade é calculada conforme a **Equação 6** (ZONTOV *et al.*, 2017).

$$\text{Sensibilidade} = 100\% \frac{(\text{Amostras} - \text{Extremos})}{\text{Amostras}} \quad (\text{eq. 6})$$

onde, Amostras é a quantidade total de amostras da classe autêntica pertencentes ao conjunto de treinamento ou teste; e, Extremos é a quantidade de amostras da classe autêntica que foram atribuídas não pertencente a ela.

A especificidade mede a habilidade do modelo ao identificar corretamente as amostras pertencentes à classe não autêntica. Lembrando que o cálculo de especificidade do modelo DD-SIMCA só pode ser calculado a partir das amostras pertencentes a classe não autêntica e, portanto, os resultados de treinamento não possuem valores de especificidade (ZONTOV *et al.*, 2017). Matematicamente, é definida pela **Equação 7**:

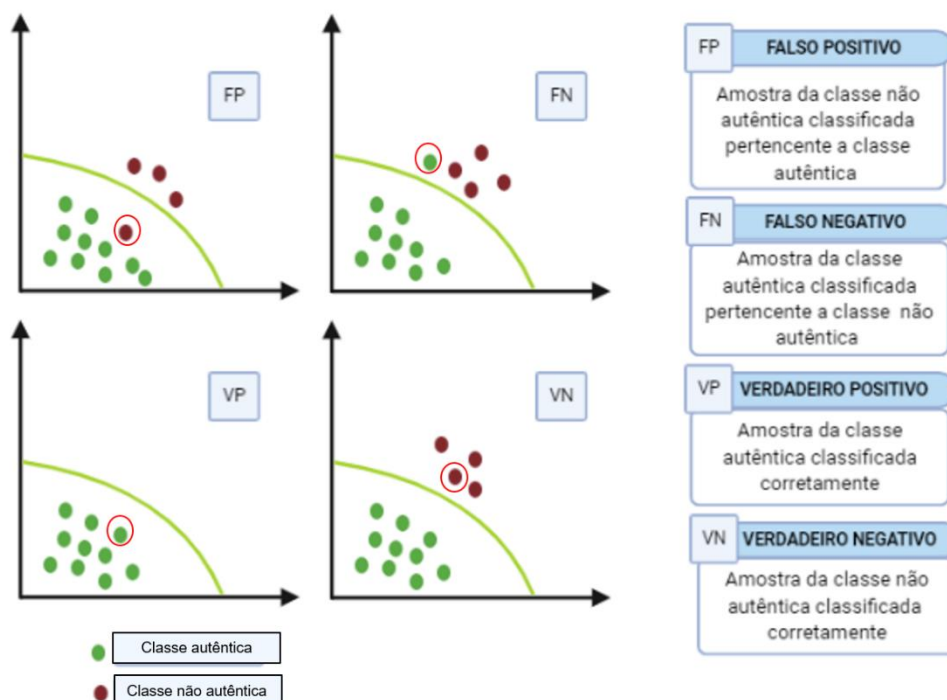
$$\text{Especificidade} = 100\% \frac{(\text{Objetos externos})}{\text{Amostras}} \quad (\text{eq. 7})$$

onde, Objetos externos representa a quantidade de acertos das amostras pertencentes à classe não autêntica e Amostras corresponde a quantidade total de amostras da classe não autêntica pertencentes ao conjunto teste.

A acurácia é um parâmetro estatístico empregado para avaliar o desempenho global dos modelos qualitativos. O cálculo se dá a partir do número de amostras, classificadas corretamente (independente da classe), dividido pelo número total de amostras (SANTANA *et al.*, 2020).

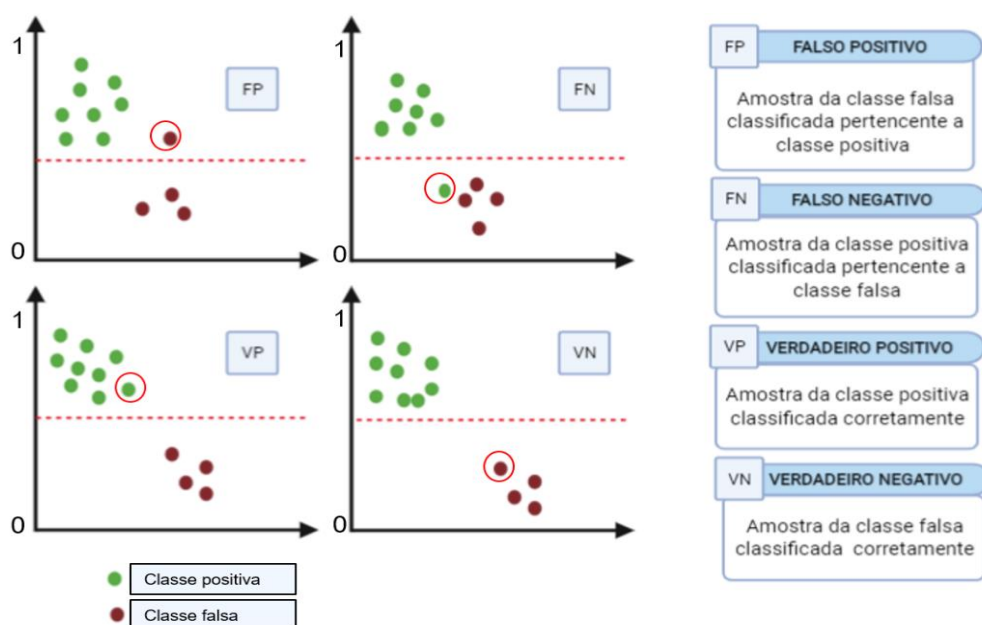
$$\text{Acurácia} = 100\% \frac{VP+VN}{VP+VN+FP+FN} \quad (\text{eq. 8})$$

onde, os termos VP (verdadeiro positivo), VN (verdadeiro negativo), FP (falso positivo) e FN (falso negativo) são explicados na **Figura 4 e 5**.



**Figura 4.** Representação gráfica dos parâmetros de avaliação do modelo de classificação (Criada com BioRender.com)

Já no método PLS-DA e *i*SPA-PLS-DA podem ser calculados os valores correspondentes aos parâmetros de desempenho tanto no conjunto de treinamento quanto no conjunto de teste. A sensibilidade total representa a proporção das amostras verdadeiras positivas. A especificidade representa a proporção das amostras verdadeiras negativas (RODIONOVA; POMERANTSEV, 2020).



**Figura 5.** Representação gráfica dos parâmetros de avaliação do modelo de classificação para o PLS-DA e *i*SPA-PLS-DA (Criada com BioRender.com)

*Capítulo 4*  
**METODOLOGIA**

---

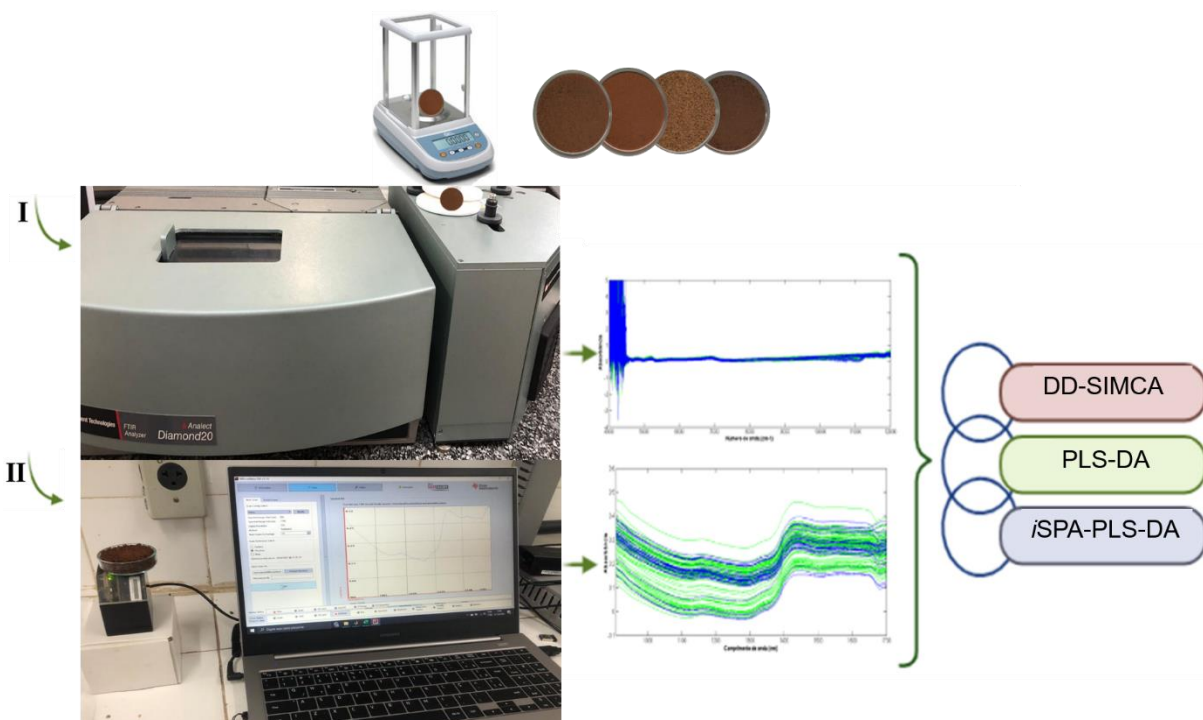
## 4 METODOLOGIA

### 4.1 Aquisição de amostras

Foram adquiridas 103 amostras de cafés solúveis em estabelecimentos comerciais nas cidades de João Pessoa – PB, Arapiraca – AL e Barreiras – BA, das quais eram produzidas por diferentes processos de secagem (liofilização, aglomeração e atomização) e pontos de torra, incluindo produtos regulares (com cafeína) e descafeinados, com 15 marcas e todas pertencentes a diferentes lotes de fabricação, a fim de assegurar variabilidade e representatividade aos modelos. O conjunto é composto por 30 amostras de cafés solúveis descafeinados, 42 amostras de cafés solúveis tradicionais e 31 amostras de cafés solúveis extraforte.

### 4.2 Aquisição dos espectros NIR

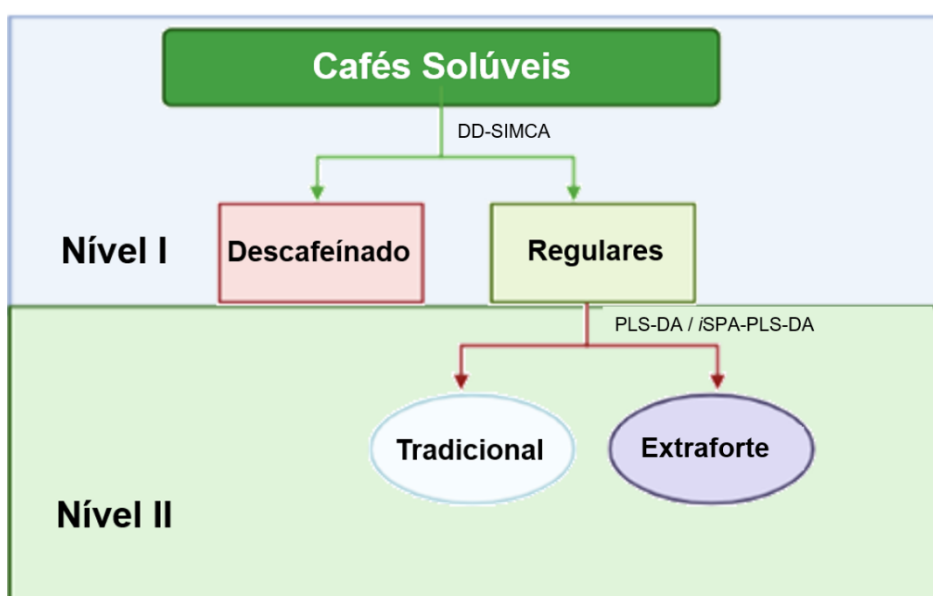
Um diagrama esquemático do procedimento de pesagem, das medidas dos espectros NIR e dos tratamentos quimiométricos utilizados é mostrado na **Figura 6**. As cento e três amostras de cafés solúveis foram pesadas, 8g, em placas de petri lisas de poliestireno (60 × 15 mm) para aquisição dos espectros NIR por meio de dois equipamentos diferentes. Todas as análises foram realizadas em triplicata e à temperatura ambiente ( $23 \pm 1^\circ\text{C}$ ). No experimento (I) foi usado um espectrofotômetro de bancada FT-NIR com faixa espectral de 4000 a 12000  $\text{cm}^{-1}$ , resolução de 4  $\text{cm}^{-1}$ , e 16 varreduras modelo Analyzer Diamond 20, Applied Instrument Technologies no modo reflectância difusa e detector de InGaAs. Para experimento (II), os espectros foram obtidos a partir de um NIR portátil modelo DLP NIRscan Nano Evaluation Module, marca Texas Instruments na faixa de 900 a 1700nm (11111 a 5882  $\text{cm}^{-1}$ ) com resolução digital de 228 pontos e 32 varreduras usando transformação de Hadamard. Para ambos os experimentos, o branco das amostras foi realizado empregando uma pastilha de politetrafluoretileno (PTFE).



**Figura 6.** Diagrama esquemático do processo de pesagem, obtenção dos espectros NIR e ferramentas quimiométricas

### 4.3 Procedimento Quimiométrico e Software

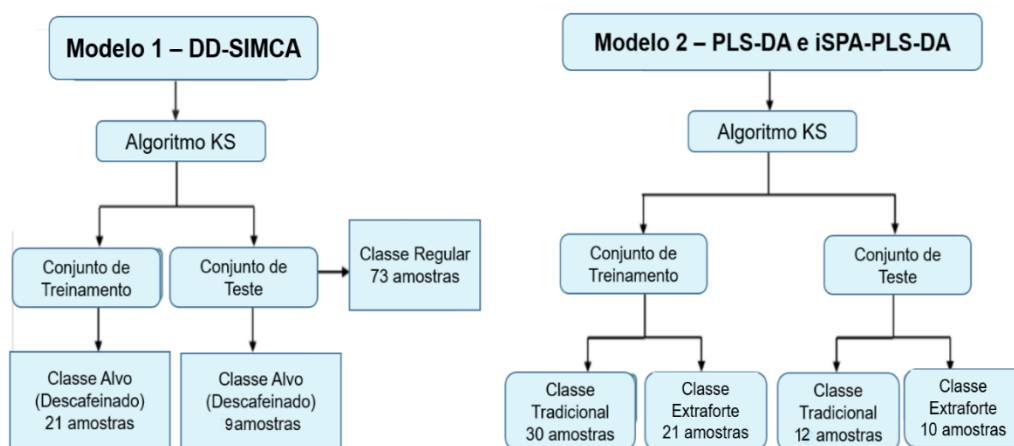
As informações contidas nos espectros foram utilizadas para a classificação dos cafés solúveis aliada às técnicas de reconhecimento de padrões, utilizando dois níveis de classificação diferente, como representado na **Figura 7**.



**Figura 7.** Representação da classificação de cafés solúveis em dois níveis (Criada com BioRender.com)

No primeiro nível, os cafés solúveis foram classificados de acordo com o tipo, produtos regulares (categoria composta por produtos cafeinados) (SOARES *et al.*, 2021; MARCUCCI; BENASSI, 2013) e descafeinados. Em seguida, no segundo nível, os cafés solúveis regulares foram classificados a respeito do grau de torra, tradicionais e extraforte, como representado na **Figura 7**.

A análise exploratória para ambos os níveis de classificação foi realizada empregando Análise de Componentes Principais (PCA). A seleção de amostras para compor os conjuntos de treinamento (70%) e teste (30%) foi realizada utilizando o algoritmo Kennard-Stone (KS), como mostrado no diagrama da **Figura 8**. Os cálculos do PLS-DA e DD-SIMCA foram realizados utilizando o pacote *Classification Toolbox* 5.1 (BALLABIO; CONSONNI, 2013) e *DD-SIMCA Tool* (ZONTOV *et al.*, 2017), enquanto *iSPA-PLS-DA* foi executado empregando as rotinas desenvolvidas por Fernandes *et al.* (2016). Todos os procedimentos quimiométricos foram implementados em Matlab® 2011b (Mathworks, USA).



**Figura 8.** Conjuntos de amostras de treinamento e teste em cada classe para os dois modelos de classificação

Para o primeiro nível de classificação ilustrado na **Figura 8**, o DD-SIMCA foi utilizado com o objetivo de classificar as amostras com baixo teor de cafeína. Nesta etapa, apenas as amostras pertencentes a classe autêntica (classe descafeinado) foram consideradas. Um total de 21 amostras foram selecionadas para compor o conjunto de treinamento. As 9 amostras restantes de cafés solúveis descafeinado foram inseridas juntas as amostras de cafés solúveis regulares (73 amostras) para formar o conjunto de teste.

Para construção do modelo DD-SIMCA foi adotado o nível de significância de *outliers* de 0,05 para definição da área de aceitação usando distribuição qui-quadrado no modo robusto (leva em consideração possíveis *outliers* no modelo). A escolha do número de componentes principais (PCs) foi obtida analisando os valores da sensibilidade para o conjunto de treinamento, utilizando validação cruzada completa, em função do número de PCs adicionado ao modelo. O número de PCs escolhido deve ter o menor número de amostras extremas (falso negativo) ou *outliers* (amostras discrepantes a respeito da classe alvo) para o conjunto de treinamento (MANUEL *et al.*, 2021; POMERANTSEV; RODIONOVA, 2020b).

Para o segundo nível de classificação, PLS-DA e *i*SPA-PLS-DA foram utilizados para identificar amostras que possuíam grau de torra tradicional frente amostras de torra extraforte. O conjunto de dados foi então dividido em conjuntos de treinamento com 51 amostras (30 tradicional e 21 extraforte) e, teste com 22 amostras (12 tradicional e 10 extraforte).

Para os modelos PLS-DA e *i*SPA-PLS-DA a escolha do número de variáveis latentes (LV's) foi baseada no menor valor da taxa de erro de classificação para o conjunto de treinamento empregando validação cruzada completa. A taxa de erro geral para as classes (C) é definida pela **Equação 9**, que corresponde ao somatório da taxa de falso negativo (FN*R*<sub>*i*</sub>, do inglês: *False Negative Rate*) e falso positivo (FP*R*<sub>*i*</sub>, do inglês: *False Positive Rate*) dividido pelo número de classes (FERNANDES *et al.*, 2016).

$$ER = \frac{1}{C} \sum_{i=1}^C ER_i = \frac{FN_{Ri} + FP_{Ri}}{2} \quad (\text{eq. 9})$$

*Capítulo 5*  
**RESULTADOS E  
DISCUSSÃO**

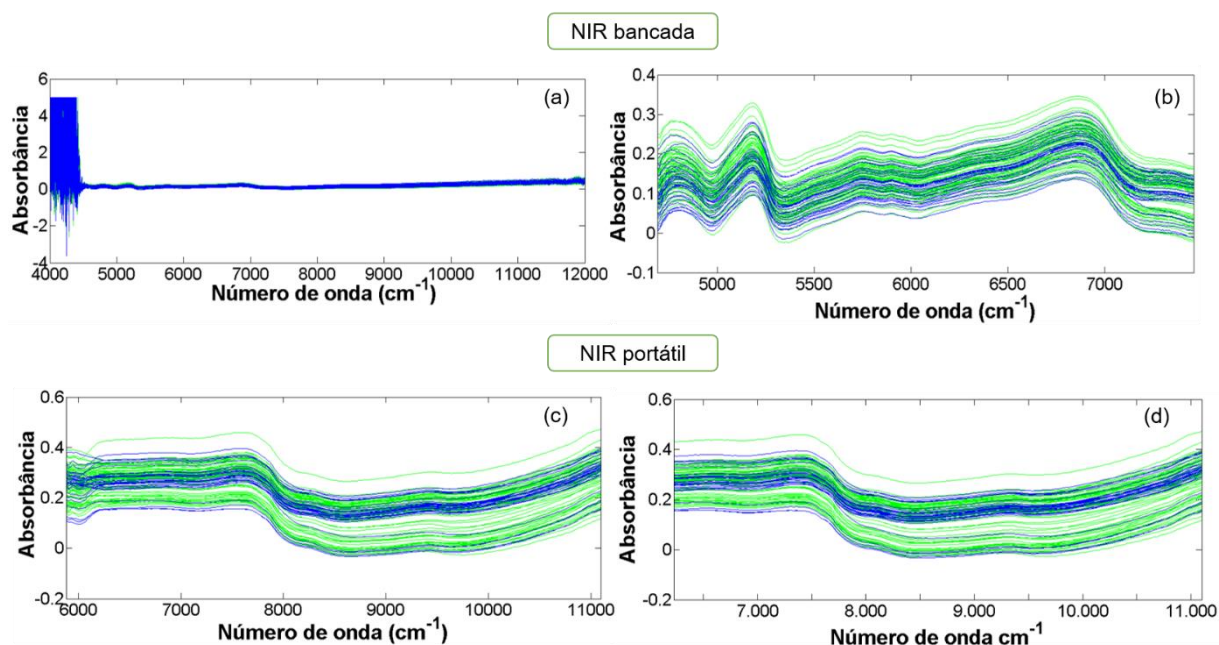
---

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Classificação de cafés solúveis descafeinado usando espectroscopia NIR e DD-SIMCA

#### 5.1.1 Espectros NIR bancada e portátil

Os espectros das 103 amostras de cafés solúveis, obtidos na região do NIR de bancada (4000 a 1200  $\text{cm}^{-1}$ ) e portátil (5882 a 11111  $\text{cm}^{-1}$ ), são apresentados nas **Figuras 9a e 9c**. Como pode ser observado, na faixa espectral do NIR de bancada 4000 a 4690  $\text{cm}^{-1}$  apresenta sinal de absorbância alto e ruidoso devido a saturação do detector, e na faixa de 7460 a 12000  $\text{cm}^{-1}$  representa uma região não informativa, enquanto no NIR portátil há presença do sinal ruidoso no início do espectro (5882 - 6211  $\text{cm}^{-1}$ ). Com isso, fez-se necessário diminuir o número de variáveis por seleção a priori para melhorar o desempenho na construção do modelo, representado na **Figura 9b e 9d**.



**Figura 9.** Espectros NIR (a,c) originais (b,d) cortados para as 103 amostras de café solúveis regulares (verde) e descafeínados (azul)

Para análise de dados foi selecionado para NIR de bancada a faixa espectral 4690 a 7460  $\text{cm}^{-1}$  (**Figura 9b**) e portátil 6211 a 11111  $\text{cm}^{-1}$  (**Figura 9d**). Ao investigar os espectros da **Figura 9b**, foi possível perceber a existência de sinais intensos de

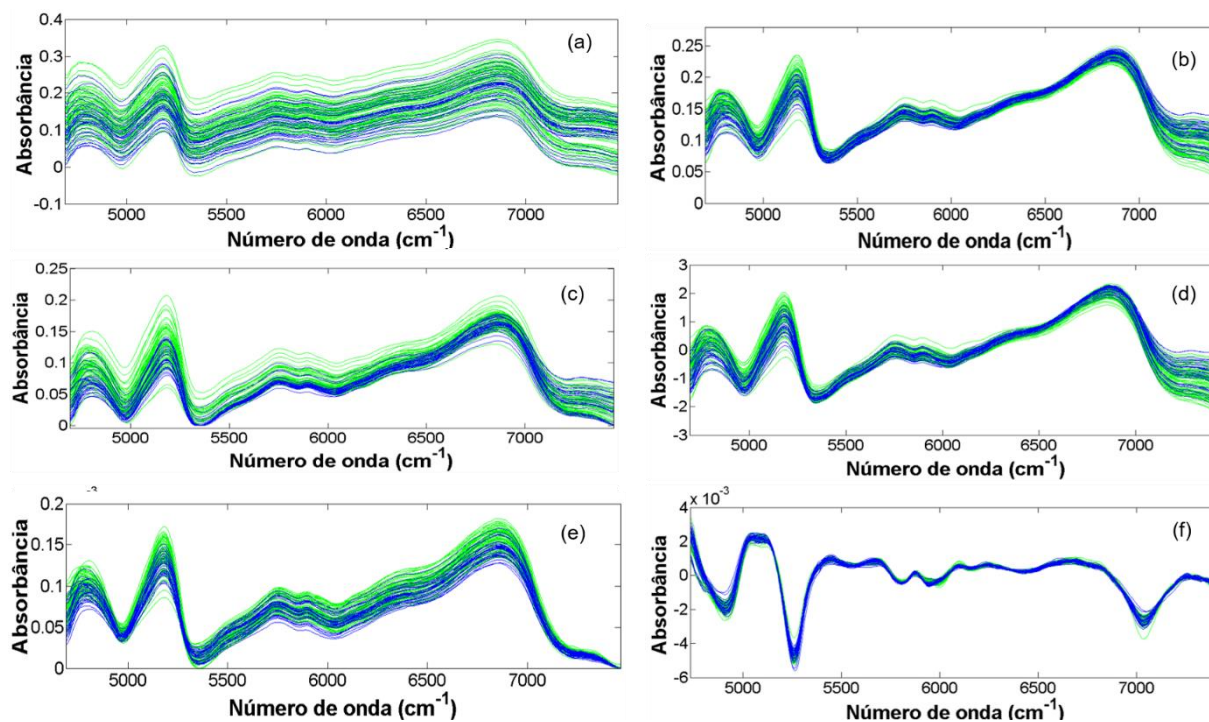
bandas de combinação e sobretoms em regiões possivelmente relacionadas a água entre 5000 a 5200  $\text{cm}^{-1}$  que corresponde a banda de combinação de estiramento e deformação do grupo hidroxila (O-H) e de 6750-7200  $\text{cm}^{-1}$  refere-se ao primeiro sobretom da vibração de estiramento de O-H. Em 5200-5300  $\text{cm}^{-1}$  pode estar associada a vibração do segundo sobretom da carbonila (C=O) nos grupos funcionais de CONH que provavelmente está relacionada a cafeína (CRAIG *et al.*, 2015; ARAUJO *et al.*, 2020).

Além disso, foram observadas bandas vibracionais de estiramento e deformação do grupo -CH em torno de 7200  $\text{cm}^{-1}$ , primeiro sobretom de -CH<sub>2</sub> (~5750  $\text{cm}^{-1}$ ) e -CH<sub>3</sub> (~5800  $\text{cm}^{-1}$ ) e segundo sobretom de -CH (9091  $\text{cm}^{-1}$ ), geralmente causadas pelos principais componentes encontrados no café, como: cafeína, ácidos clorogênicos, carboidratos e alguns compostos aromáticos. Em torno de 5660-5830  $\text{cm}^{-1}$  é reportado na literatura a presença de bandas referentes ao conteúdo de lípidios devido a presença de ácidos graxos de cadeia longa em café arábica e robusta que dão origem ao primeiro sobretom de -CH<sub>2</sub> e em 4600-4900  $\text{cm}^{-1}$  a presença das bandas de combinação C-N, C-H (CRAIG *et al.*, 2015).

### 5.1.1.2 Pré-Processamentos

Nos espectros apresentados na **Figura 9b e 9d** é possível notar a presença de ruídos e variações sistemáticas na linha de base, devido ao fenômeno de espalhamento de luz causado por diferentes granulometrias decorrentes dos seus processos de secagem (pó, granulado ou liofilizado) do café solúvel. Dessa forma, foi aplicada uma série de pré-processamentos para atenuar estes problemas. Inicialmente, foi empregado nos espectros do NIR de bancada o algoritmo de suavização baseado em uma média móvel (MM) com janela de 9 pontos combinado com diferentes pré-processamento, como: Correção de espalhamento multiplicativo (MSC), Variação normal padrão (SNV), Correção de linha de base linear (LBC) e offset (BO). Além desses, o filtro derivativo Savitzky-Golay com primeira derivada, polinômio de segunda ordem e janela de 23 pontos foi utilizado nos dados antes da suavização por média móvel de 9 pontos.

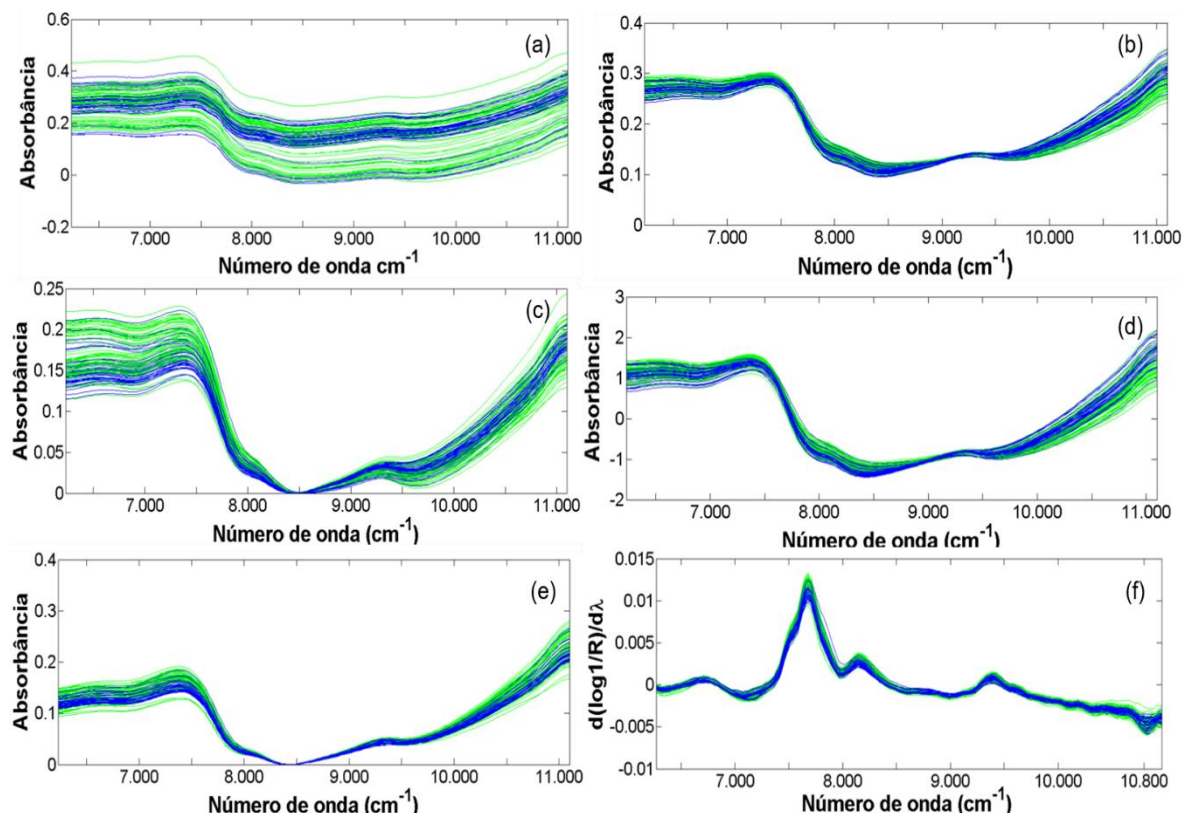
Na **Figura 10** é apresentada os espectros NIR utilizando os diferentes tipos de pré-processamentos e suas combinações.



**Figura 10.** Espectros (a) originais e pré-processados das classes regular (verde) e descafeinado (azul) com: (a) Suavização Média Móvel (MM) com janela de 9 pontos e MSC, (c) MM e BO, (d) MM e SNV, (e) MM+LBC+BO, (f) derivação Savitzky-Golay (dSG) com primeira derivada, janela de 23 pontos e polinômio de segunda ordem (●: Regular, ●: Descafeinado)

Para os espectros obtidos no NIR portátil foi aplicado a média móvel com janela de 7 pontos. As variações sistemáticas na linha de base foram reduzidas empregando outras técnicas de pré-processamento, com MSC, SNV, BO e LBC. Foi utilizado também a derivada Savitzky-Golay (primeira derivada, janela de 7 pontos e polinômio de segunda ordem) conforme a **Figura 11**.

Após a realização dos diversos tipos de pré-processamentos nos espectros (**Figura 10 e 11**), foi observada a redução do espalhamento dos espectros, ruído e a correção da linha de base. A partir da etapa de pré-processamento, o algoritmo PCA foi usado para a análise exploratória dos dados.

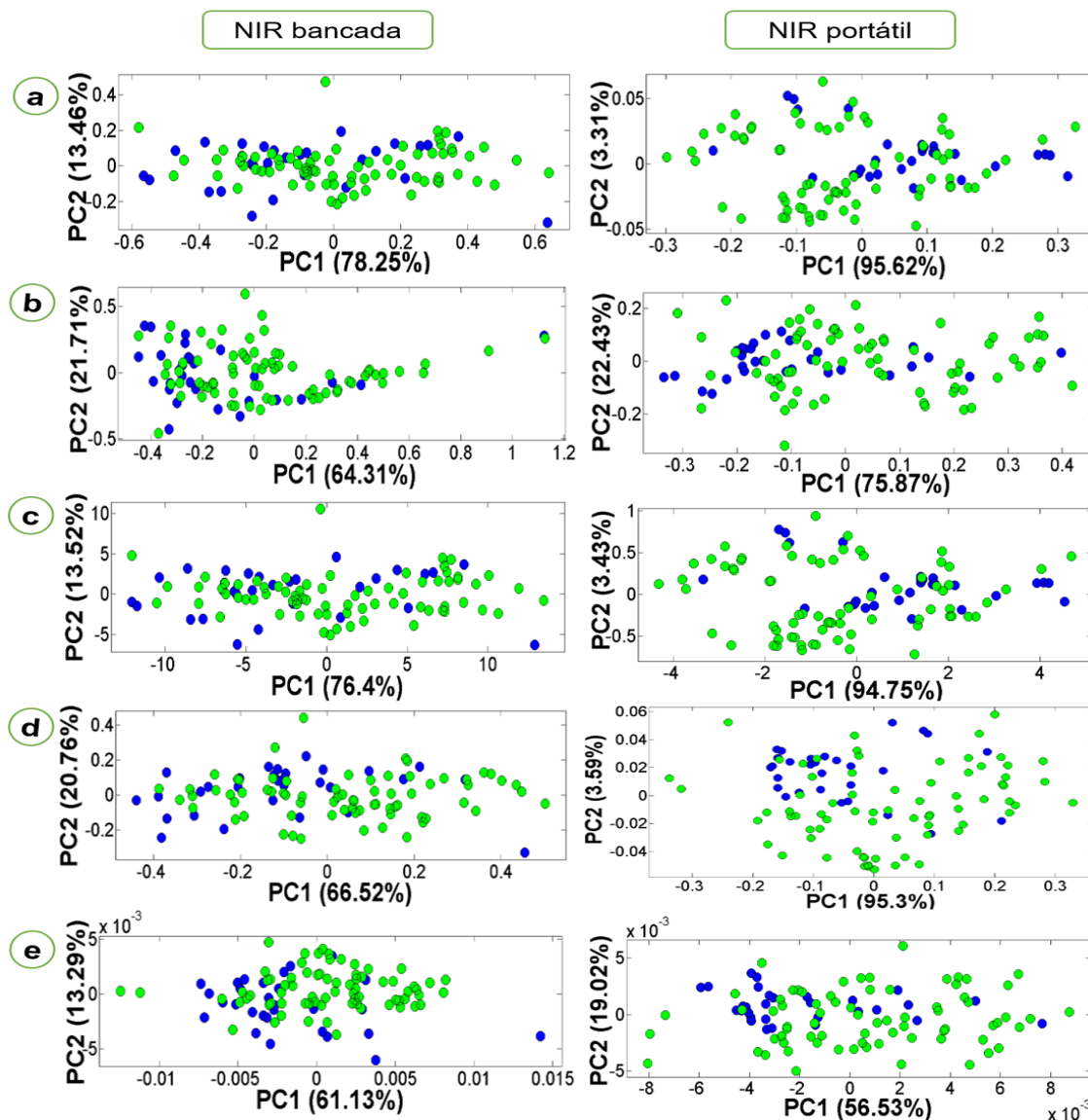


**Figura 11.** Espectros (a) originais e pré-processados das classes regular (verde) e descafeinado (azul) com: (b) Suavização Média Móvel (MM) com janela de 7 pontos e MSC, (c) MM e BO, (d) MM e SNV, (e) MM+LBC+BO, (f) derivação Savitzky-Golay ( $dX_7$ ) com primeira derivada, janela de 7 pontos e polinômio de segunda ordem (●: Regular, ●: Descafeinado)

### 5.1.1.3 Análise por componentes principais

Na **Figura 12** é ilustrada os gráficos de escores do NIR de bancada e portátil das duas primeiras PCs obtidos por meio dos diferentes pré-tratamentos das amostras de cafés solúveis, com o propósito de avaliar a capacidade discriminante das mesmas em relação ao tipo das amostras de cafés descritas no rótulo.

Contudo, foi possível observar uma ligeira tendência de separação entre os valores positivos de escores em PC1 das amostras regulares na **Figura 12e** referente ao equipamento de bancada. Porém, há uma forte sobreposição dos tipos de cafés regular e descafeinado e uma dispersão substancial. Esse comportamento pode ser causado devido ao processo de descafeinação implementado por cada indústria, além dos fatores ambientais que influenciam diretamente na característica/qualidade do produto final.



**Figura 12.** Gráfico dos escores de PC1 versus PC2 para as 103 amostras de cafés solúveis utilizando os espectros pré-processados: **(a)** Suavização Média Móvel (MM) e MSC, **(b)** MM e BO, **(c)** MM e SNV, **(d)** MM+LBC+BO, **(e)** derivação Savitzky-Golay (dSG) com primeira derivada e polinômio de segunda ordem (●: Regular, ●: Descafeinado)

Assim, foi empregado o classificador “*one class*” para fins de classificação, incluindo novas amostras no conjunto de teste como pertencentes ou não a classe descafeinado (que não foram utilizadas para construção do modelo).

#### 5.1.1.4 Classificação

Para o primeiro nível de classificação, o modelo DD-SIMCA foi construído utilizando apenas amostras representativas da classe alvo (produtos comerciais descafeinado), independentemente de se ter o conhecimento do teor de cafeína das amostras da classe regular, que podem variar de acordo com a proporção dos blends

de café robusta e arábica. Na literatura, o teor de cafeína para cafés solúveis comerciais apresentou maior variabilidade, de 2,84 a 5,82 g de cafeína 100 g<sup>-1</sup> (VIGNOLI; BASSOLI; BENASSI, 2011). Por esse motivo, o método DD-SIMCA foi escolhido para verificar a conformidade dos produtos descafeinados para segurança alimentar do consumidor, pois as amostras pertencentes a esta classe formam um conjunto mais uniforme.

Na **Tabela 2** apresenta os resultados obtidos para os modelos de classificação DD-SIMCA em termos de sensibilidade, especificidade e acurácia para os diversos tipos de pré-processamentos nos espectros NIR de bancada e portátil.

**Tabela 2** – Resultados dos parâmetros de desempenho sensibilidade, especificidade e acurácia para DD-SIMCA nos conjuntos de treinamento e teste

Método de Classificação		DD-SIMCA							
		NIR bancada				NIR portátil			
Pré-Processamento	Conjunto de amostra	PCs	Sens.	Espec.	Acurácia	PCs	Sens.	Espec.	Acurácia
MM+BO	Treinamento	8	100	-	100	8	100	-	98,06
	Teste		100	100			77,78	100	
MM+MSC	Treinamento	8	90,48	-	98,06	7	95,24	-	99,03
	Teste		100	100			100	100	
MM+SNV	Treinamento	7	90,48	-	95,15	8	100	-	100
	Teste		66,67	100			100	100	
MM+BO+LBC	Treinamento	7	90,48	-	94,18	6	90,48	-	87,38
	Teste		55,56	100			100	84,93	
dSG	Treinamento	7	100	-	99,03	7	95,24	-	92,23
	Teste		88,89	100			100	90,41	

PCs: componentes principais; Sens: sensibilidade; Espec: especificidade; MM: média Móvel; SNV: variação normal padrão; MSC: correção de espalhamento multiplicativo; BO: correção de linha de base *offset*; LBC: correção de linha de base linear; dSG: derivada Savitsky-Golay.

A princípio, a escolha do número de componentes principais foi adotada através do modelo que fornecesse o número total das amostras de treinamento dentro da área de aceitação, ou seja, máxima taxa de verdadeiros positivos (NEVES e POPPI, 2020). Portanto, para alcançar valores máximos de sensibilidade, o modelo DD-SIMCA para os espectros pré-processados para o NIR de bancada e portátil selecionaram acima de 7 componentes principais. A quantidade de PCs selecionadas podem estar associadas à complexidade da matriz, além da variabilidade dos diferentes lotes e marcas.

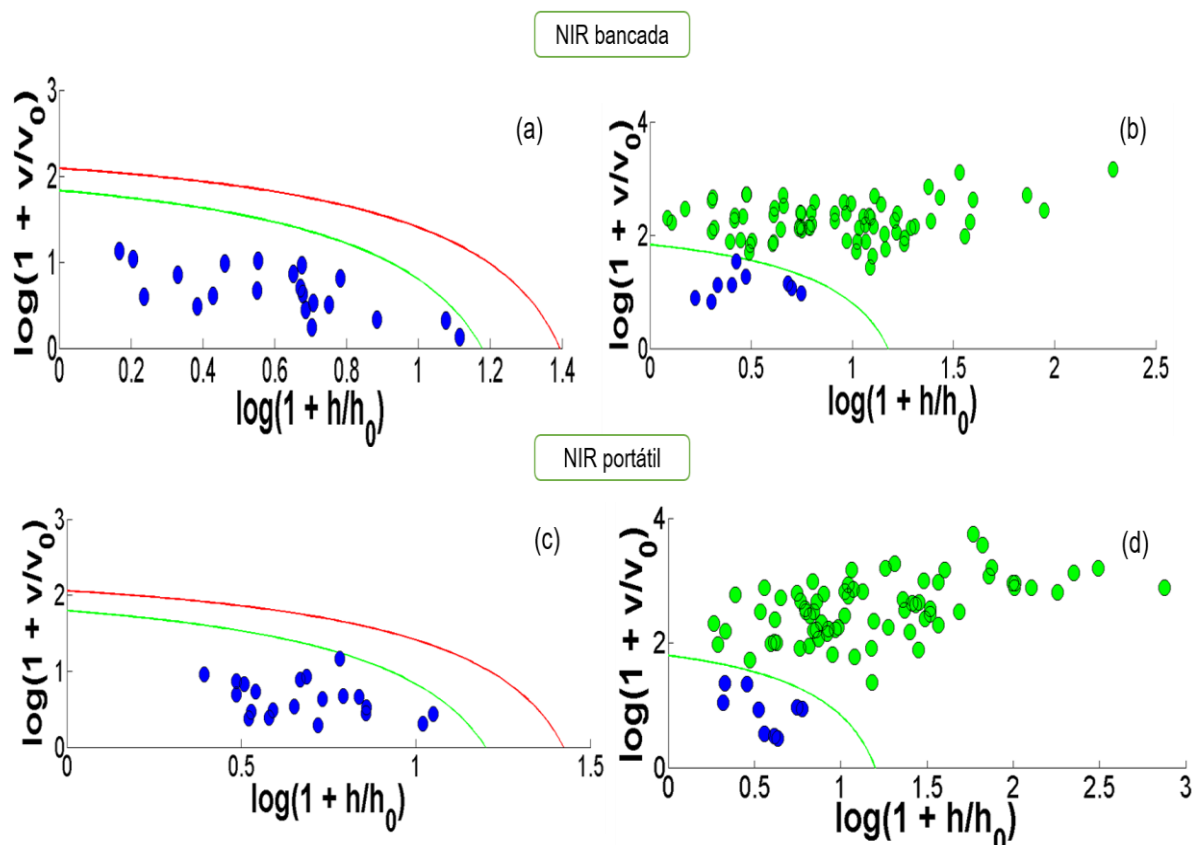
Ao analisar a **Tabela 2**, observa-se que para o equipamento de bancada, a combinação da Média Móvel com a correção de linha de base *offset* (MM+BO) promoveu o melhor modelo devido aos valores máximos de sensibilidade tanto para o conjunto de treinamento e teste, como também para especificidade e acurácia no conjunto de teste. De igual modo, foi encontrado na literatura que o uso da combinação desses pré-processamentos (MM+BO) obteve melhores resultados nas análises direta de leite por Pereira *et al.* (2021) e de cafés torrados e moídos por Araújo *et al.* (2021).

O uso desse pré-processamento pode ser explicado devido à alta capacidade preditiva quando comparado com a derivada Savitzky-Golay. Além disso, a derivada requer cálculos mais complexos e seleção cuidadosa na escolha do tamanho da janela, o que pode ocasionar perdas de informações importantes (TIBOLA *et al.*, 2018).

Ainda de acordo com os resultados apresentados na **Tabela 2**, para o NIR portátil verifica-se que o modelo desenvolvido DD-SIMCA com o pré-processamento média móvel e variação normal padrão (MM+SNV) apresentou resultado satisfatório, alcançando 100% de sensibilidade nos conjuntos de treinamento e teste, 100% de especificidade para previsão, com oito componentes principais. Em métodos de *one class*, o cálculo de especificidade é determinado apenas no conjunto de teste, uma vez que o conjunto de treinamento contém apenas amostras da classe alvo (descafeinado).

Neste caso, a discriminação da classe descafeinada em relação a classe regular pode ser atribuída as bandas de absorção dos grupos funcionais da cafeína, ácidos clorogênicos e carboidratos, porque, segundo Ribeiro e colaboradores (2010) identificaram nas amostras descafeinada menores concentrações desses compostos em comparação com os cafés torrados tradicionais e escuros. Bertone e colegas (2016) observaram também que no processo de descafeinação, à uma ligeira modificação na composição química dos grãos de café, sendo diferenciadas das amostras não descafeinada. Por sua vez, Santos e colegas (2012) foram capazes de discriminar amostras de cafés torrados e moídos de acordo com o tipo (tradicional, gourmet e descafeinada) usando espectroscopia NIR associado ao SIMCA e PLS-DA, obtendo 90,8 % e 96,3% de acerto, respectivamente.

A **Figura 13** mostra o gráfico de área de aceitação para o DD-SIMCA nos conjuntos de treinamento e teste para os pré-processamentos MM+BO e MM+SNV obtidos pelos equipamentos de bancada e portátil, respectivamente.



**Figura 13.** Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) para o algoritmo DD-SIMCA desenvolvidos no NIR de bancada (a, b) e portátil (c, d) (●) descafeinado (●) regular. As linhas verdes representam a borda entre as amostras regulares e extremas; a linha vermelha é a fronteira da área de aceitação

Para o modelo DD-SIMCA, o gráfico de distância dos escores (SD) *versus* distâncias ortogonais auxilia nos resultados de classificação, dado em escala logarítmica, no intuito de melhorar a visualização e identificar se uma amostra pertence ou não a classe autêntica/alvo. A combinação dessas duas distâncias é descrita em termos da variável  $c$ , distância global, onde assume valores que seguem uma distribuição qui-quadrado e fornece a construção da área de decisão de dois níveis, com os limiares extremos e *outliers* (ZONTOV *et al.*, 2017).

Caso amostras apresentem valores menores que a variável  $c$ , abaixo do primeiro limiar (linha verde nas **Figuras 13**), passa a ser considerada como pertencente à classe alvo, enquanto amostras com valores de  $c$  entre o primeiro e

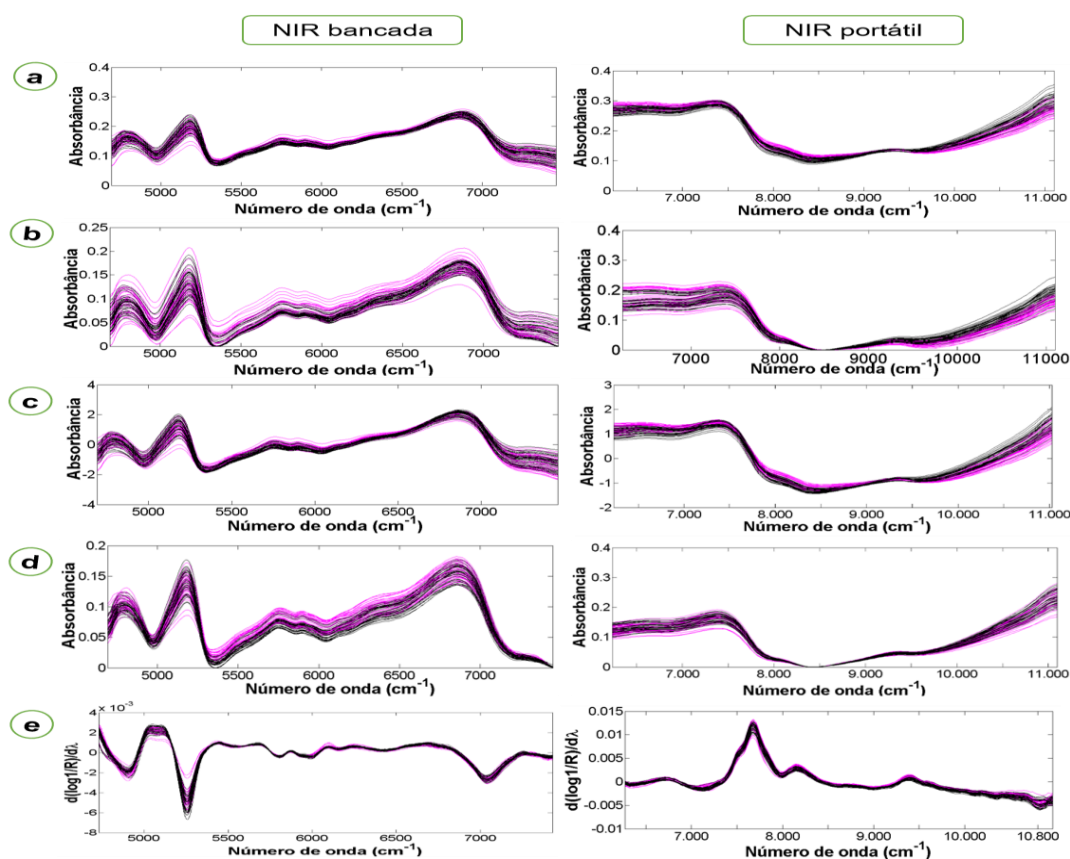
segundo limiar exibem um comportamento extremo. Amostras com valores de  $c$  maiores que o segundo limiar (linha vermelha na **Figura 13a e 13c**) são vistas como *outliers* (RODIONOVA *et al.*, 2014; ZONTOV *et al.*, 2017). Então, as amostras de café solúvel descafeinado no conjunto de treinamento e de teste (**Figura 13**) foram todas classificadas como pertencentes a classe e se posicionaram dentro da área de aceitação do modelo para um valor de significância de 0,05. Por outro lado, no conjunto de teste (**Figura 13b e 13d**), as amostras regulares (círculo verde) não pertencentes à classe autêntica e que contém teores distintos de cafeína (classe regular) se posicionaram fora da área de aceitação, resultando em uma taxa de especificidade de 100%, sem presença do erro falso positivo.

Percebe-se que os modelos *one class* construídos para ambos os equipamentos foram capazes de identificar corretamente todas as amostras, indicando que pode ser utilizado como uma ferramenta auxiliar ou alternativa aos métodos convencionais para verificação dos cafés solúveis comerciais. Considerando o baixo custo do instrumento portátil, pode ser indicado para realização das análises durante o processamento industrial, uma vez que o equipamento é portátil e de fácil manuseio.

## 5.2 Classificação de cafés solúveis tradicional e extraforte usando espectroscopia NIR associada a PLS-DA e *i*SPA-PLS-DA

### 5.2.1 Espectroscopia NIR bancada e portátil

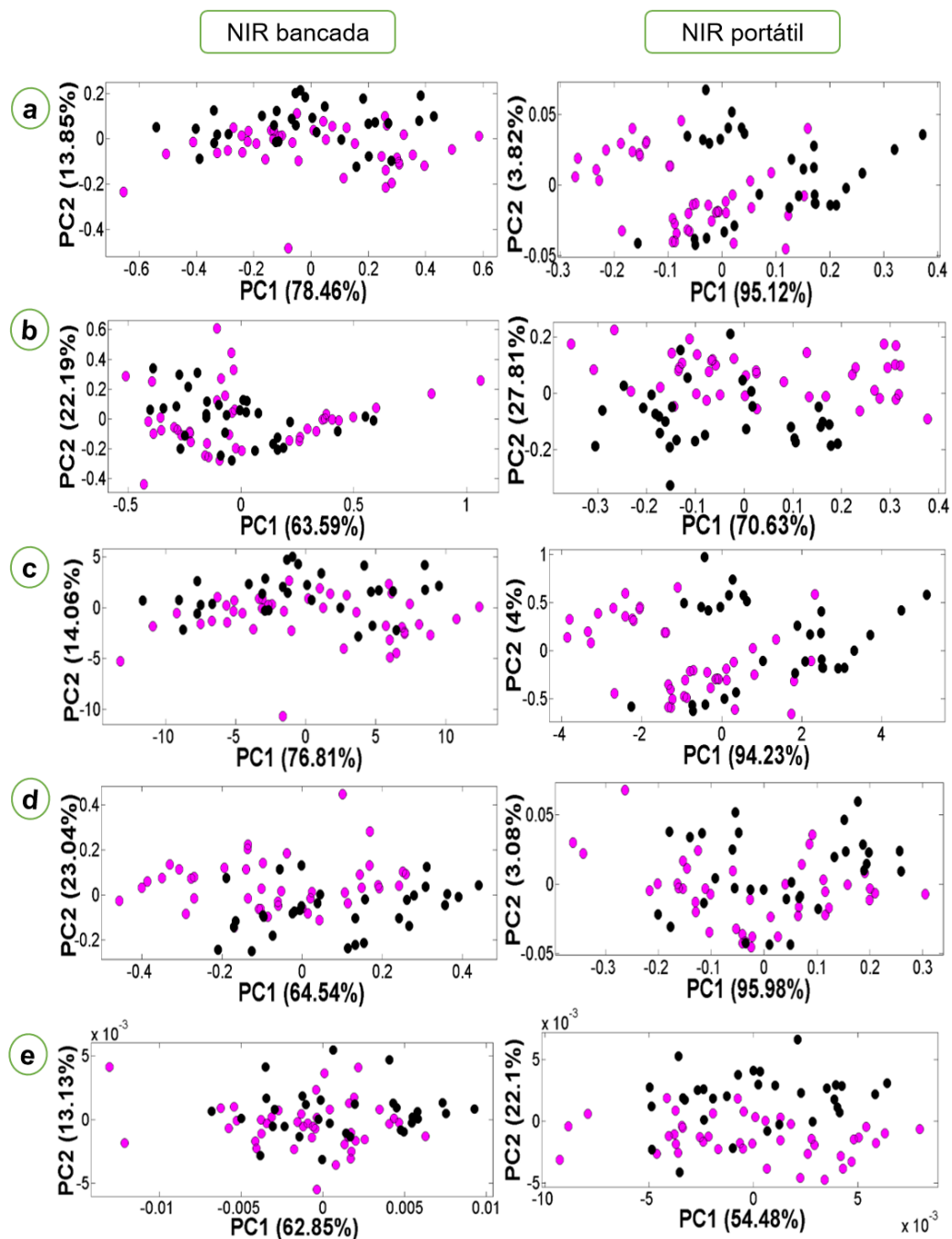
As amostras antes atribuídas a classe de café regulares são usadas no segundo nível de classificação empregando os métodos PLS-DA (*full spectrum*) e *i*SPA-PLS-DA (seleção de intervalos de variáveis) para discriminar amostras de cafés solúveis tradicionais e extraforte de acordo com seu grau de torrefação. Nesta seção, serão discutidos os resultados obtidos para os espectros no NIR de bancada e portátil utilizando a mesma faixa de trabalho e os mesmos pré-processamentos nos espectros NIR dos cafés solúveis descafeinados do primeiro nível de classificação. Na **Figura 14** são apresentados os espectros pré-processados das 73 amostras de cafés solúveis tradicionais e extraforte.



**Figura 14.** Espectros pré-processados das classes tradicional e extraforte com: (a) Suavização Média Móvel (MM) e MSC, (b) MM e BO, (c) MM e SNV, (d) MM+LBC+BO, (e) derivação Savitzky-Golay (dSG) com primeira derivada e polinômio de segunda ordem (●: Tradicional, ●: Extraforte)

### 5.2.1.1 Análise por componentes principais

A análise de componente principal foi empregada para detectar formação de agrupamentos e fornecer uma visão natural da diversidade das amostras de cafés solúveis tradicional e extraforte. A partir disso, foi gerado os gráficos de escores para os diferentes pré-processamentos como mostrado na **Figura 15**.



**Figura 15.** Gráfico dos escores de PC1 versus PC2 para as 73 amostras de cafés solúveis utilizando os espectros pré-processados: (a) Suavização Média Móvel (MM) e MSC, (b) MM e BO, (c) MM e SNV, (d) MM+LBC+BO, (e) derivação Savitzky-Golay (dSG) com primeira derivada e polinômio de segunda ordem (●: Tradicional, ●: Extraforte)

Após a avaliação dos gráficos de escores apresentados na **Figura 15**, no espectro de bancada, pode-se concluir que não houve formação de agrupamentos em nenhum pré-processamento entre as amostras de cafés solúveis tradicionais e extraforte. Já para o equipamento portátil, observa-se que nos gráficos de escores para os pré-processamento MM+BO (**Figura 15b**) e dSG\_7 (**Figura 15e**) a uma pequena tendência de separação entre as amostras das duas classes. Contudo, sobreposições entre as duas classes ainda são observadas em todos os pré-processamento dos diferentes equipamentos.

### 5.2.1.2 Classificação

De acordo com a **Figura 7**, o segundo nível de classificação se dá pela discriminação de café solúvel tradicional e extraforte, onde a indústria cafeeira, normalmente, mantém ou difere a proporção dos blends de café (Robusta ou Arábica) com intensidades de torra distintas para se obter cafés com qualidades diferentes para comercialização (BAQUETA *et al.*, 2019). Dessa forma, a torra escura do café extraforte, serve também para mascarar grãos defeituosos e possíveis adulterações com cascas e paus resultando em uma bebida de qualidade inferior (OLIVEIRA, 2018; COUTO *et al.*, 2020). Portanto, a metodologia proposta com a utilização do PLS-DA e  $\lambda$ SPA-PLS-DA pode ser útil na discriminação das amostras de café extraforte e tradicional, evitando rotulagens fraudulentas.

De início, o modelo PLS-DA foi desenvolvido a partir dos espectros pré-processados para os equipamentos de bancada e portátil das duas classes de cafés solúveis e o número ótimo de variáveis latentes foi determinado usando validação cruzada completa.

Para os espectros obtidos no equipamento de bancada, as informações do modelo PLS-DA incluindo acurácia, sensibilidade e especificidade são mostradas no **Anexo 1**. Ao analisar a Tabela do **Anexo 1**, pode-se observar que o melhor desempenho foi utilizando a combinação da Média Móvel com 9 pontos e correção de espalhamento multiplicativo (MSC), com sete variáveis latentes. Para este modelo (**Tabela 3**), uma amostra de café extraforte foi classificada incorretamente na classe tradicional no conjunto de treinamento, e todas as amostras de café extraforte no conjunto de teste foram classificadas corretamente, e uma amostra de café tradicional foi classificada incorretamente tanto no conjunto de treinamento como no de teste.

**Tabela 3** – Matriz de confusão obtido pelo algoritmo PLS-DA para os espectros NIR de bancada das 73 amostras de cafés solúveis comerciais

Método de Classificação	Pré- Processamento Conjunto de amostras	MM+MSC (7*)			
		Treinamento		Teste	
		TD	EF	TD	EF
	TD	29	1	11	1
	EF	1	20	-	10
Full PLS-DA	Sensibilidade (%)	96,7	95,2	91,7	100
	Especificidade (%)	95,2	96,7	100	91,7
	Acurácia (%)	96,1		95,5	
	Acurácia (%)	95,8			

(\*) variáveis latentes, TD: café tradicional, EF: café extraforte.

Para os espectros obtidos no equipamento portátil, o melhor resultado foi obtido através da combinação MM+BO com treze variáveis latentes, conforme mostrado no **Anexo 2**. Na **Tabela 4** é apresentado os resultados da classificação e a matriz de confusão para este modelo. Como pode ser visto, no conjunto de treinamento todas as amostras da classe tradicional foi classificada corretamente e 1 amostra da classe extraforte como tradicional, enquanto no conjunto de teste 3 amostras tradicional foram classificadas como extraforte e todas as amostras da classe extraforte foram classificadas corretamente.

**Tabela 4** – Matriz de confusão obtido pelo algoritmo PLS-DA para os espectros NIR portátil das 73 amostras de cafés solúveis comerciais

Método de Classificação	Pré- Processamento Conjunto de amostras	MM+BO (13*)			
		Treinamento		Teste	
		TD	EF	TD	EF
	TD	30	0	9	3
	EF	1	20	0	10
Full PLS-DA	Sensibilidade (%)	100	95,2	75	100
	Especificidade (%)	95,2	100	100	75
	Acurácia (%)	98,04		86,36	
	Acurácia (%)	92,2			

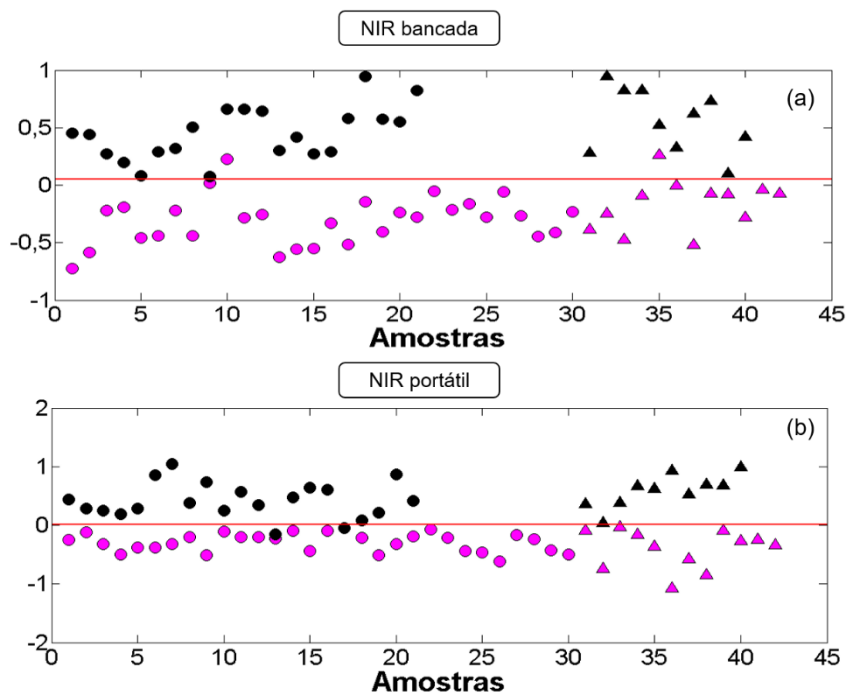
(\*) variáveis latentes, TD: café solúvel tradicional, EF: café solúvel extraforte.

Embora os resultados tenham bons desempenhos de classificação, acima de 90%, para os diversos tipos de pré-processamento, foi empregado o método *i*SPA-PLS-DA para seleção de intervalos das variáveis espectrais a fim de identificar regiões com informações químicas relevantes para discriminar a classe extraforte da tradicional e remover variáveis não informativas ou correlacionadas. Portanto, o algoritmo *i*SPA-PLS-DA foi testado com diferentes números de intervalos ( $w = 5, 10, 15, 20, 25$  e  $30$ ) e pré-processamentos como mostrado no **Anexo 1 e 2**.

A **Tabela 5** apresenta resumidamente os resultados dos modelos *i*SPA-PLS-DA. O equipamento de bancada mostrou uma ligeira melhoria no desempenho de classificação quando comparado com o *Full* PLS-DA, alcançando 96,7% de acurácia, porém identificou algumas variáveis espectrais relevantes para discriminação das classes estudadas. Todas as amostras de cafés solúveis extraforte do conjunto de treinamento e teste foram classificadas corretamente. Enquanto para as amostras pertencentes a classe tradicional, apenas duas foram classificadas erroneamente como classe extraforte, conforme é ilustrado na **Figura 16a**.

Por outro lado, para o instrumento portátil, o modelo *i*SPA-PLS-DA construído para os dados espectrais pré-processado com média móvel de 7 pontos e correção de linha de base *offset* melhorou significativamente a capacidade preditiva quando comparado com os resultados obtidos por PLS-DA/MM+BO, atingindo 98% de acurácia, e fornecendo um modelo mais parcimonioso com a redução do número de variáveis latentes e selecionando apenas as características espectrais relevantes, em vez do espectro completo.

Através do gráfico dos valores preditos (**Figura 16b**), observa-se que o modelo *i*SPA-PLS-DA/MM+BO para o equipamento portátil classificou corretamente todas as amostras de cafés solúveis tradicional em ambos os conjuntos de treinamento e teste. No entanto, duas amostras de café solúvel extraforte foram classificadas como tradicional no conjunto de treinamento. Todavia, para o conjunto de teste todas as amostras da classe extraforte foram classificadas corretamente, indicando que o modelo proposto associado ao equipamento NIR pode ser utilizado como ferramenta analítica para avaliação rápida dos cafés solúveis comerciais.



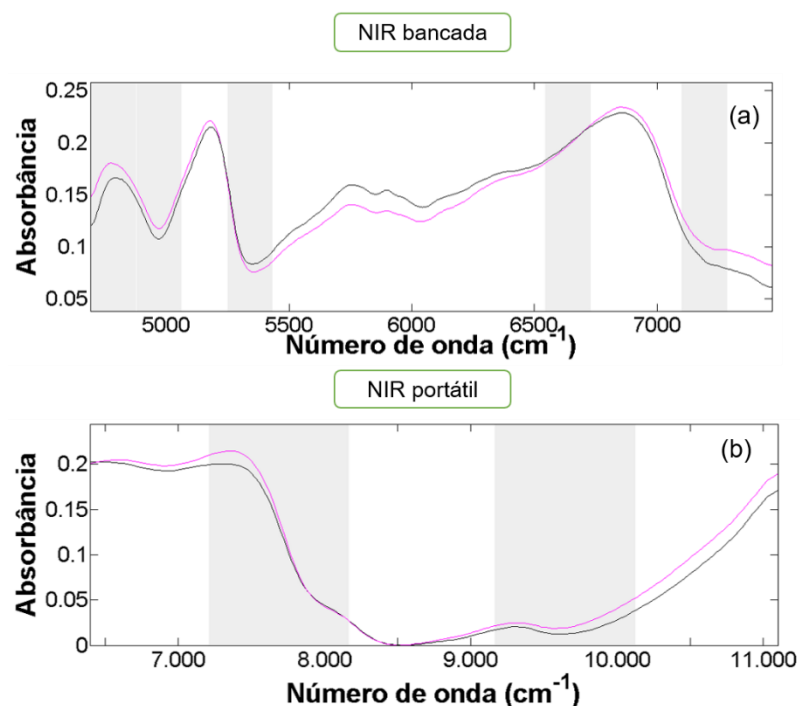
**Figura 16.** Resultados utilizando o equipamento NIR de bancada (a) e portátil (b) para previsão do modelo de classificação *iSPA-PLS-DA* das amostras do conjunto de treinamento (●: Tradicional, ●: Extraforte) e teste (▲: Tradicional, ▲: Extraforte). A linha vermelha representa o limiar (*threshold*) entre as classes

A **Figura 17** mostra as regiões espectrais responsáveis pela discriminação dos tipos de café solúveis, para ambos os equipamentos. Para o instrumento de bancada (**Figura 17a**), o *iSPA-PLS-DA* selecionou 5 intervalos de variáveis, quanto ao instrumento portátil (**Figura 17b**) foi selecionado apenas 2 regiões espectrais. Possivelmente, a distinção entre as duas classes deve-se ao nível de torra que provoca a degradação e a formação de compostos que absorvem em determinadas regiões (HERAWATI *et al.*, 2018). Como regiões ( $5000 - 5100 \text{ cm}^{-1}$  e  $7000-7200 \text{ cm}^{-1}$ ) relacionadas a absorção da água, pois a intensidade de absorbância da classe extraforte é menor que a classe tradicional, o que pode estar associada ao grau de torra, que quando submetido a uma torra acentuada possuem menores teores de umidade (ESTEBAN-DÍEZ; GONZÁLEZ-SÁIZ; PIZARRO, 2004; BAQUETA *et al.*, 2020; VILLALÓN-LÓPEZ *et al.*, 2018).

De acordo com Pascoa *et al.* (2015), os ácidos clorogênicos e sacarose são responsáveis pela absorção na região em torno de  $4700-5000 \text{ cm}^{-1}$  (banda de combinação de C-C e C-H). A diferença da intensidade de absorbância nessa região indica que as concentrações desses compostos nas amostras extraforte, mais torrado, são menores do que nos cafés tradicionais devido a maior degradação térmica desses componentes durante a torrefação, corroborando com o estudo de Monteiro e Trugo

(2005) que quando os grãos de café são submetidos a uma torra extraforte (torra escura) apresentam menores teores de ácidos clorogênicos.

No estudo de Baqueta *et al.* (2020) mostrou que a região entre 9090 a 10000  $\text{cm}^{-1}$  pode ser atribuída ao segundo sobretom  $-\text{CH}$  dos constituintes do café como cafeína e carboidratos, observa-se na **Figura 17b** que a classe tradicional apresenta maior absorvância em relação a classe extraforte, possivelmente devido á menor degradação térmica desses componentes no café (CORREIA *et al.*, 2018).



**Figura 17.** Intervalos selecionados pelo (a) modelo *i*SPA-PLS-DA ( $w=20$ ) com MM+MSC para o equipamento de bancada e (b) o modelo *i*SPA-PLS-DA ( $w=5$ ) com MM+BO para o equipamento portátil

Apesar dos resultados serem promissores para o modelo *i*SPA-PLS-DA, uma análise dos valores preditos (**Figura 16**) demonstraram que algumas amostras foram classificadas incorretamente ou apresentaram valores preditos bem próximos do limiar de classificação. Para essas amostras incertas, a análise sensorial e de cor pode ser realizada pelos analistas, como frequentemente é feito na indústria cafeeira, pois considerando a eficácia do modelo proposto, o número de amostras a serem analisadas seriam reduzidas, resultando numa otimização do processo e redução dos erros subjetivos, para monitorar a qualidade e o nível de torra durante e depois da produção, bem como do produto acabado.

**Tabela 5** – Resultados da classificação para o grau de torra em cafés solúveis usando espectros NIR de bancada e portátil pré-processados e algoritmo de classificação iSPA-PLS-DA

Método de Classificação		Pré-Processamento		MM + SNV (8*)				MM + MSC (7*)				MM + BO (10*)				MM + BO + LBC (8*)				dSG_23 (4*)			
		Amostras		Treinam.		Teste		Treinam.		Teste		Treinam.		Teste		Treinam.		Teste		Treinam.		Teste	
		Intervalos		(20/3) <sup>a</sup>		(15/5) <sup>a</sup>		(10/2) <sup>a</sup>		(5/4) <sup>a</sup>		(15/12) <sup>a</sup>											
NIR bancada	iSPA-PLS-DA	<b>TD</b>	30	-	10	2	29	1	11	1	30	-	11	1	29	1	11	1	29	1	11	1	
		<b>EF</b>	-	21	1	9	-	21	-	10	-	21	1	9	1	20	-	10	2	19	-	10	
		Sens (%)	100	100	83,3	90,0	96,7	100	91,7	100	100	100	91,7	90,0	96,7	95,2	91,7	100	96,7	90,5	91,7	100	
		Espec (%)	100	100	90,0	83,3	100	96,7	100	91,7	100	100	90,0	91,7	95,2	96,7	100	91,7	90,5	96,7	100	91,7	
		Acurácia (%)	100		86,4		98,0		95,5		100		90,9		96,1		95,5		94,2		95,5		
		Acurácia (%)	93,2				<b>96,7</b>				95,4				95,8				94,8				
Método de Classificação		Pré-Processamento		MM + SNV (7*)				MM + MSC (8*)				MM + BO (10*)				MM + BO + LBC (10*)				dSG_7 (9*)			
		Amostras		Treinam.		Teste		Treinam.		Teste		Treinam.		Teste		Treinam.		Teste		Treinam.		Teste	
		Intervalos		(5/1) <sup>a</sup>		(20/3) <sup>a</sup>		(5/2) <sup>a</sup>		(15/7) <sup>a</sup>		(20/13) <sup>a</sup>											
NIR portátil	iSPA-PLS-DA	<b>TD</b>	29	1	10	2	28	2	11	1	30	-	12	-	30	-	12	-	30	-	11	1	
		<b>EF</b>	2	19	-	10	1	20	-	10	2	19	-	10	-	21	3	7	1	20	-	10	
		Sens (%)	96,7	96,7	83,3	100	93,3	95,2	91,7	100	100	90,5	100	100	100	100	100	70,0	100	95,2	91,7	100	
		Espec (%)	90,4	90,4	100	83,3	95,2	93,3	100	91,7	90,5	100	100	100	100	100	100	70,0	100	95,2	100	100	91,7
		Acurácia (%)	94,1		90,1		94,1		95,4		96,1		100		100		86,4		98,0		95,4		
		Acurácia (%)	92,5				94,7				<b>98,0</b>				93,2				96,7				

TD: Café Tradicional, EF: Café Extraforte, Sens: Sensibilidade, Espec: Especificidade, (\*): Variáveis Latentes; MM: Média Móvel; SNV: Variação normal padrão; MSC: Correção de espalhamento multiplicativo; LBC: Correção de linha de base linear; (BO) offset; dX\_23: derivada Savitzky-Golay com 23 pontos; dSG\_7: derivada Savitzky-Golay com 7 pontos. <sup>a</sup> Número de intervalos selecionados por SPA (ex: iSPA-PLS-DA (5/2): 2 intervalos foram selecionados de 5 pelo algoritmo SPA)

*Capítulo 6*  
**Conclusões e  
Perspectivas**

---

## 6 CONCLUSÕES E PERSPECTIVAS

### 6.1 Conclusões

O presente estudo apresentou pela primeira vez a aplicação da espectroscopia NIR para classificação e diferenciação de cafés solúveis comerciais usando apenas uma medida direta da amostra. Além disso, esta metodologia foi associada no desenvolvimento de classificação de dois níveis, utilizando técnicas quimiométricas (DD-SIMCA, PLS-DA e *i*SPA-PLS-DA) para avaliação dos cafés solúveis comerciais.

No primeiro nível de classificação foi demonstrado o potencial da espectroscopia NIR juntamente com DD-SIMCA para classificação das amostras de cafés solúveis descafeinado. Em ambos os equipamentos, NIR de bancada e NIR portátil apresentaram excelente capacidade preditiva, classificando corretamente todas as amostras.

No segundo nível de classificação foi desenvolvido uma metodologia baseada em espectroscopia NIR e técnicas quimiométricas de classificação (PLS-DA e *i*SPA-PLS-DA) para discriminação das classes tradicional e extraforte. Para os espectros NIR de bancada o modelo empregado para seleção de intervalos (*i*SPA-PLS-DA/MM+MSC) obteve desempenho de classificação muito próximo ao obtido no PLS-DA/MM+MSC alcançando 96,7% de acurácia, porém foi possível identificar as variáveis relevantes para discriminação das classes estudadas. Os espectros NIR portátil obteve uma melhoria significativa na capacidade preditiva utilizando *i*SPA-PLS-DA/MM+BO quando comparado com PLS-DA/MM+BO, com 98% de acurácia.

Desse modo, pode-se inferir que a técnica NIR de bancada e portátil associada as ferramentas quimiométricas de classificação é uma alternativa promissora para análise direta do café solúvel comercial em função dos diferentes procedimentos industriais, ou seja, processo de descafeinação e grau de torra.

### 6.2 Perspectivas

Embora o método proposto tenha proporcionado resultados satisfatórios, sugerimos as seguintes propostas para dar continuidade no desenvolvimento deste trabalho:

- Quantificar os teores de umidade, cafeína e ácidos clorogênicos em amostras de cafés solúveis para correlacionar com os resultados obtidos pela Espectroscopia NIR.
- Classificar simultaneamente o tipo e o grau de torra de cafés solúveis empregando imagens digitais, DD-SIMCA, PLS-DA e iSPA-PLS-DA.
- Avaliação da análise de cor do café solúvel por imagens digitais.

## REFERÊNCIAS

ABIC. Associação Brasileira da Indústria de Café. Recomendações Técnicas. Disponível em: <https://www.abic.com.br/recomendacoes-tecnicas/recomendacoes-tecnicas-da-abic/>. Acesso em: 20 nov. 2020.

ABICS. Associação Brasileira da Indústria de Café Solúvel, 2018. Disponível em: [https://www.abics.com.br/noticia.php?noticia=79&industria\\_de\\_soluvel\\_e\\_produtores\\_de\\_conilon\\_os\\_desafios\\_sao\\_comuns#:~:text=O%20uso%20de%20conilon%20na,imbat%C3%ADvel%20quando%20comparado%20ao%20ar%C3%A1bica.](https://www.abics.com.br/noticia.php?noticia=79&industria_de_soluvel_e_produtores_de_conilon_os_desafios_sao_comuns#:~:text=O%20uso%20de%20conilon%20na,imbat%C3%ADvel%20quando%20comparado%20ao%20ar%C3%A1bica.) Acesso em: 20 nov. 2020.

ABICS. Associação Brasileira da Indústria de Café Solúvel, 2021. Disponível em: [https://www.abics.com.br/noticia.php?noticia=245&exportacao\\_de\\_cafe\\_soluvel\\_do\\_brasil\\_bate\\_recorde\\_em\\_2020.](https://www.abics.com.br/noticia.php?noticia=245&exportacao_de_cafe_soluvel_do_brasil_bate_recorde_em_2020.) Acesso em: 10 fev. 2021.

ABICS. Associação Brasileira da Indústria de Café Solúvel. Manual de Café Solúvel para baristas, 2020. Disponível em: [https://www.abics.com.br/noticia.php?noticia=233&abics\\_lanca\\_manual\\_de\\_cafe\\_soluvel\\_para\\_baristas.](https://www.abics.com.br/noticia.php?noticia=233&abics_lanca_manual_de_cafe_soluvel_para_baristas.) Acesso em: 10 fev. 2021.

AGTRON. Specialty Coffee Association. Roast color classification system. **Long Beach: SCAA**, 2010. p. 9

ALLEGRETTO, I.; MARANGONI, B.; MANZARI, P.; PORFIDO, C.; TERZANO, R.; DE PASCALE, O.; SENESI, G. S. Macro-classification of meteorites by portable energy dispersive X-ray fluorescence spectroscopy (pED-XRF), principal component analysis (PCA) and machine learning algorithms. **Talanta**, v. 212, p. 120785, 2020.

ANANDHARAMAKRISHNAN, C. Spray-freeze-drying of coffee. In: GRUMEZESCU, A. M.; HOLBAN, A. Ma. (eds.). **Caffeinated and Cocoa Based Beverages**. The Science of Beverages (v. 10, p. 337–366). Duxford: Elsevier, p. 337-366, 2019.

ARAÚJO, C. S.; VIMERCATI, W.C.; MACEDO, L.L.; FERREIRA, A.; PREZOTTI, L.C.; TEIXEIRA, L. J. Q.; SARAIVA, S. H. Predicting the Electric Conductivity and Potassium Leaching of Coffee by NIR Spectroscopy Technique. **Food Analytical. Methods**, v. 13, p. 2312–2320, 2020.

ARAÚJO, T. K. L.; NÓBREGA, R. O.; FERNANDES, D. D. S.; ARAÚJO, M. C. U.; DINIZ, P. H. G. D.; SILVA, E. C.; Non-destructive authentication of Gourmet ground roasted coffees using NIR spectroscopy and Digital Images. **Food Chemistry**, v. 364, 2021.

ASSIS, C.; OLIVEIRA, L. S.; SENA, M. M. Variable Selection Applied to the Development of a Robust Method for the Quantification of Coffee Blends Using Mid Infrared Spectroscopy. **Food Analytical Methods**, v. 11, p. 578-588, 2017.

BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models. PLS-DA. **Analytical Methods**, v. 5, 3790–3798, 2013.

BARBIN, D. F.; FELICIO, A.L.S.M.; SUN, D.; NIXDORF, S.L.; HIROOKA, E.Y. Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. **Food Research International**, v. 61, p. 23–32, 2014.

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. **Journal of Food Science**, v. 84, n. 6, p. 1247-1255, 2019.

BAQUETA, M. R.; COQUEIRO, A.; MARÇO, P. H.; VALDERRAMA, P. Quality Control Parameters in the Roasted Coffee Industry: a Proposal by Using MicroNIR Spectroscopy and Multivariate Calibration. **Food Analytical Methods**, v. 13, n. 1, p. 50-60, 2019.

BAQUETA, M. R.; COQUEIRO, A.; MARÇO, P. H.; VALDERRAMA, P. Multivariate classification for the direct determination of cup profile in coffee blends via handheld near-infrared spectroscopy. **Talanta**, v. 222, p. 121526, 2020.

BERRUETA, L. A.; ALONSO-SALCES, R. M.; HÉBERGER, K. Supervised pattern recognition in food analysis. **Journal of Chromatography A**, v. 1158, p. 196–214, 2007.

BERTONE, E.; VENTURELLO, A.; GIRAUDO, A.; PELLEGRINO, G.; GEOBALDO, F. Simultaneous determination by NIR spectroscopy of the roasting degree and Arabica/Robusta ratio in roasted and ground coffee. **Food Control**, v. 59, p. 683–689, 2016.

BLITTERSDORFF, M. V.; KLATT, C. The Grind—Particles and Particularities. **The Craft and Science of Coffee**, p. 311–328, 2017.

BRASIL. ANVISA. **Resolução RDC nº 277**, de 22 de setembro de 2005. Regulamento Técnico para Café, Cevada, Chá, Erva-Mate e Produtos Solúveis. Diário Oficial [da] República Federativa do Brasil, Brasília, DF: Ministério da Saúde, 2005. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/inspecao/produtos-vegetal/legislacao-1/biblioteca-de-normas-vinhos-e-bebidas/resolucao-rdc-no-277-de-22-de-setembro-de-2005.pdf/view>. Acesso em: 10 fev. 2021.

BRIANDET, R.; KEMSLEY, E. K.; WILSON, R. H. Approaches to Adulteration Detection in Instant Coffees using Infrared Spectroscopy and Chemometrics. **Journal of the Science of Food and Agriculture**, v. 71, n. 3, p. 359–366, 1996.

CAPORASO, N.; WHITWORTH, M. B.; CUI, C.; FISK, I. D. Variability of single bean coffee volatile compounds of Arabica and robusta roasted coffees analysed by SPME-GC-MS. **Food Research International**, 2018.

CID, M.C.; PEÑA, M. P. Coffee: Analysis and Composition. **Encyclopedia of Food and Health**, p. 225–231, 2016.

CORNEJO-BÁEZ, A. A.; PEÑA-RODRÍGUEZ, L. M.; ALVAREZ-ZAPATA, R.; VAZQUEZ-HERNÁNDEZ, M.; SÁNCHEZ-MEDINA, A. Chemometrics: a

complementary tool to guide the isolation of pharmacologically active natural products. **Drug Discovery Today**, v. 25, n. 1, p. 27-37, 2020.

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**, v. 176, p. 59–68, 2018.

COUTO, J.M.; F.A.; RESSUTE, J. B.; BELLUCO, C. Z., NASCIMENTO, M. G. Color discrimination and categorization of the roasted coffee toast and commercial ground standard. **Brazilian Journal of Development**, v.6, n.2, p. 7863-7869, 2020.

CRAIG, A. P.; FRANCA, A. S.; OLIVEIRA, L. S.; IRUDAYARAJ, J.; ILELEJI, K. Fourier transform infrared spectroscopy and near infrared spectroscopy for the quantification of defects in roasted coffees. **Talanta**, v. 134, p. 379–386, 2015.

ESTEBAN-DÍEZ, I.; GONZÁLEZ-SÁIZ, J. M.; PIZARRO, C. Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy. **Analytica Chimica Acta**, v. 525, n. 2, p. 171–182, 2004.

ESTEBAN-DÍEZ, I.; GONZÁLEZ-SÁIZ, J.M.; SAENZ-GONZALEZ, J. M.; PIZARRO, C. Coffee varietal differentiation based on near infrared spectroscopy. **Talanta**, v. 71, n. 1, p. 221–229, 2007

ESTEKI, M.; SHAHSAVARI, Z.; SIMAL-GANDARA, J. Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. **Food Control**, v. 91, p. 100-112, 2018.

FRANCA, A. S. Coffee: Decaffeination. **Encyclopedia of Food and Health**, Academic Press, p. 232–236, 2016.

FERNANDES, D. D. S.; ALMEIDA, V. E; PINTO, L.; VERAS, G.; GALVÃO, R. K H; GOMES, A.; ARAUJO, M. C. U. The successive projections algorithm for interval selection in partial least squares discriminant analysis. **Analytical Methods**, v. 41, p. 1-9, 2016.

FERNANDES, S. M.; PEREIRA, R. G. F. A.; PINTO, N. A. V. D.; NERY, M. C.; PÁDUA, F. R. M. Constituintes químicos e teor de extrato aquoso de cafés arábica (*Coffea arabica* L.) e conilon (*Coffea canephora* Pierre) torrados. **Ciência e Agrotecnologia**, v. 27, p. 1076–1081, 2003.

INSTITUTO ADOLFO LUTZ. **Métodos físico-químicos para análises de alimentos**. 4. ed. São Paulo: Instituto Adolfo Lutz, 2008. p. 1020

JIMÉNEZ-CARVELO, A. M.; CUADROS-RODRÍGUEZ, L. The occurrence: A meaningful parameter to be considered in the validation of multivariate classification-based screening methods – Application for authenticating virgin olive oil. **Talanta**, v. 208, p. 120-467, 2019.

GHIRIŞAN, A.; MICLĂUŞ, V. Comparative study of spray-drying and freeze-drying on the soluble coffee properties. **Studia Universitatis Babeş-Bolyai Chemia**, v. 62, n. 4, p. 309–316, 2017.

GRANATO, D.; SANTOS, J. S.; ESCHER, G. B.; FERREIRA, B. L.; MAGGIO, R. M. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. **Trends in Food Science & Technology**, v. 72, p. 83-90, 2018.

HASKELL, C.; KENNEDY, D.; MILNE, A.; WESNES, K.; SCHOLEY, A.B. Caffeine at levels found in decaffeinated beverages is behaviourally active. **Appetite**, v. 50, n. 2-3, p. 559, 2008.

HERAWATI, D.; GIRIWONO, P. E.; DEWI, F. N. A.; KASHIWAGI, T.; ANDARWULAN, N. Critical roasting level determines bioactive content and antioxidant activity of Robusta coffee beans. **Food Science and Biotechnology**, v. 28, p. 7-14, 2018.

HUANG, J.; ROMERO-TORRES, S.; MOSHGBAR, M. Practical considerations in data pré-processing for NIR and Raman spectroscopy. **American Pharmaceutical Review**, v. 13, n. 6, p. 116-127, 2010.

KALSCHNE, D. L., SILVA, N. K., CANAN, C., BENASSI, M. T., FLORES, E. L. M., LEITE, O. D. Main minerals and organic compounds in commercial roasted and ground coffee: an exploratory data analysis. **Química Nova**, v. 44, n. 1, p. 70-75, 2021.

KREPPER, G.; ROMEO, F.; FERNANDES, D. D. S.; DINIZ, P. H. G. D.; ARAÚJO, M. C. U.; NEZIO, M. S.; PISTONESI, M. F.; CENTURIÓN, M. E. Determination of fat content in chicken hamburgers using NIR spectroscopy and the Successive Projections Algorithm for interval selection in PLS regression (iSPA-PLS). **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v.189, p. 300–306, 2018.

LUCA, S.; FILIPPIS, M.; BUCCI, R.; MAGRÌ, A. D.; MAGRÌ, A. L.; MARINI, F. Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD, NIR and chemometrics. **Microchemical Journal**, v. 129, p. 348–361, 2016.

MALAQUIAS, J. V.; CELESTINO, S. M. C.; XAVIER, M. F. F. Optimization of the roasting conditions of arabica coffee cultivated in the cerrado area of Brazil. **Brazilian Journal of Food Technology**, Campinas, v. 21, e2016162, 2018.

MANLEY, M.; BAETEN, V. Spectroscopic Technique: Near Infrared (NIR) Spectroscopy. **Modern Techniques for Food Authentication**, p. 51–102, 2018.

MANUEL, M. N.B.; SILVA, A. C.; LOPES, G.S.; RIBEIRO, L.P.D. One-class classification of special agroforestry brazilian coffee using NIR spectrometry and chemometric tools. **Food Chemistry**, 2021.

MÁQUINA, A. D. V.; SITO, B. V.; SANTANA, F. B.; SANTOS, D. Q.; NETO, W. B. Determination of Adulteration of the B10 Blend of Diesel and Crambe Biodiesel Using Proton Nuclear Magnetic Resonance ( $^1\text{H}$  NMR) Spectroscopy with a Data Driven Soft Independent Modeling of Class Analogy (DD-SIMCA) Model. **Analytical Letters**, 2020.

MARCUCCI, C. T.; BENASSI, M. T. Teores de trigonelina, ácido 5-cafeoilquínico, cafeína e melanoidinas em cafés solúveis comerciais brasileiros. **Química Nova**, v. 36, n. 4, p. 544-548, 2013.

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. **Computers and Electronics in Agriculture**, v. 121, p. 313–319, 2016.

MEINHART, A.D.; BIZZOTTO, C.S.; BALLUS, C.A.; PRADO, M.A.; BRUNS, R.E.; FILHO, J.T.; GODOY, H.T. Optimisation of a CE method for caffeine analysis in decaffeinated coffee. **Food Chemistry**, v. 120, n. 4, p. 1155–1161, 2010.

MILANI, M. I.; ROSSINI, E. L.; CATELANI, T. A.; PEZZA, L.; TOCI, A. T.; PEZZA, H. R. Authentication of roasted and ground coffee samples containing multiple adulterants using NMR and a chemometric approach. **Food Control**, v. 112, p. 107104, 2020.

MONTEIRO, M. C.; TRUGO, L.C. Determinação de compostos bioativos em amostras comerciais de café torrado. **Química Nova**, v. 28, n.4, 2005.

MONTEIRO, P. I.; SANTOS, J. S.; BRIZOLA, V. R. A.; DEOLINDO, C. T. P.; KOOT, A.; BOERRIGTER-EENLING, R.; VAN RUTH, S.; GEORGOULI, K.; KOIDIS, A.; GRANATO, D. Comparison between proton transfer reaction mass spectrometry and near infrared spectroscopy for the authentication of Brazilian coffee: A preliminary chemometric study. **Food Control**, v. 91, p. 276–283, 2018.

MORAIS, T. C. B.; RODRIGUES, D. R.; SOUTO, U. T.C.P.; LEMOS, S. G. A simple voltammetric electronic tongue for the analysis of coffee adulterations. **Food chemistry**, v. 273, p. 31-38, 2019.

NEVES, M. G.; POPPI, R. J. Authentication and identification of adulterants in virgin coconut oil using ATR/FTIR in tandem with DD-SIMCA one class modeling. **Talanta**, v. 219, p. 121-338, 2020.

NIELSEN, S. S. **Food Analysis**. 5. ed. New York (USA): Springer, 2017.

OLIVEIRA, R. R. de. A arte de torrar café. *In: A reinvenção da cafeicultura no Paraná*. **Revista Engenharia e Agronomia do Norte Pioneiro**, ed. 1, n. 1, p. 36-37, nov. 2018.

OLIVERI, P.; SIMONETTI, R. **Chemometrics for Food Authenticity**. Applications Advances in Food Authenticity Testing, 2016.

PÁSCOA, R. N. M. J.; SARRAGUÇA, M. C.; MAGALHÃES, L. M.; SANTOS, J. R.; RANGEL, A. O. S. S.; LOPES, J. A. Use of Near-Infrared Spectroscopy for Coffee Beans Quality Assessment. *In*: PREEDY, Victor R. (ed.) - **Coffee in Health and Disease Prevention**. London: Academic Press, 2015. Chapter 103, p. 933-942.

PASQUINI, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. **Journal Brazilian Chemistry Society**, v. 14, n. 2, p. 198-212, 2003.

PASQUINI, C. Near Infrared Spectroscopy: a mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8-36, 2018.

PEREIRA, E. V. S., FERNANDES, D. D. S., ARAÚJO, M. C. U., DINIZ, P. H. G. D, MACIEL, M. I. S. *In-situ* authentication of goat milk in terms of its adulteration with cow milk using a low-cost portable NIR spectrophotometer. **Microchemical Journal**, v. 163, 2021.

PIETSCH, A. Decaffeinating coffee - processes and quality. *In*: Follmer B. **The Craft and Science of coffee**. Academic Press, Elsevier, 2017. p. 225-243.

POISSON, L.; BLANK, I.; DUNKEL, A.; HOFMANN, T. The Chemistry of Roasting—Decoding Flavor Formation. *In*: Follmer B. **The Craft and Science of coffee**. Academic Press, Elsevier, 2017. p. 273-309.

RIBEIRO, J. S.; SALVA, T. J.; FERREIRA, M. M.C. Chemometric studies for quality control of processed Brazilian coffees using DRIFTS. **Journal of Food Quality**, v. 33, p. 212-227, 2010.

RODIONOVA, O. Y.; BALYKLOVA, K. S.; TITOVA, A. V.; POMERANTSEV, A. L. Quantitative risk assessment in classification of drugs with identical API content. **Journal of Pharmaceutical and Biomedical Analysis**, v. 98, p. 186-192, 2014.

RODIONOVA, O. Y.; BALYKLOVA, K. S.; TITOVA, A. V.; POMERANTSEV, A. L. Application of NIR spectroscopy and chemometrics for revealing of the 'high quality fakes' among the medicines. **Forensic Chemistry**, v. 8, p. 82–89, 2018.

RODIONOVA, O. Y.; POMERANTSEV, A. L. Chemometric tools for food fraud detection: the role of target class in non-targeted analysis. **Food Chemistry**, v. 317, p. 126-448, 2020a.

RODIONOVA, O. Y.; POMERANTSEV, A. L. Popular decision rules in SIMCA: Critical review. **Journal of Chemometrics**, v. 34, p.1-21, 2020b.

RODIONOVA, O. Y.; TITOVA, A. V.; POMERANTSEV, A. L. Discriminant analysis is an inappropriate method of authentication. **TrAC Trends in Analytical Chemistry**, v. 78, p. 17-22, 2016.

ROSTAGNO, M. A.; MANCHÓN, N.; D'ARRIGO, M.; GUILLAMÓN, E.; VILLARES, A.; GARCÍA-LAFUENTE, A.; RAMOS, A.; MARTÍNEZ, J.A. Fast and simultaneous determination of phenolic compounds and caffeine in teas, mate, instant coffee, soft

drink, and energetic drink by high-performance liquid chromatography using a fused-core column. **Analytica Chimica Acta**, v. 685, n. 2, p. 204-211, 2011.

SANTANA, F. B.; SOUZA, A. M.; ALMEIDA, M. R.; BREITKREITZ, M. C.; FILGUEIRAS, P. R.; SENA, M. M.; POPPI, R. J. Experimento didático de quimiometria para classificação de óleos vegetais comestíveis por espectroscopia no infravermelho médio combinado com análise discriminante por mínimos quadrados parciais: um tutorial, parte V. **Química Nova**, v. 43, n. 3, p. 371-381, 2020.

SANTOS, K. M.; MOURA, M. F. V.; AZEVEDO, F. G.; LIMA, K. M. G.; RAIMUNDO, I. M.; PASQUINI, C. (2012). Classification of Brazilian Coffee Using Near-Infrared Spectroscopy and Multivariate Calibration. **Analytical Letters**, v. 45, n. 7, p. 774–781, 2012.

SCHENKER, S.; ROTHGEB, T. The Roast—Creating the Beans' Signature. *In*: Follmer B. **The Craft and Science of coffee**. Academic Press, Elsevier, 2017. p. 245–271, 2017.

SHAH, A.; CHAUHAN, Y.; CHAUDHURY, B. Análise de componentes principais baseada na construção e avaliação do índice de criptomoeda. **Expert Systems with Applications**, v. 163, p. 113-796, 2020.

SILVA, C. S.; PIMENTEL, M. F.; AMIGO, J. M.; HONORATO, R. S.; PASQUINI, C. Detecting semen stains on fabrics using near infrared hyperspectral images and multivariate models. **TrAC Trends in Analytical Chemistry**, v. 55, p. 23-25, 2017.

SOARES, M. J.; SAMPAIO, G. M.; GUIZELLINI, S. G.; FIGUEIRA, M. S.; PINAFFI, A. C. C.; FREITAS, R.A.M.S.; SHAHIDI, F.; CAMARGO, A. C.; TORRES, E. A. F. S. Regular and decaffeinated espresso coffee capsules: Unravelling the bioaccessibility of phenolic compounds and their antioxidant properties in milk model system upon in vitro digestion. **LWT – Food Science and Technology**, v. 135, p. 110-255, 2021.

SOUSA, R. M. N.; CANUTO, G.A.P.; DIAS, R.C.E.; BENASSI, M.T. Teores de compostos bioativos em café torrados e moídos comerciais. **Química Nova**, v. 33, n. 4, p. 885-890, 2010.

SOUZA, A. M.; POPPI, R.J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. **Química Nova**, v. 35, p. 223-229, 2012.

SOUTO, U. T. C. P.; BARBOSA, M. F.; DANTAS, H. V.; PONTES, A. S.; LYRA, W. S.; DINIZ, P. H. G. D.; ARAÚJO, M. C. U.; SILVA, E. C. Identification of adulteration in ground roasted coffees using UV-Vis spectroscopy and SPA-LDA. **LWT - Food Science and Technology**, v. 63, p. 1037-1041, 2015a.

SOUTO, U. T. C. P.; BARBOSA, M. F.; DANTAS, H. V.; PONTES, A. S.; LYRA, W. S.; DINIZ, P. H. G. D.; ARAÚJO, M. C. U.; SILVA, E. C. Screening for coffee adulteration using digital images and SPA-LDA. **Food Analytical Methods**, v. 8, p. 1515-1521, 2015b.

SUNARHARUM, W. B.; WILLIAMS, D. J.; SMYTH, H. E. Complexity of coffee flavor: A compositional and sensory perspective. **Food Research International**, v. 62, p. 315–325, 2014.

TIBOLA, C. S.; MEDEIROS, E. P.; SIMEONE, M. L. F.; OLIVEIRA, M. A. Espectroscopia no Infravermelho próximo para avaliar indicadores de qualidade tecnológica e contaminantes em grãos. **Embrapa**, 2018. Disponível em: [https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1106595/1/ID445392018LVe\\_spectroscopia.pdf](https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1106595/1/ID445392018LVe_spectroscopia.pdf). Acesso em: 16 jan. 2021.

VILLALÓN-LÓPEZ, N.; SERRANO-CONTRERAS, J. I.; TÉLLEZ-MEDINA, D. I.; ZEPEDA, L. G. An <sup>1</sup>H NMR-based metabolomic approach to compare the chemical profiling of retail samples of ground roasted and instant coffees. **Food Research International**, v. 106, p. 263–270, 2018.

VIGNOLI, J.A.; BASSOLI, D. G.; BENASSI, M. T. Antioxidant activity, polyphenols, caffeine and melanoidins in soluble coffee: The influence of processing conditions and raw material. **Food Chemistry**, v. 124, p. 863–868, 2011.

YASHIN, A.; YASHIN, Y.; XIA, X.; NEMZER, B. Chromatographic Methods for Coffee Analysis: **A Review**. **Journal of Food Research**, v. 6, n. 4, p. 60-82, 2017.

WANG, X.; LIM, L. Physicochemical Characteristics of Roasted Coffee. *In*: PREEDY, Victor R. Academic Press (ed.). **Coffee in Health and Disease Prevention**, p. 247-254, 2015.

WEI, F.; TANOKURA, M. Chemical changes in the components of coffee beans during roasting. *In*: PREEDY, Victor R. Academic Press (ed.). **Coffee in Health and Disease Prevention**, p. 83-91, 2015.

ZABOT, GIOVANI, L. Decaffeination using supercritical carbon dioxide. *In*: INAMUDDIN; ASIRI, A. M.; ISLOOR, A. M. **Green Sustainable Process for Chemical and Environmental Engineering and Science**, p. 255–278, 2020.

ZAREEF, M.; CHEN, Q.; HASSAN, M. M.; ARSLAN, M.; HASHIM, M. M.; AHMAD, W.; KUTSANEDZIE, F. Y. H.; AGYEKUM, A. A. An Overview on the Applications of Typical Non-linear Algorithms Coupled with NIR Spectroscopy in Food Analysis. **Food Engineering Reviews**, v. 12, p. 173-190, 2020

ZHANG, K.; ZHANG, A.; LIU, R.; ZHANG, H.; LIN, H.; ZHANG, P.; HUANG, P.; WANG, Z. Identifying muscle hemorrhage in rat cadavers with advanced decomposition by FT-IR microspectroscopy combined with chemometrics. **Legal Medicine**, v. 47, p. 101-748, 2020.

ZONTOV, Y. V.; RODIONOVA, O.Y.; KUCHERYAVSKIV, S.V.; POMERANTSEY, A.L. DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach. **Chemometrics and Intelligent Laboratory Systems**, v. 167, p. 23–28, 2017.

## Anexo 1

**Anexo 1.** Resultados da classificação para o grau de torra em cafés solúveis usando espectros NIR de bancada pré-processados e algoritmos de classificação PLS

	Conjunto de amostras	Treinamento			Teste		
		Parâmetros					
	VL	Sens.	Espec.	Acurácia	Sens.	Espec.	Acurácia
<b>MM + SNV</b>							
PLS-DA	10	100	93,33	96,08	90,00	83,33	86,36
iSPA-PLS -DA (5/2) <sup>a</sup>	10	100	100	100	90,00	75,00	81,82
iSPA-PLS -DA (10/2) <sup>a</sup>	10	100	100	100	80,00	66,67	72,73
iSPA-PLS -DA (15/4) <sup>a</sup>	10	100	100	100	80,00	66,67	72,73
iSPA-PLS-DA (20/3) <sup>a</sup>	8	100	100	100	90,00	83,33	86,36
iSPA-PLS -DA (25/7) <sup>a</sup>	10	100	100	100	80,00	75,00	77,27
iSPA-PLS-DA (30/10) <sup>a</sup>	9	100	96,67	98,04	80,00	83,33	81,82
<b>MM + MSC</b>							
PLS-DA	7	95,24	96,67	96,08	100	91,67	95,45
iSPA-PLS -DA (5/4) <sup>a</sup>	7	95,24	96,67	96,08	100	91,67	95,45
iSPA-PLS -DA (10/6) <sup>a</sup>	8	100	96,67	98,04	90,00	91,67	90,91
iSPA-PLS -DA (15/5) <sup>a</sup>	7	100	96,67	98,04	100	91,67	95,45
iSPA-PLS-DA (20/8) <sup>a</sup>	8	100	96,67	98,04	100	83,33	90,91
iSPA-PLS -DA (25/9) <sup>a</sup>	7	100	96,67	98,04	90,00	83,33	86,36
iSPA-PLS-DA (30/8) <sup>a</sup>	8	100	100	100	90,00	83,33	86,36
<b>MM + BO</b>							
PLS-DA	10	95,24	96,67	96,08	90,00	91,67	90,91
iSPA-PLS -DA (5/2) <sup>a</sup>	10	100	96,67	98,04	90,00	75,00	81,82
iSPA-PLS -DA (10/2) <sup>a</sup>	10	100	100	100	90,00	90,91	90,91
iSPA-PLS -DA (15/3) <sup>a</sup>	10	100	100	100	80,00	91,67	86,36
iSPA-PLS-DA (20/5) <sup>a</sup>	10	100	100	100	100	75,00	86,36
iSPA-PLS -DA (25/4) <sup>a</sup>	10	100	100	100	90,00	83,33	86,36
iSPA-PLS-DA (30/18) <sup>a</sup>	10	100	100	100	100	83,33	90,91
<b>MM + BO + LBC</b>							
PLS-DA	8	95,24	96,67	96,08	100	91,67	95,45
iSPA-PLS -DA (5/4) <sup>a</sup>	8	95,24	96,67	96,08	100	91,67	95,45
iSPA-PLS -DA (10/4) <sup>a</sup>	8	100	93,33	96,08	100	83,33	90,91
iSPA-PLS -DA (15/2) <sup>a</sup>	7	100	96,67	98,04	90,00	66,67	77,27
iSPA-PLS -DA (20/4) <sup>a</sup>	8	100	96,67	98,04	80,00	75,00	77,27
iSPA-PLS -DA (25/2) <sup>a</sup>	8	100	100	100	90,00	83,33	86,36
<b>dX_23</b>							
PLS-DA	4	95,24	90,00	92,16	70,00	83,33	77,27
iSPA-PLS -DA (5/4) <sup>a</sup>	4	90,48	90,00	90,20	100	91,67	95,45
iSPA-PLS -DA (10/6) <sup>a</sup>	4	95,24	86,67	90,20	80,00	83,33	81,82
iSPA-PLS -DA (15/12) <sup>a</sup>	4	90,48	96,67	94,12	100	91,67	95,45
iSPA-PLS-DA (20/4) <sup>a</sup>	3	95,24	90,00	92,16	80,00	83,33	81,82
iSPA-PLS -DA (25/20) <sup>a</sup>	4	90,48	96,67	94,12	80,00	91,67	86,36
iSPA-PLS-DA (30/5) <sup>a</sup>	3	95,24	93,33	94,12	80,00	75,00	77,27

VLs: Variáveis Latentes; Sens: Sensibilidade; Espec: Especificidade; MM: Média Móvel janela de 9 pontos; SNV: Variação normal padrão; MSC: Correção de espalhamento multiplicativo; LBC: Correção de linha de base linear; (BO) offset; dX\_23: derivada Savitzky-Golay com janela de 23 pontos. <sup>a</sup>: Número de intervalos selecionados por SPA (ex: iSPA-PLS-DA (5/2): 2 intervalos foram selecionados de 5 pelo algoritmo SPA)

## Anexo 2

**Anexo 2.** Resultados da classificação para o grau de torra em cafés solúveis usando espectros NIR portátil pré-processados e algoritmos de classificação PLS

	Conjunto de amostras	Treinamento			Teste		
		Parâmetros					
		VL	Sens.	Espec.	Acurácia	Sens.	Espec.
<b>MM + SNV</b>							
PLS-DA	7	90,48	90,00	90,20	80,00	66,67	72,73
iSPA-PLS -DA (5/1) <sup>a</sup>	7	90,48	96,67	94,12	100	83,33	90,91
iSPA-PLS -DA (10/9) <sup>a</sup>	7	90,48	100	96,08	90,00	58,33	72,73
iSPA-PLS -DA (15/13) <sup>a</sup>	7	90,48	100	96,08	80,00	58,33	68,18
iSPA-PLS-DA (20/13) <sup>a</sup>	7	90,48	100	96,08	90,00	58,33	72,73
<b>MM + MSC</b>							
PLS-DA	8	85,71	93,33	90,20	80,00	66,67	72,73
iSPA-PLS -DA (5/3) <sup>a</sup>	7	90,48	96,67	94,12	80,00	66,67	72,73
iSPA-PLS -DA (10/8) <sup>a</sup>	8	85,71	96,67	92,16	80,00	66,67	72,73
iSPA-PLS -DA (15/9) <sup>a</sup>	8	95,24	96,67	96,08	90,00	75,00	81,82
iSPA-PLS-DA (20/3) <sup>a</sup>	8	95,24	93,33	94,12	100	91,67	95,45
<b>MM + BO</b>							
PLS-DA	13	100	96,77	98,04	100	75,00	86,36
iSPA-PLS -DA (5/2) <sup>a</sup>	10	90,48	100	96,08	100	100	100
iSPA-PLS -DA (10/3) <sup>a</sup>	9	90,48	96,67	94,12	100	100	95,45
iSPA-PLS -DA (15/4) <sup>a</sup>	8	95,24	93,33	94,12	80,00	83,33	81,82
iSPA-PLS-DA (20/9) <sup>a</sup>	8	95,24	93,33	94,12	80,00	83,33	81,82
<b>MM + BO + LBC</b>							
PLS-DA	13	100	96,67	98,04	70,00	91,67	81,82
iSPA-PLS -DA (5/1) <sup>a</sup>	9	100	100	100	70,00	91,67	81,82
iSPA-PLS -DA (10/3) <sup>a</sup>	13	100	100	100	70,00	91,67	81,82
iSPA-PLS -DA (15/7) <sup>a</sup>	10	100	100	100	70,00	100	86,36
iSPA-PLS -DA (20/12) <sup>a</sup>	9	100	96,67	98,04	70,00	91,67	81,82
<b>dX_7</b>							
PLS-DA	9	90,48	96,67	94,12	90,00	75,00	81,82
iSPA-PLS -DA (5/2) <sup>a</sup>	9	95,24	96,67	96,08	100	75,00	86,36
iSPA-PLS -DA (10/8) <sup>a</sup>	8	95,24	100	98,04	90,00	75,00	75,00
iSPA-PLS -DA (15/4) <sup>a</sup>	9	95,24	100	98,04	80,00	75,00	77,27
iSPA-PLS-DA (20/13) <sup>a</sup>	9	95,24	100	98,04	100	91,67	95,45

VLs: Variáveis Latentes; Sens: Sensibilidade; Espec: Especificidade; MM: Média Móvel janela de 7 pontos; SNV: Variação normal padrão; MSC: Correção de espalhamento multiplicativo; LBC: Correção de linha de base linear; (BO) offset; dX\_7: derivada Savitzky-Golay com janela de 7 pontos. <sup>a</sup>: Número de intervalos selecionados por SPA (ex: iSPA-PLS-DA (5/2): 2 intervalos foram selecionados de 5 pelo algoritmo SPA)