



**UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA  
DEPARTAMENTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

# **DISSERTAÇÃO DE MESTRADO**

**METODOLOGIAS ANALÍTICAS BASEADAS EM IMAGENS  
DIGITAIS E ESPECTROSCOPIA NIR PARA VERIFICAÇÃO  
DE ADULTERAÇÃO EM PÁPRICA DOCE**

**RUTH BEZERRA ARANHA**

**João Pessoa – PB – Brasil**

**Novembro/2024**

**UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA  
DEPARTAMENTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

## **DISSERTAÇÃO DE MESTRADO**

### **METODOLOGIAS ANALÍTICAS BASEADAS EM IMAGENS DIGITAIS E ESPECTROSCOPIA NIR PARA VERIFICAÇÃO DE ADULTERAÇÃO EM PÁPRICA DOCE**

**RUTH BEZERRA ARANHA\***

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

Orientador: **Prof. Dr. Márcio José Coelho de Pontes**

Coorientador: **Dr. Sófacles Figueredo Carreiro Soares**

**\*Bolsista FAPESQ (EDITAL N° 16/2022)**

**João Pessoa – PB – Brasil  
Novembro/2024**

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

A662m Aranha, Ruth Bezerra.

Metodologias analíticas baseadas em imagens digitais e espectroscopia NIR para verificação de adulteração em páprica doce / Ruth Bezerra Aranha. - João Pessoa, 2024.

88 f. : il.

Orientação: Márcio José Coelho de Pontes.

Coorientação: Sófacles Figueredo Carreiro Soares.  
Dissertação (Mestrado) - UFPB/CCEN.

1. Química analítica. 2. Especiarias. 3. Segurança de alimentos. 4. Quimiometria. 5. Imagens digitais. 6. Páprica doce - Adulteração. I. Pontes, Márcio José Coelho de. II. Soares, Sófacles Figueredo Carreiro. III. Título.

UFPB/BC

CDU 543(043)

# “Metodologias analíticas baseadas em imagens digitais e espectroscopia NIR para verificação de adulteração em páprica doce”.

Dissertação de Mestrado apresentada pela discente **Ruth Bezerra Aranha** e aprovada pela banca examinadora em 01 de novembro de 2024.

Documento assinado digitalmente  
 **MARCIO JOSE COELHO DE PONTES**  
Data: 05/11/2024 15:20:44-0300  
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Márcio José Coelho de Pontes DQ/UFPB  
Orientador/Presidente

Documento assinado digitalmente  
 **SOFACLES FIGUEREDO CARREIRO SOARES**  
Data: 05/11/2024 09:13:55-0300  
Verifique em <https://validar.iti.gov.br>

Dr. Sófacles Figueredo Carreiro Soares CT/UFPB  
Co-Orientador

Documento assinado digitalmente  
 **EVERALDO PAULO DE MEDEIROS**  
Data: 04/11/2024 18:09:36-0300  
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Everaldo Paulo de Medeiros EMBRAPA/Campina Grande-PB  
Examinador externo

Documento assinado digitalmente  
 **RICARDO ALEXANDRE CAVALCANTI DE LIMA**  
Data: 04/11/2024 17:50:30-0300  
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Ricardo Alexandre Cavalcanti de Lima DQ/UFPB  
Examinador interno

**Dedico este trabalho a um eterno amigo,  
Thiago Marinho (*in memoriam*). Sei que esse  
era o seu sonho e o realizo por você, não na  
mesma qualidade, mas com o mesmo sonho!**

## AGRADECIMENTOS

Ao Deus eterno! Obrigada pela oportunidade de iniciar e finalizar este trabalho, meu Senhor.

Ao meu lindo e admirável esposo, Zildomar Aranha, por todo suporte e apoio que me foi dado. Sem os seus abraços nos momentos que eu me desesperei, nada disso seria possível!

A minha família:

- Aos meus avós, Eduardo e Severina, que me deram a base e o suporte necessário para seguir meus objetivos;
- À minha mãe, Rubia Bezerra, por nunca ter me abandonado e meu padrasto, João Nunes, por ser uma pessoa importante na família;
- Aos meus irmãos mais novos, Moab Bezerra e Gabriel Lucas, por causa de vocês me esforço o máximo para concluir meus planos;
- Ao meu tio, Gilmar Fernandes, por sempre ter dado apoio e suporte para a minha educação.

Ao Prof. Dr. Márcio José Coelho, pela orientação e grande contribuição durante a minha jornada acadêmica.

Ao meu coorientador, Dr. Sófacles Soares, pelas orientações, por todo tempo dedicado a esta pesquisa e por todos os conselhos!

Aos professores, Dr. Everaldo Paulo, Dr. Ricardo Alexandre, Dra. Simone Simões e Dr. Edvan Cirino, por todas as contribuições e correções para a dissertação.

Aos colegas e amigos do LAQA, por todas as conversas, risadas, abraços e conhecimentos compartilhados.

A Universidade Federal da Paraíba, por me fornecer capacitação e formação.

A Fundação de Apoio à Pesquisa do Estado da Paraíba (EDITAL N° 16/2022), pela bolsa concedida.

**“Porque sou EU que conheço os planos que tenho para vocês”, diz o Senhor, “planos de fazê-los prosperar e não de lhes causar dano, planos de dar-lhes esperança e um futuro.”**  
**Bíblia Sagrada - Jeremias 29:11**

# **METODOLOGIAS ANALÍTICAS BASEADAS EM IMAGENS DIGITAIS E ESPECTROSCOPIA NIR PARA VERIFICAÇÃO DE ADULTERAÇÃO EM PÁPRICA DOCE**

## **RESUMO**

As especiarias são produtos feitos a partir de partes de uma mesma espécie de vegetal, sendo utilizadas para dar sabor, cor, aroma e características a alimentos e bebidas. A páprica é uma especiaria mundialmente conhecida, produzida a partir do pimentão vermelho. Sua comercialização ocorre em pequenas quantidades, mas tem um valor relativamente elevado, levando a práticas de adulteração por comerciantes e produtores com o objetivo de obter ganhos econômicos. Para identificar essas adulterações, é necessário desenvolver metodologias de detecção. Neste sentido, este trabalho propõe o desenvolvimento de metodologias analíticas, baseadas em espectrometria NIR e imagens digitais, para a detecção de adulteração em páprica doce com amido e com urucum. Para esta finalidade, foram criados e comparados os modelos SPA-LDA, GA-LDA e PLS-DA, utilizando espectros NIR com diferentes métodos de pré-processamento e os histogramas das imagens digitais geradas a partir dos diversos canais de cores. Na classificação utilizando NIR, para as amostras adulteradas com amido, todos os modelos apresentaram resultados satisfatórios, com taxa de classificação correta de 100%, exceto para o modelo GA-LDA utilizando os dados pré-processados com segunda derivada de Savitzky–Golay com Janela de 11 pontos, que alcançou uma taxa de classificação correta de 94%. Os modelos PLS-DA construídos para a classificação das amostras adulteradas com urucum, classificaram todas as amostras corretamente quando foi utilizada primeira derivada de Savitzky–Golay com janela de 7 pontos e segunda derivada de Savitzky–Golay com Janela de 11 pontos. Já para os modelos LDA, os dados sem pré-processamento alcançaram 100% de taxa de classificação correta utilizando o SPA na seleção de variáveis, enquanto os modelos GA-LDA apresentaram melhor desempenho com a correção de linha de base offset e 2ª Der. S.G 11\*. Na classificação usando as imagens digitais, quando o PLS-DA foi construído para a identificação da adulteração com amido, a taxa de classificação correta foi de 0,9400 e com o urucum como adulterante obteve 0,8000. O melhor resultado do SPA-LDA para os dois adulterantes foi alcançado quando o canal G (verde) foi utilizado, com taxa de classificação correta de 0,9400 para o amido e 0,8200 para o urucum. Para o modelo GA-LDA, com o amido como adulterante, a combinação RGB + HSI apresentou resultados mais satisfatórios, com taxa de classificação correta de 92%. Já para o adulterante urucum, os melhores resultados foram alcançados com o canal G, com 88% de taxa de classificação correta. Por fim, mediante os resultados obtidos nessa pesquisa, as metodologias propostas apresentam vantagens, pois utilizam técnicas rápidas e de baixo custo, sendo alternativas promissoras para a identificação de adulterações de páprica doce com amido ou com urucum.

**Palavras-chave:** Especiarias; Segurança de alimentos; Quimiometria; Imagens Digitais.

# **ANALYTICAL METHODOLOGIES BASED ON DIGITAL IMAGES AND NIR SPECTROSCOPY FOR VERIFYING ADULTERATION IN SWEET PAPRIKA SOLD**

## **ABSTRACT**

Spices are products made from parts of the same plant species and are used to give flavor, color, aroma and characteristics to foods and beverages. Paprika is a world-renowned spice produced from red peppers. It is sold in small quantities, but has a relatively high value, leading to adulteration practices by traders and producers with the aim of obtaining economic gains. In order to identify these adulterations, it is necessary to develop detection methodologies. In this sense, this work proposes the development of analytical methodologies, based on NIR spectrometry and digital images, for the detection of adulteration in sweet paprika with starch and annatto. For this purpose, the SPA-LDA, GA-LDA and PLS-DA models were created and compared, using NIR spectra with different preprocessing methods and the histograms of digital images generated from the different color channels. In the classification using NIR, for samples adulterated with starch, all models presented satisfactory results, with a correct classification rate of 100%, except for the GA-LDA model using the pre-processed data with Savitzky–Golay second derivative with 11-point window, which achieved a correct classification rate of 94%. The PLS-DA models built for the classification of samples adulterated with annatto classified all samples correctly when the Savitzky–Golay first derivative with 7-point window and Savitzky–Golay second derivative with 11-point window were used. As for the LDA models, the data without pre-processing achieved a 100% correct classification rate using SPA in variable selection, while the GA-LDA models presented better performance with the offset baseline correction and 2nd Der. S.G 11\*. In the classification using digital images, when the PLS-DA was built to identify adulteration with starch, the correct classification rate was 0.9400 and with annatto as adulterant it obtained 0.8000. The best result of the SPA-LDA for both adulterants was achieved when the G (green) channel was used, with a correct classification rate of 0.9400 for starch and 0.8200 for annatto. For the GA-LDA model, with starch as adulterant, the RGB + HSI combination presented more satisfactory results, with a correct classification rate of 92%. For the annatto adulterant, the best results were achieved with the G channel, with an 88% correct classification rate. Finally, based on the results obtained in this research, the proposed methodologies present advantages, since they use fast and low-cost techniques, being promising alternatives for the identification of sweet paprika adulterations with starch or annatto.

**Keywords: Spices; Food safety; Chemometrics; Digital Images.**

## LISTA DE FIGURAS

|  |    |
|--|----|
| Figura II-1 - Formas de medição mais comuns para amostras sólidas empregados pela espectroscopia NIR.....  | 27 |
| Figura II-2 - Mecanismo óptico do espectrômetro NIR.....   | 29 |
| Figura II-3 - Cubo RGB.....  | 31 |
| Figura II-4 - Componentes R, G e B fragmentadas e imagem com os três canais.....   | 32 |
| Figura II-5 - Sólido HSI.....  | 33 |
| Figura III-1 – Ilustração da aquisição dos espectros NIR.....  | 41 |
| Figura III-2 - Esquema do preparo de amostras e de aquisição dos dados.....  | 41 |
| Figura III-3 - Amostras decompostas nas componentes vermelho (R), verde (G) e azul (B).....  | 42 |
| Figura IV-1 – a) Espectros NIR de páprica doce e adulterada com amido sem nenhum pré-processamento na faixa de 900 a 1700 nm e (b) espectro médio da páprica doce e do amido puro.....       | 46 |
| Figura IV-2 - Gráfico de PC1 versus PC2 aplicada aos espectros sem pré-processamento NIR para Páprica doce e das amostras adulteradas com amido.....   | 47 |
| Figura IV-3 - Gráfico de PC1 versus PC2 aplicada aos espectros NIR com correção de linha de base <i>offset</i> para Páprica adulterada com amido.....  | 48 |
| Figura IV-4 - Gráfico da taxa de erro de validação cruzada <i>versus</i> o número de variáveis latentes incluídas no modelo para os modelos PLS-DA usando diferentes pré-processamentos..... | 49 |
| Figura IV-5 - Gráfico dos valores estimados pelo PLS-DA versus o índice de cada amostra não adulteradas e adulterada com amido utilizada nos conjuntos de treinamento e teste.....           | 50 |
| Figura IV-6 - Espectros médios do subconjunto de treinamento com as variáveis selecionadas pelo SPA-LDA e GA-LDA.....  | 51 |
| Figura IV-7 - Gráfico dos scores da função discriminante de Fisher (FD1) versus o número de amostras.....  | 53 |
| Figura IV-8 - Espectros NIR da páprica sem adulteração e adulterada com urucum sem nenhum pré-processamento na faixa de 900 a 1700 nm.....   | 55 |
| Figura IV-9 - Gráfico de PC1 versus PC2 aplicada aos espectros sem pré-processamento NIR para Páprica adulterada com urucum.....   | 56 |
| Figura IV-10 - Gráfico de PC1 versus PC2 aplicada aos espectros NIR com correção de linha de base <i>offset</i> para Páprica adulterada com urucum.....                                      | 57 |
| Figura IV-11 - Taxa de erro de validação cruzada versus o número de variáveis latentes incluídas nos modelos PLS-DA.....   | 58 |
| Figura IV-12 - Valores de classe estimados pelo PLS-DA versus o gráfico de amostras de treinamento e teste das amostras não adulteradas e adulteradas com urucum.....                        | 58 |
| Figura IV-13 - Espectros médios do subconjunto de treinamento com as variáveis selecionadas pelo SPA-LDA e GA-LDA.....   | 60 |

|  |    |
|--|----|
| Figura IV-14 - Gráfico dos scores da função discriminante de Fisher (FD1) versus o número de amostras.<br>.....  | 62 |
| Figura IV-15 - Histograma com todas as escalas de cores da Páprica doce não adulterada e adulterada com amido.....   | 64 |
| Figura IV-16 - Gráfico dos scores da PC1 x PC2 x PC3 aplicada ao histograma completo para as amostras de páprica doce não adulteradas e adulterada com amido .....                   | 66 |
| Figura IV-17 - Taxa de erro de validação cruzada versus o número de variáveis latentes .....   | 67 |
| Figura IV-18 - Gráfico dos valores estimados pelo PLS-DA versus o índice de cada amostra não adulteradas e adulterada com amido utilizada nos conjuntos de treinamento e teste.....  | 68 |
| Figura IV-19 - Gráfico de scores obtido por SPA-LDA e GA-LDA para as amostras de páprica doce não adulteradas e adulterada com amido .....   | 69 |
| Figura IV-20 - Histograma com todas as escalas de cores da Páprica doce não adulterada e adulterada com Urucum.....  | 73 |
| Figura IV-21 - Gráfico dos scores da PC1 x PC2 x PC3 aplicada ao histograma completo para as amostras de páprica doce não adulteradas e adulterada com Urucum .....                  | 74 |
| Figura IV-22 - Taxa de erro de validação cruzada versus o número de variáveis latentes .....   | 75 |
| Figura IV-23 - Gráfico dos valores estimados pelo PLS-DA versus o índice de cada amostra não adulteradas e adulterada com Urucum utilizada nos conjuntos de treinamento e teste..... | 76 |
| Figura IV-24 - Gráfico de scores obtido por SPA-LDA e GA-LDA para as amostras de páprica doce não adulteradas e adulterada com Urucum .....  | 78 |

## LISTA DE TABELAS

|  |    |
|--|----|
| Tabela II-1- Regiões espectrais no Infravermelho .....   | 25 |
| Tabela III-1 - Número de amostras de treinamento e teste usadas na construção dos modelos.....   | 43 |
| Tabela IV:1 - Matriz de confusão obtida para os modelos PLS-DA para o conjunto de teste. Entre parênteses são apresentados o número de variáveis latentes utilizados. ....                       | 50 |
| Tabela IV-2 - Matriz de confusão obtida para os modelos SPA-LDA e GA-LDA construídos com os espectros NIR das amostras de pprica e pprica com amido.....                                       | 54 |
| Tabela IV-3 - Parmetros de desempenho dos resultados dos modelos SPA-LDA e GA-LDA a classe das amostras no adulteradas .....   | 55 |
| Tabela IV-4 - Matriz de confuso obtida para os modelos PLS-DA para o conjunto de treinamento. Entre parênteses so apresentados o nmero de variveis latentes utilizados. ....                 | 59 |
| Tabela IV-5 - Matriz de confuso obtida para os modelos PLS-DA para o conjunto de teste. Entre parênteses so apresentados o nmero de variveis latentes utilizados. ....                       | 60 |
| Tabela IV-6 - Matriz de confuso obtida para os modelos SPA-LDA e GA-LDA construídos com os espectros NIR das amostras de pprica e pprica com Urucum.....                                      | 63 |
| Tabela IV-7 - Parmetros de desempenho dos resultados dos modelos SPA-LDA e GA-LDA.....  | 64 |
| Tabela IV-8 - Matriz de confuso obtida para os modelos PLS-DA para o conjunto de teste do histograma de cor. Entre parênteses so apresentados o nmero de variveis latentes utilizados .....  | 69 |
| Tabela IV-9 - Matriz de confuso obtida para os modelos SPA-LDA e GA-LDA construídos com histograma de cores das amostras de pprica e pprica com amido.....                                    | 71 |
| Tabela IV-10 - Parmetros de desempenho dos modelos SPA-LDA e GA-LDA para o conjunto de teste .....  | 72 |
| Tabela IV-11 - Matriz de confuso obtida para os modelos PLS-DA para o conjunto de teste do histograma de cor. Entre parênteses so apresentados o nmero de variveis latentes utilizados ..... | 77 |
| Tabela IV-12 - Matriz de confuso obtida para os modelos SPA-LDA e GA-LDA construídos com histograma de cores das amostras de pprica e pprica com urucum .....                                 | 79 |
| Tabela IV-13 - Parmetros de desempenho dos modelos SPA-LDA e GA-LDA para o conjunto de teste .....  | 80 |

## LISTA DE ABREVIATURAS E SIGLAS

**ANVISA** - Agência Nacional de Vigilância Sanitária

**CMY** - Cyan, Magenta, Yellow

**Der. S.G** – Derivada Savitzky–Golay

**DPI** - Dots Per Inch

**EMA** - Economically Motivated Adulteration

**FD** - Funções Discriminantes

**FDA** - Food and Drug Administration

**FIR** - Far Infrared

**FT** - Fourier Transform

**GA** – Genetic Algorithm

**HPLC** - High Performance Liquid Chromatography

**HSI** - Hue, Saturation, Intensity

**KS** - Kennard-Stone

**LDA** - Linear Discriminant Analysis

**LVF** - Linear Variable Filter

**MIR** - Mid Infrared

**MSC** - Multiple Scattering Correction

**NIR** - Infravermelho Próximo

**PCA** - Análise De Componentes Principais

**PLS-DA** - Partial Least Squares Discriminant Analysis

**PPI** - Pixels Per Inch

**REM** - Radiação Eletromagnética

**RGB** - Red, Green, Blue

**SNV** - Standard Normal Variate

**SPA** - Successive Projection Algorithm

**TCC** – Taxa de Classificação Correta

**VL** - Variáveis Latentes

# SUMÁRIO

|   |           |
|---|-----------|
| <b>CAPÍTULO I : INTRODUÇÃO E OBJETIVOS.....</b>                                       | <b>15</b> |
| <b>1.1. CARACTERIZAÇÃO DA PROBLEMÁTICA.....</b>                                       | <b>16</b> |
| <b>1.2. OBJETIVOS .....</b>   | <b>19</b> |
| 1.2.1    Objetivo Geral .....   | 19        |
| 1.2.2    Objetivos Específicos .....  | 19        |
| <b>CAPÍTULO II : FUNDAMENTAÇÃO TEÓRICA .....</b>                                      | <b>20</b> |
| <b>2.1. FRAUDES EM ALIMENTOS .....</b>  | <b>21</b> |
| <b>2.2. ESPECIARIAS .....</b>   | <b>22</b> |
| 2.2.1.    Páprica .....   | 22        |
| 2.2.2.    Adulteração de especiarias .....  | 23        |
| <b>2.3. ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO .....</b>                             | <b>24</b> |
| 2.3.1.    Princípios da espectroscopia no infravermelho próximo .....                 | 24        |
| 2.3.2.    Modos de medidas dos sinais e instrumentação do NIR portátil.....           | 26        |
| 2.3.2.1. <i>Modos de medidas dos sinais para amostras em partículas</i> .....         | 26        |
| 2.3.2.2. <i>Instrumentação do espectrofotômetro NIR portátil</i> .....                | 27        |
| <b>2.4. IMAGENS DIGITAIS.....</b>   | <b>29</b> |
| 2.4.1.    Sistema de cores .....  | 30        |
| 2.4.1.1. <i>Sistema RGB</i> .....   | 31        |
| 2.4.1.2. <i>Sistema CMY e CMYK</i> .....  | 32        |
| 2.4.1.3. <i>Sistema HSI</i> .....   | 32        |
| <b>2.5. QUIMIOMETRIA.....</b>   | <b>34</b> |
| 2.5.1.    Pré-processamento.....  | 34        |
| 2.5.2.    Técnicas de Reconhecimento de Padrões .....                                 | 35        |
| 2.5.2.1. <i>Análise por componentes principais (PCA)</i> .....                        | 35        |
| 2.5.2.2. <i>Análise discriminante pelos mínimos quadrados parciais (PLS-DA)</i> ..... | 36        |
| 2.5.2.3. <i>Análise discriminante linear (LDA)</i> .....                              | 36        |
| 2.5.3.    Seleção de variáveis.....   | 37        |
| 2.5.3.1. <i>Algoritmo genético (GA)</i> .....   | 37        |
| 2.5.3.2. <i>Algoritmo das Projeções Sucessivas (SPA)</i> .....                        | 38        |

|  |           |
|--|-----------|
| <b>CAPÍTULO III : METODOLOGIA .....</b>  | <b>39</b> |
| <b>3.1. AQUISIÇÃO E PREPARO DAS AMOSTRAS .....</b>   | <b>40</b> |
| <b>3.2. REGISTRO DOS ESPECTROS NIR.....</b>  | <b>40</b> |
| <b>3.3. REGISTRO DAS IMAGENS DIGITAIS .....</b>  | <b>41</b> |
| <b>3.4. SOFTWARE E PROCEDIMENTO QUIMIOMÉTRICO .....</b>  | <b>42</b> |
| <b>CAPÍTULO IV : RESULTADOS E DISCUSSÃO.....</b>   | <b>45</b> |
| <b>4.1. AVALIAÇÃO DOS ESPECTROS NIR .....</b>  | <b>46</b> |
| 4.1.1. Adulteração da Páprica Doce com Amido .....   | 46        |
| 4.1.1.1. <i>Análise das Componentes Principais (PCA) para os espectros NIR das amostras sem adulteração e adulterada com amido.....</i>  | <i>47</i> |
| 4.1.1.2. <i>Métodos de Reconhecimento de Padrões Supervisionados aplicados as amostras não adulteradas e adulteradas com amido.....</i>  | <i>48</i> |
| 4.1.2. Adulteração da Páprica Doce com Urucum .....  | 55        |
| 4.1.2.1. <i>Análise das Componentes Principais (PCA) para os espectros NIR das amostras sem adulteração e adulterada com urucum.....</i> | <i>56</i> |
| 4.1.2.2. <i>Métodos de Reconhecimento de Padrões Supervisionados aplicados as amostras não adulteradas e adulteradas com urucum.....</i> | <i>57</i> |
| <b>4.2. AVALIAÇÃO DAS IMAGENS DIGITAIS.....</b>  | <b>64</b> |
| 4.2.1. Adulteração da Páprica Doce com Amido .....   | 64        |
| 4.2.1.1. <i>Análise das Componentes Principais (PCA) do histograma de cores da páprica e páprica com amido.....</i>                      | <i>65</i> |
| 4.2.1.2. <i>Métodos de Reconhecimento de Padrões Supervisionados.....</i>  | <i>66</i> |
| 4.2.2. Adulteração da Páprica Doce com Urucum .....  | 72        |
| 4.2.2.1. <i>Análise das Componentes Principais (PCA) do histograma de cores da páprica doce e adulterada com Urucum.....</i>             | <i>74</i> |
| 4.2.2.2. <i>Métodos de Reconhecimento de Padrões Supervisionados.....</i>  | <i>75</i> |
| <b>CAPÍTULO V : CONCLUSÃO .....</b>  | <b>81</b> |
| <b>PERSPECTIVAS.....</b>   | <b>84</b> |
| <b>REFERÊNCIAS .....</b>   | <b>85</b> |

## CAPÍTULO I : INTRODUÇÃO E OBJETIVOS

---



## 1.1. CARACTERIZAÇÃO DA PROBLEMÁTICA

As especiarias são feitas a partir de partes específicas de vegetais, como raízes, cascas, folhas, flores, frutos, sementes e talos. A Resolução 276 da Diretoria Colegiada da Anvisa (Agência Nacional de Vigilância Sanitária), de setembro de 2005, especifica quais partes das plantas devem ser processadas para se tornarem especiarias (BRASIL, 2005). Elas são utilizadas para fins culinários, conservação de alimentos e na elaboração de misturas farmacêuticas (CARDIN et al., 2021). Na cozinha, frequentemente servem para realçar o sabor dos pratos, trazendo um toque especial a diversas receitas.

A páprica é uma das especiarias amplamente utilizadas em todo o mundo, sendo obtida a partir do pimentão vermelho (*Capsicum annuum L.*) maduro, que passa pelo processo de secagem e depois é moído. O pimentão é nativo das Américas Central e do Sul, e possui uma composição rica em água, proteínas, gorduras totais, carboidratos e fibras (LAGUNA et al., 2011; LIMA et al., 2011).

Embora sejam utilizadas em pequenas quantidades, as especiarias estão entre os 25 alimentos mais adulterados (MOORE, SPINK, LIPP, 2012). A preocupação com a fraude desse tipo de alimento se deve à sutileza da adulteração, tornando difícil a identificação dessa ação pelos sentidos humanos (RODRIGUES, 2017).

A fraude envolvendo alimentos está se tornando cada vez mais comum e chegando aos consumidores com mais frequência. Com o passar do tempo, identificar produtos adulterados tem se tornado mais desafiador, pois as modificações nos alimentos podem ser sutis e imperceptíveis visualmente. O impacto dessa atividade é considerado grave. No Código Penal Brasileiro, a Lei nº 7.209, em seu artigo 272, estabelece penas de reclusão de até 8 anos para quem pratica a adulteração de alimentos (BRASIL, 1940). Tal prática, pode envolver a adição de substâncias nocivas, como corantes ilegais ou produtos químicos que podem ser tóxicos e causar reações alérgicas nos consumidores (GALVIN-KING et al., 2018).

Segundo Banti (2020) pode ser difícil encontrar itens como farinha, óleo, leite, chá, café e especiarias totalmente isentos de adulterantes. Além destes, diversos alimentos e bebidas são vulneráveis à adulteração, especialmente quando se tornam escassos no mercado.

Estudos já publicados identificaram adulterações em amostras comerciais de especiarias, comprovando esse tipo de infração e enfatizando a necessidade de fiscalização. Pesquisas realizadas por cientistas do Instituto Adolfo Lutz (PAULI-YAMADA et al., 2021) revelaram que, ao analisar amostras comerciais de páprica, 30% estavam adulteradas. O estudo



identificou que os adulterantes mais frequentes eram o amido de milho (*Zea mays L.*) e o urucum (*Bixa orellana*).

Nesse contexto, é extremamente importante encontrar formas de identificar fraudes em alimentos, para que os responsáveis por essas atividades ilegais possam ser identificados e punidos, além de garantir a segurança dos alimentos para os consumidores. Técnicas Analíticas têm um papel fundamental para identificar essas ações fraudulentas.

A técnica cromatográfica tem uma ampla aplicação e permite a separação, identificação e determinação de componentes químicos em misturas complexas (SKOOG et al., 2005). No entanto, possui uma instrumentação complexa, necessita de procedimentos de preparo de amostra e trata-se de uma técnica destrutiva para a amostra.

A espectroscopia na região do infravermelho próximo (NIR) ocorre principalmente por meio de sobretons, combinações e ressonâncias de modos vibracionais fundamentais (PASQUINI, 2018). Esta técnica vem se tornando amplamente utilizada nas indústrias, pois oferece uma análise rápida, não destrutiva e de baixo custo, permitindo a avaliação de amostras com pouco ou nenhum pré-tratamento. Com o avanço da tecnologia, a espectrometria NIR tornou-se mais acessível, com equipamentos cada vez menores, incluindo modelos portáteis, que apresentam funcionalidade comparável à dos equipamentos de bancada (PASQUINI, 2018).

Outra técnica que vem sendo bastante utilizada para a detecção de adulteração em alimentos é a análise química utilizando o processamento de imagens digitais. Após capturadas, essas imagens são processadas e os dados extraídos, fazendo com que seja possível relacionar a resposta dos sistemas de cores com alterações que ocorrem com as amostras. As imagens podem ser capturadas por meio de câmeras (ARAÚJO et al., 2018), webcams (FRANÇA, 2020), scanners (PEREIRA, 2022) e até mesmo celulares (RIOS; BREITKREITZ, 2021).

A quimiometria é essencial tanto na espectroscopia NIR quanto na análise de imagens digitais. Esses dados frequentemente contêm uma grande quantidade de informações, tornando necessário um tratamento adequado para sua interpretação. A quimiometria pode ser descrita como a aplicação de ferramentas matemáticas e estatísticas para resolver problemas relacionados a dados químicos (FERREIRA, 2015). Na análise de dados multivariados, a quimiometria é uma ferramenta fundamental, pois permite identificar variações e tendências na matriz de dados estudados.

Em face do exposto, torna-se de elevada importância o desenvolvimento de metodologias analíticas rápidas e seguras, como a espectrometria NIR e o processamento de



imagens digitais, para a detecção de adulteração em alimentos, em especial, especiarias, que vem se tornando cada vez mais alvo de ações fraudulentas.

## 1.2. OBJETIVOS

### 1.2.1 Objetivo Geral

Desenvolver metodologias analíticas baseadas na espectroscopia na região do infravermelho próximo e processamento de imagens digitais para a detecção de adulteração de páprica doce com amido ou urucum.

### 1.2.2 Objetivos Específicos

- Investigar os diferentes níveis de adulteração avaliando de acordo com os sinais das misturas por meio da espectrometria NIR e imagens digitais;
- Registrar os espectros de NIR utilizando um equipamento portátil;
- Capturar as imagens digitais das amostras por meio de um scanner;
- Extrair o histograma de cores das imagens digitais obtidas;
- Realizar um estudo exploratório dos dados através da análise de componentes principais (PCA) aos espectros NIR;
- Aplicar métodos de reconhecimento de padrões supervisionados LDA (SPA e GA) e PLS-DA aos espectros NIR para classificação de páprica não adulteradas e adulterada com amido e com urucum;
- Aplicar a análise de componentes principais (PCA) para realização de análise exploratória do histograma de cores das imagens digitais;
- Aplicar métodos de reconhecimento de padrões supervisionados LDA (SPA e GA) e PLS-DA ao histograma de cores completo, canais de cores separadamente e combinações para classificação de páprica não adulterada e adulterada com amido ou urucum.



## **CAPÍTULO II : FUNDAMENTAÇÃO TEÓRICA**

---



## 2.1.FRAUDES EM ALIMENTOS

A fraude alimentar tem se tornado cada vez mais comum na vida dos consumidores. Com o tempo, a identificação de produtos adulterados vem se tornando mais difícil, pois as alterações podem ser sutis e não perceptíveis aos sentidos humanos. Esses eventos não são recentes; Silva (2018) menciona que já na época dos gregos e romanos, documentos históricos relatavam a adulteração de diversos produtos, incluindo vinho, vinagre e farinha de trigo.

De acordo com a FDA (*Food and Drug Administration*), a adulteração econômica intencional (EMA - Economically Motivated Adulteration) ocorre quando alguém intencionalmente deixa de fora, remove ou substitui um ingrediente ou parte de um alimento para fazê-lo parecer melhor ou de maior valor (Food and Drug Administration, 2024).

A adulteração de alimentos é um crime que pode acarretar diversos problemas, apresentando um impacto sério ao colocar a saúde dos consumidores em risco. Essa prática criminosa pode ter repercussões regionais, nacionais ou internacionais, representando um risco significativo para a segurança dos alimentos e causando perdas financeiras diretas para consumidores e empresas. No Código Penal Brasileiro, a Lei nº 7.209 estabelece, no artigo 272, uma pena de reclusão de até 8 anos para práticas de adulteração de alimentos (BRASIL, 1940). No entanto, um Projeto de Lei de 2017 propõe o aumento das penas para crimes de fraude, falsificação e adulteração de alimentos e bebidas, elevando a reclusão de 8 para até 15 anos (PL 7.664/2017, 2017). O impacto econômico exato da fraude alimentar é difícil de estimar, pois tais ações são frequentemente planejadas para evitar a detecção, conforme indica a FDA (FOOD AND DRUG ADMINISTRATION, 2024).

Diversos alimentos e bebidas são vulneráveis à adulteração, principalmente quando se tornam escassos do mercado. Atualmente, o consumidor ser enganado por produtos falsificados e não ter ciência disto é mais frequente do que se imagina. Desta forma, pode ser difícil encontrar itens como farinha, leguminosas, óleo, frutas, vegetais, leite, doces, especiarias, chá, café, mel e sucos de frutas totalmente isentos de adulterantes (BANTI, 2020). Neste sentido, encontrar formas de identificar fraude alimentar é extremamente importante para que os autores destas atividades ilegais possam ser identificados e punidos, e, principalmente, garantir a segurança dos alimentos aos consumidores.

A fraude alimentar pode causar sérios problemas de saúde e, em alguns casos, até levar à morte. Um tipo de fraude que ocorre em especiarias é quando o produto com alto valor de



mercado é misturada com outros materiais vegetais não relacionados. Outro tipo de fraude é o uso de corantes para conferir uma certa coloração às especiarias, especialmente quando a cor afeta fortemente a percepção de qualidade (FOOD AND DRUG ADMINISTRATION, 2024).

## 2.2.ESPECIARIAS

As especiarias são produtos derivados de partes específicas oriundas de diversos tipos de plantas, que podem ser: raízes, rizomas, bulbos, cascas, folhas, flores, frutos, sementes e talos, conforme a RDC 276. Por outro lado, os temperos são obtidos pela mistura de ervas com outros ingredientes, podendo ser ou não fermentados, e são usados para adicionar sabor ou aroma aos alimentos e bebidas (BRASIL, 2005). Desta forma, podemos entender que a especiaria tem uma forma mais pura e específica, diferente do tempero, que pode ser a mistura de várias especiarias.

As especiarias são empregadas para fins culinários, conservação de alimentos, e no preparo de misturas farmacêuticas (CARDIN, et al., 2021). Na cozinha muitas vezes é usada para dar sabor aos alimentos conferindo o diferencial em algumas receitas. Historicamente, seu papel foi mais amplo. No Antigo Egito, especiarias como canela, anis, cominho e manjerona eram componentes essenciais no processo de embalsamamento. Já na Europa medieval foram usadas para a conservação dos alimentos (MENESES, 2019).

### 2.2.1. Páprica

A páprica é uma especiaria utilizado mundialmente, produzida a partir do pimentão vermelho maduro, que passa por um processo de desidratação em estufa e em seguida é moído. Comercialmente, podemos encontrar a páprica doce, páprica defumada e páprica picante. O pimentão vermelho (*Capsicum annuum L.*) é nativo das Américas Central e do Sul e é cultivado em regiões tropicais e temperadas, ocupando, no Brasil, uma das dez posições entre as hortaliças mais importantes economicamente (DE OLIVEIRA, LUZ, 1998). O pimentão pertencente à família das Solanáceas, que inclui batata, tomate, jiló, berinjela e pimentas (MELO et al., 2009), sendo rico em água, proteínas, gorduras totais, carboidratos e fibras. Também contém em menor quantidade nutrientes como cálcio, magnésio, fósforo, ferro, potássio, vitamina C e vitaminas do complexo B. Além disso, ele acrescenta cor, aroma e sabor a diversas receitas (LAGUNA et al., 2011; LIMA et al., 2011).



### 2.2.2. Adulteração de especiarias

Apesar das especiarias serem usadas em pequenas quantidades, a garantia de um produto puro, que assegura qualidade e proporciona características sensoriais específicas, tem gerado preocupações entre os profissionais de segurança de alimentos, uma vez que esses alimentos podem ser consumidos sem cozimento ou preparo mais rigoroso. Além disso, devido ao seu alto valor comercial, elas se tornam alvos potenciais de adulteração (KUCHARSKA-AMBROŻEJ; KARPINSKA, 2020).

A principal causa da adulteração de especiarias é o aumento da demanda ou a falta de oferta, frequentemente associada por fatores naturais. Recentemente, a motivação por trás das fraudes com especiarias tem se tornado predominantemente econômica, com o objetivo de obter margens de lucro mais altas. A adulteração pode envolver a adição de substâncias nocivas, como corantes ilegais ou produtos químicos que podem ser tóxicos para os consumidores, além de causar reações alérgicas, pois o adulterante não é informado no rótulo (GALVIN-KING, et al., 2018). Corantes à base de chumbo e outros corantes industriais que podem causar problemas de saúde, como câncer, foram encontrados em especiarias como pó de chili, cúrcuma e cominho (OSMAN, 2019; FOOD AND DRUG ADMINISTRATION, 2024).

Em um levantamento entre o período de 1980-2010, as especiarias estiveram entre os 25 alimentos mais adulterados, representando uma parte com 10% dos registros (MOORE, SPINK, LIPP, 2012). Na literatura é possível localizar alguns estudos que encontraram, através de metodologias analíticas, as adulterações mais sofisticadas, que são aquelas que não podem ser descobertas ao olhar ou ingerir o produto (RODRIGUES, 2017).

Em um estudo conduzido por Forsyth et al. (2019), foi identificado que o cromato de chumbo (II) ( $PbCrO_4$ ) foi adicionado ao açafrão para intensificar a sua cor amarela. Das amostras de nove produtores analisadas, apenas dois não apresentaram a presença deste contaminante. A equipe responsável pelo estudo solicitou uma intervenção urgente, pois os produtores do açafrão não estão cientes dos riscos que essa adulteração pode representar para a saúde humana.

Segundo Johnson e colaboradores (2014) já foram encontradas em amostras de açafrão adulterantes como glicerina, pó de sândalo, tartrazina (um corante amarelo conhecido como E102), sulfato de bário e bórax. Alguns registros indicam ainda a adição de amido, sementes de mamão, trigo sarraceno, farinha, galhos e painço à pimenta-do-reino moída.



Black et al. (2016) investigaram a adulteração do orégano comercial, identificando a presença de folhas de oliveira, murta, sumagre, esteva e avelã. Os resultados mostraram que cerca de 24% das amostras de orégano testadas estavam adulteradas, com níveis de adulteração variando de 30% a mais de 70%. Em duas amostras, o orégano estava praticamente ausente. Em um outro estudo Wielogorska e colaboradores (2018) também analisaram amostras de orégano e descobriram que 50% delas continham algum tipo de adulterante.

De Lima e colaboradores (2020) desenvolveram uma metodologia que utiliza a espectroscopia no infravermelho próximo para detectar adulterações em amostras comerciais de cominho em pó, utilizando amido e farinha de mandioca como adulterantes. Os autores ressaltaram que a espectroscopia NIR apresentou resultados satisfatórios, demonstrando ser uma técnica eficaz para essa aplicação.

Em estudos conduzidos por pesquisadores do Instituto Adolfo Lutz (PAULI-YAMADA et al., 2021), foi constatado que 13, das amostras 43 de páprica analisadas, apresentaram adulteração. Neste estudo, foi identificado que os adulterantes mais comuns encontrados foram o amido de milho (*Zea mays L.*) e o urucum (*Bixa orellana*).

A elaboração de metodologias analíticas para detectar adulteração em alimentos tem se tornado alvo de estudos científicos, propondo recursos analíticos que podem ajudar as organizações responsáveis pelas fiscalizações a identificarem possíveis irregularidades na qualidade dos alimentos. Técnicas como a espectroscopia na região do infravermelho e de imagens digitais são propostas analíticas com um potencial em detecção de adulteração em alimentos.

## 2.3. ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO

### 2.3.1. Princípios da espectroscopia no infravermelho próximo

A radiação na região do infravermelho foi descoberta em 1800 por um astrônomo alemão chamado Frederick William Herschel, que estava investigando qual cor da radiação solar produzia calor. Ele empregou a radiação do sol a um prisma de vidro, com o objetivo de decompor a luz branca em seus comprimentos de onda visíveis. Com o auxílio de um termômetro, pôde verificar a temperatura produzida por cada cor. Herschel notou que a variação mais intensa de temperatura foi alcançada quando o termômetro foi colocado logo acima da região vermelha do espectro, chamando esta radiação de “raios de aquecimento”, do inglês,



*heating rays*. Ele realizou muitos experimentos detalhados sobre seu comportamento e estudou a energia solar e a luz, estabelecendo uma base sólida para o conhecimento moderno da radiação infravermelha (RING, 2000).

A radiação nessa faixa de energia promove as variações de energia molecular associadas às transições vibracionais de estiramento e deformação dos ângulos as ligações. Essas transições são provocadas pela absorção ou emissão de um fóton, não sendo suficientemente energético para causar uma transição eletrônica (PAVIA et al., 2021; LIMA et al., 2009).

A região espectral do infravermelho envolve radiação com comprimentos de onda de 780 a 100.000 nm. Dentro dessa faixa o infravermelho pode ser classificado em três regiões diferentes. Na Tabela II:1 podemos encontrar a região espectral para o infravermelho próximo (NIR – Near Infrared), médio (MID – Middle Infrared) e distante (FIR - FAR Infrared) (SKOOG; HOLLER; NIEMAN, 2002).

Tabela II:1- Regiões espectrais no Infravermelho

| Região   | Intervalo de Comprimento de Ondas (nm) | Intervalo de Número de Ondas, $\text{cm}^{-1}$ | Região de Frequência, Hz                    |
|----------|--|--|---|
| Próximo  | 780 – 2500                             | 12.800 a 4.000                                 | $3,8 \times 10^{14}$ a $1,2 \times 10^{14}$ |
| Médio    | 2500 - 50000                           | 4.000 a 200                                    | $1,2 \times 10^{14}$ a $6,0 \times 10^{12}$ |
| Distante | 50000 - 100000                         | 200 a 10                                       | $6,0 \times 10^{12}$ a $3,0 \times 10^{11}$ |

Fonte: Skoog, Holler e Nieman (2002)

A espectrometria do infravermelho próximo é um tipo de espectroscopia vibracional que emprega uma REM cuja energia dos fótons ( $h\nu$ ) no faixa de  $2,65 \times 10^{-19}$  a  $7,96 \times 10^{-20}$  J, que corresponde à faixa de comprimento de onda de 780 a 2.500 nm, como apresentando na Tabela II:1.

A interação da radiação eletromagnética na faixa espectral NIR ocorre principalmente por meio de sobretons, combinações e ressonâncias de (ou envolvendo) modos vibracionais fundamentais. Esses modos ativos de NIR estão predominantemente relacionados a modos vibracionais anarmônicos presentes em grupos funcionais moleculares que incluem um átomo relativamente com maior massa, ligado a um átomo de hidrogênio, como C, N, O e S. Além disso, vibrações em ligações químicas fortes entre átomos mais pesados, como C=O, também podem ser detectadas através do NIR. Esta técnica possui uma vasta aplicação nos estudos de



compostos orgânicos com presença em suas estruturas moleculares ligações C-H, N-H, O-H, S-H (PASQUINI, 2003; PASQUINI, 2018).

São chamadas de sobretons as transições  $0 \rightarrow n$  ( $n > 1$ ). As frequências vibracionais fundamentais de uma molécula correspondem à transição  $0 \rightarrow 1$ , sendo o primeiro sobreton relacionado à transição  $0 \rightarrow 2$ , o segundo sobreton à transição  $0 \rightarrow 3$  e assim por diante.

Além dos sobretons, também podem ser observadas combinações de diferentes transições vibracionais. No entanto, a probabilidade dessas transições diminui significativamente com sua ordem, e geralmente as bandas de absorção correspondentes às vibrações de sobretons ou combinações têm intensidade muito menor do que suas equivalentes fundamentais. Ao contrário do MIR, a região NIR contém quase exclusivamente bandas de absorção que podem ser atribuídas às vibrações de sobretons e combinações. Devido à sobreposição dessas bandas a especificidade da espectroscopia NIR, para fins de interpretação espectral, torna-se mais complexa, sendo essa uma das principais razões pelas quais a técnica foi negligenciada durante anos. Com o surgimento da Quimiometria, a técnica NIR começou a ser utilizada de maneira mais ampla e hoje tornou-se uma técnica analítica de extrema importância no controle de qualidade em vários setores industriais (BURNS; CIURCZAK, 2007).

A espectroscopia NIR é, portanto, amplamente adotada na agricultura e em várias indústrias. Tendo como destaque o curto tempo até o resultado e não ser necessário o preparo de amostra, acarretando um número maior de análises em um menor tempo, quando comparada a técnicas convencionais, como a cromatografia de alta eficiência (HPLC) (BEĆ, et al., 2021).

Com o avanço da tecnologia digital (eletrônica e de software), a espectroscopia NIR tornou-se uma técnica mais acessível, levando ao surgimento de espectrômetros NIR portáteis e autônomos sendo visto como um avanço em várias funções estabelecidas, com a mesma funcionalidade dos equipamentos de bancadas (PASQUINI, 2018).

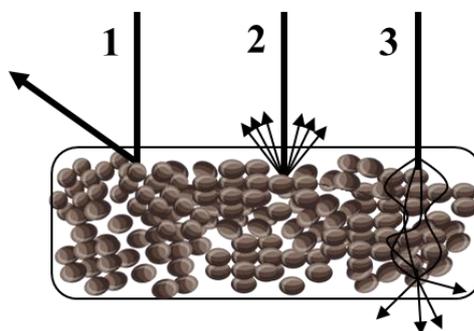
### 2.3.2. Modos de medidas dos sinais e instrumentação do NIR portátil

#### 2.3.2.1. Modos de medidas dos sinais para amostras em partículas

A Figura II-1 representa os modos de interação da radiação eletromagnética com amostras sólidas que são empregados pela espectroscopia NIR.



Figura II-1 - Formas de medição mais comuns para amostras sólidas empregados pela espectroscopia NIR



Fonte: Autor, 2024.

A radiação eletromagnética monocromática produzida por um instrumento NIR pode interagir com amostras sólidas através de refletância especulativa (Figura II-1-1), refletância difusa (Figura II-1-2) e transmitância através de meios de dispersão (Figura II-1-3). Dentre estas, a refletância difusa, se destaca pela simplicidade na aquisição das informações espectrais na região do NIR (PASQUINI, 2003).

A Refletância Difusa ocorre em superfícies como pós, grãos ou superfícies ásperas, em que a luz é refletida em várias direções, diferentemente da Refletância Especulativa, em que a luz é refletida em um ângulo igual ao ângulo de incidência. Além dessa diferença, a refletância difusa envolve processos mais complexos, como reflexão, refração e difração, e é frequentemente modelada por teorias que consideram a distribuição aleatória das superfícies (TORRENT; BARRÓN, 2008).

Quando a análise é de amostras em pó ou grãos, a refletância difusa permite que a luz interaja mais intimamente com a amostra devido a múltiplas reflexões e dispersões, aumentando a probabilidade de absorção e, portanto, a quantidade de informação espectroscópica obtida (TORRENT; BARRÓN, 2008).

#### 2.3.2.2. Instrumentação do espectrofotômetro NIR portátil

A estrutura de um espectrofotômetro NIR portátil segue um esquema geral de qualquer instrumento utilizado em espectroscopia de absorção óptica. Os principais componentes incluem uma fonte de luz, um seletor de comprimento de onda e um detector (BEĆ, et al., 2021).

A lâmpada de tungstênio-halogênio é uma fonte de radiação bem conhecida e amplamente utilizada em espectrofotômetros NIR de bancada. Trata-se de uma fonte de



radiação térmica, na qual um filamento sofre aquecimento resistivo por uma corrente elétrica que o atravessa. Esse processo gera luz de alta intensidade, com um perfil de emissão espectral que varia conforme a temperatura do filamento e da parede interna da lâmpada.

O seletor de comprimento de onda determina se será necessário usar um detector de pixel único, ou seja, um único elemento de resolução espectral, ou um detector de matriz complexa no espectrômetro. Ao contrário dos espectrômetros NIR de bancada, que são predominantemente instrumentos de transformada de Fourier (FT) equipados com um interferômetro de Michelson, os dispositivos portáteis apresentam uma grande diversidade nesse aspecto (BEĆ, et al., 2021). Os instrumentos NIR portáteis apresentam uma variação expressiva de seletores de comprimento de onda no mercado. Podem ser encontrados dispositivos que utilizam interferômetro de Fabry–Pérot, multiplexação Hadamard, filtro variável linear (LVF) ou matriz de microespelhos digitais.

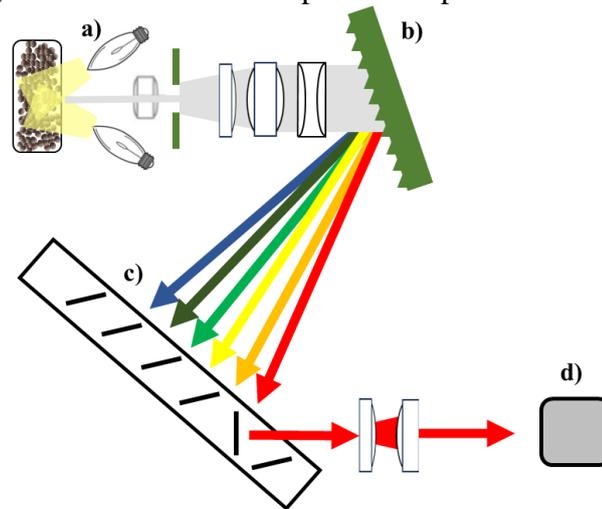
A multiplexação Hadamard foca o feixe de luz em uma fenda e, após passar por uma rede óptica associada, é codificado pela máscara de Hadamard antes de ser direcionado a um detector de pixel único. Ao chegar ao detector, o espectro é recuperado por meio da transformação de Hadamard. Esses equipamentos não possuem partes móveis, e o seletor de comprimento de onda com transformada de Hadamard é amplamente utilizado em espectrofotômetros NIR portáteis (BEĆ, et al., 2021, PASQUINI, 2003).

Dois tipos de detectores são comumente utilizados em espectrômetros NIR portáteis. Os diodos de silício fotovoltaicos, que abrange a faixa de 700 – 1.100 nm, que necessitam de filtros de seleção de comprimento de onda para eliminar o risco de o detector responder à luz solar. Já os fotodetectores de InGaAs, abrangendo a faixa de 900 a 1.700 nm, quando comparados a outros detectores, apresentam um tempo de resposta rápido, boa eficiência quântica e baixo ruído escuro, permitindo tempos de varredura curtos com boa relação sinal-ruído (S/N) (BEĆ, et al., 2021).

A instrumentação básica para o equipamento NIR-S-G1 da InnoSpectra, utilizado neste estudo, é apresentada na Figura II-2.



Figura II-2 - Mecanismo óptico do espectrômetro NIR



Fonte: Autor, 2024.

No mecanismo óptico do espectrômetro NIR-S-G1 da InnoSpectra, a radiação interage com a amostra e passa por uma fenda de entrada (Figura II-2a), um conjunto de lentes de colimação e uma grade de difração (Figura II-2b), um Dispositivo de Microespelho Digital (DMD) (Figura II-2c), um conjunto de lentes receptoras de luz e um detector InGaAs de elemento único (Figura II-2d).

## 2.4. IMAGENS DIGITAIS

Uma imagem pode ser definida, matematicamente, como uma função bidimensional  $f(x, y)$ , determinada numa certa região, em que  $x$  e  $y$  são coordenadas espaciais. O valor de  $f$  em qualquer par de coordenadas  $(x, y)$  é chamada de intensidade da imagem, que é uma quantidade escalar positiva, cujo significado físico é determinado pela origem da imagem (GONZALEZ, WOODS, 2010). A representação digital de imagens é realizada pelo armazenamento dos valores de intensidade de cor em cada par de coordenadas  $(x, y)$  (SERGYÁN, 2007). Tal imagem é composta de um número finito de elementos, cada um com localização e valor específicos. Cada ponto na grade bidimensional da imagem digital é denominado elemento de imagem ou pixel, que é o termo mais utilizado para representar os componentes de uma imagem digital, quanto maior for o número de pixels por dada área, maior será resolução da imagem, (GONZALEZ, WOODS, 2010; ACHARYA, 2005).

Na matriz que representa uma imagem, a sua dimensão é proporcional a resolução e aumentar esse parâmetro é desejável para definição de detalhes, melhor reprodução de cores,



brilho e contraste (MACHADO; SOUKI, 2004). Dessa forma, aumentando a resolução mais detalhes em um histograma obtido desta matriz, pode-se aumentar o poder de discriminação.

A resolução de uma imagem, quando capturada digitalmente, é dada por PPI (*pixels per inch*). Quando impressa, torna-se definida por DPI (*dots per inch*), que é dada pelo número de pontos por polegada. Quanto maior o número de pixels por polegada, mais informações haverá. Uma imagem que apresenta uma resolução de 300 PPI, é composta por 300 pixels de largura e 300 pixels de altura, comportando um total de 90 000 pixels (300x300 ppi). Sendo 300 ppi uma resolução utilizada, justamente por conter um grande número de informação que quando impressa são suficientes para os olhos humanos perceberem uma boa qualidade. Diante disto, quanto melhor a resolução, mais informação o arquivo vai conter e, conseqüentemente, maior será o espaço de memória para o armazenamento dos dados (MACHADO; SOUKI, 2004).

A intensidade luminosa em cada ponto  $(x, y)$  é o produto da iluminação  $i(x, y)$  com a refletância  $r(x, y)$  (Equação II.1), associada à quantidade de luz incidente e refletida sobre o ponto  $(x, y)$ , respectivamente (GONZALEZ, WOODS, 2010).

$$f(x, y) = i(x, y)r(x, y) \quad (\text{Equação II-1})$$

Para realizar uma metodologia utilizando imagens digitais, primeiro passo é escolher o equipamento que será utilizado para a aquisição das imagens, podendo ser uma câmera (ARAÚJO, et al., 2018), webcam (França, 2020), scanner (PEREIRA, 2022) e até mesmo o celular (RIOS; BREITKREITZ, 2021) e garantir a menor interferência da luz externa para o registro das imagens. Por fim, as imagens são processadas em um dispositivo processador para a construção do modelo, seja ela de classificação ou calibração.

#### 2.4.1. Sistema de cores

A representação digital das imagens é realizada pelo armazenamento dos valores de intensidade de cor de cada pixel. O sistema de cores tem o objetivo de facilitar e padronizar especificações de cores. A maioria dos modelos de cores utilizados atualmente é direcionada a hardware, como monitores e impressoras, ou a aplicações envolvendo a manipulação de cores para a criação de imagens coloridas em uma animação (SERGYÁN, 2007; GONZALEZ, WOODS, 2010).

As informações são extraídas de uma imagem por meio de métodos de processamentos, que transcrevem para dados, possibilitando o tratamento e construções de modelos para diversas



aplicações. Métodos para analisar imagens digitais envolvem a decomposição de tal imagem em componentes separados (DEINEKO, 2021). Esta decomposição é definida pelo conjunto de escala de cores em que a imagem é representada. Dentre os canais de cores, os mais conhecidos são: RGB, CMY/CMYK e HSI (GONZALEZ, WOODS, 2010).

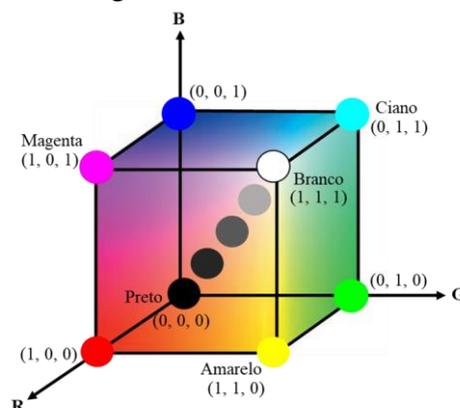
#### 2.4.1.1. Sistema RGB

O sistema RGB (*Red, Green e Blue*) é formado pela combinação dos componentes espectrais primários que inclui o vermelho, verde e azul. A escala RGB se baseia no sistema de coordenadas cartesianas que podem ser representadas por um cubo.

No cubo do RGB, como apresentado na Figura II-3, as cores primárias estão em três vértices principais, as cores secundárias ciano (0, 1, 1), magenta (1, 0, 1) e amarelo (1, 1, 0) estão em outros três vértices que são formados por dois eixos das cores primárias, o preto está na origem (0, 0, 0) e o branco (1, 1, 1) está no vértice mais distante da origem. As diferentes cores que compõem este modelo são pontos presentes no cubo.

A Escala de cinza estende-se no cubo RGB do preto até o branco, ao longo do segmento de reta que une esses dois pontos, com valores iguais para os canais R, G e B. Em termos dos componentes de uma imagem, o 0 representa o preto e 255 representa o branco (GONZALEZ, WOODS, 2010).

Figura II-3 - Cubo RGB



Fonte: Autor, 2024.

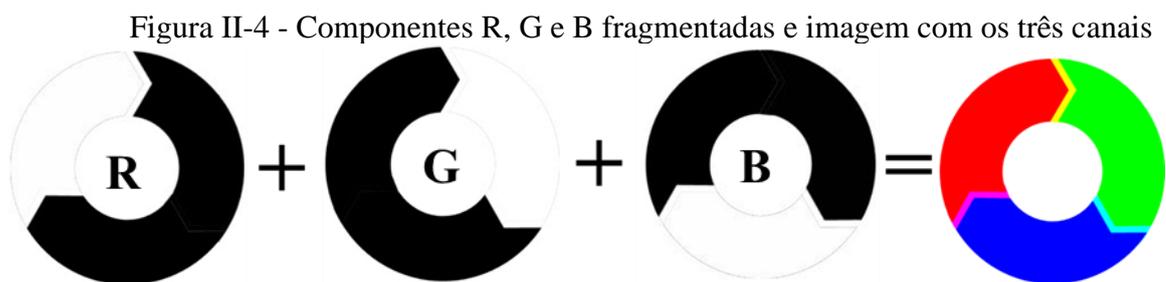
Imagens representadas por esse sistema de cores consistem em três componentes de imagens, uma para cada cor primária. Quando alimentadas em um monitor RGB, essas três imagens se combinam na tela para produzir uma imagem de cores compostas (Figura II-4).



Em uma imagem digital no sistema RGB, um pixel pode ser visto como um vetor cujas componentes representam as intensidades de vermelho, verde e azul de sua cor. A imagem colorida pode ser vista como a composição de três imagens monocromáticas e o somatório (Equação II.2) desses três componentes resulta na figura RGB (DE QUEIROZ; GOMES, 2006).

$$f(x, y) = f_R(x, y) + f_G(x, y) + f_B(x, y) \quad (\text{Equação II-2})$$

É possível observar na Figura II-4 os planos monocromáticos de uma imagem e a resultante do somatório das três componentes, como descrito na Equação II.2. No R (Figura II-4), temos a imagem decomposta na componente vermelha, sendo possível notar que a parte com a cor vermelha passa a ser branca, ou seja, presença total da respectiva cor. O mesmo ocorre para as demais componentes, G (verde) e B (azul) (Figura II-4).



Fonte: Autor, 2024.

#### 2.4.1.2. Sistema CMY e CMYK

A Escala CMY (*Cyan, Magenta, Yellow*) é composta pelas cores primárias de pigmentos, ou seja, são as cores secundárias do modelo RGB. Ao incidir uma luz branca sobre uma superfície magenta, nenhuma cor verde é refletida, isso ocorre porque a cor magenta é a ausência da cor verde da luz branca, que é composta pelo RGB. Equipamentos como impressoras e copiadoras utilizam a escala CMYK, que nada mais do que a CMY com a adição da cor preta (GONZALEZ, WOODS, 2010).

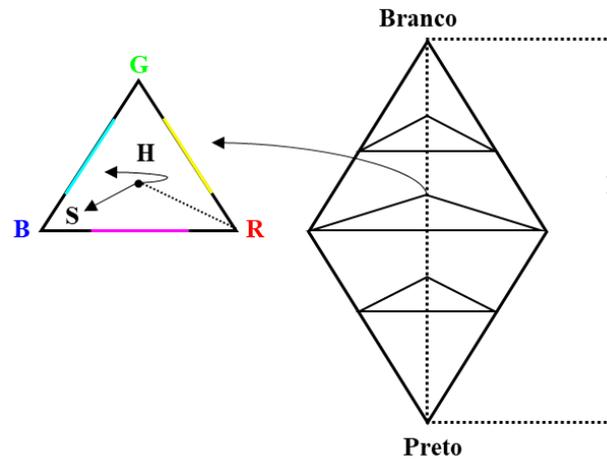
#### 2.4.1.3. Sistema HSI

O modelo HSI (matiz, saturação e intensidade) é empregado como uma ferramenta para o desenvolvimento de algoritmos de processamento de imagens, utilizando descrições de cores conforme a percepção humana. Isso ocorre porque, ao ser humano observar um objeto, ele é descrito em termos de matiz, saturação e intensidade e não pela quantidade de cada cor presente



na imagem (GONZALEZ, WOODS, 2010). Geometricamente, o modelo HSI pode ser visto como um sólido (Figura II-5), cujos cortes horizontais produzem triângulo equiláteros.

Figura II-5 - Sólido HSI



Fonte: Autor, 2024.

Na Figura II-5, podemos encontrar o sólido em forma de triângulo equilátero que representa o sistema **HSI**. O **H** forma um ângulo que inicia no vermelho e pode ser dividido em três setores: setor RG ( $0^\circ < H \leq 120^\circ$ ), setor GB ( $120^\circ < H \leq 240^\circ$ ) e setor BR ( $240^\circ < H \leq 360^\circ$ ). O **S** varia do centro do triângulo até a extremidade; quanto mais próximo do centro, maior a presença de branco, e quanto mais afastado, menor a presença. O **I** vai do preto (ausência de cor) ao branco (presença de cor), indicando a intensidade da cor.

A Matiz (**H**) está relacionada ao comprimento de onda predominante em uma mistura de ondas de luz, sendo um atributo que descreve uma cor pura (amarelo, laranja ou vermelho puros). Ela representa a cor que um observador percebe como dominante. Assim, quando afirmamos que um objeto é vermelho, laranja ou amarelo, estamos nos referindo ao seu matiz. Ao passo que a Saturação (**S**) mede a pureza relativa à cor ou o quanto que essa cor é diluída por uma quantidade de luz branca. Cores totalmente puras tem um alto valor de saturação, já cores misturadas, como rosa, por exemplo, que é formada por vermelho e branco, são menos saturadas. A Intensidade (**I**) é o brilho ou nível de energia da luz e não possui conteúdo de cor, descrevendo a sensação das cores (GONZALEZ, WOODS, 2010; MARQUES FILHO; VIEIRA NETO, 1999).

Os dados obtidos das imagens digitais, assim como os espectros NIR, requerem auxílio para sua interpretação. Esses dados contêm uma grande quantidade de informações, e é necessário utilizar ferramentas estatísticas e matemáticas para eliminar variáveis que não são



informativas. Nesse contexto, a Quimiometria é aplicada para fornecer métodos que aprimoram a interpretação dos dados.

## 2.5. QUIMIOMETRIA

A Quimiometria é a aplicação de métodos estatísticos e matemáticos para a resolução de problemas relacionados a dados químicos, se encontra em um ponto de intersecção entre estas três grandes áreas (FERREIRA, 2015). Desta forma, a quimiometria é utilizada para planejar e otimizar experimentos, extrair o maior número de informações relevantes dos dados químicos e obter conhecimento sobre estes sistemas (MASSART; et al., 1998).

Na análise de dados multivariados, a Quimiometria é uma ferramenta essencial, pois permite identificar variações e tendências na matriz de dados estudadas. Assim, as ferramentas quimiométricas tentam encontrar as relações entre as amostras e variáveis em um determinado conjunto de dados (KUMAR et al., 2014).

O potencial da quimiometria é vasto, abrangendo áreas como química e física, como estudos em cinética e equilíbrio, otimização de reações e química teórica. Ela também cobre a maioria das áreas de espectroscopia e tem aplicações diversas como monitoramento ambiental, arqueologia científica, biologia, ciência forense, monitoramento de processos industriais, geoquímica etc. (BRERETON, 2003; KUMAR et al., 2014).

### 2.5.1. Pré-processamento

O objetivo por trás da aplicação de técnicas de pré-processamento é minimizar variações indesejadas que não foram eliminadas durante a aquisição da informação dos sinais que possam prejudicar a interpretação ou capacidade preditiva do modelo resultante, sem excluir ou alterar variações quimicamente relevantes, que podem distorcer os dados e com isso, induzir os modelos ao erro (FERREIRA, 2015).

Pré-tratamentos aplicados com frequência aos dados são: centralização na média, escalonamento e autoescalonamento. Centrar os dados na média tem o objetivo de subtrair os elementos de cada linha ou coluna pela sua média. Escalonar um determinado conjunto de dados consiste em dividir os elementos da coluna pelo desvio padrão da mesma coluna. O autoescalonamento consiste em centrar os dados na média e, em seguida, efetuar o



escalonamento. O escalonamento e o autoescalonamento são utilizados quando é desejável atribuir os mesmos pesos às variáveis empregadas (PONTES, 2009).

Nos dados de origem espectrais os métodos de pré-processamento comumente utilizados incluem diferenciação e correção de dispersão, por exemplo, derivadas de primeira (1ª Derivada. S.G) e segunda (2ª Derivada S.G) ordem com filtro Savitzky–Golay, Correção de Sinal Multiplicativo (MSC) e transformação Padrão Normal da Variável (SNV) (STEVEN; TAULER R; WALCZAK, 2010).

### 2.5.2. Técnicas de Reconhecimento de Padrões

As técnicas de reconhecimento de padrões têm o objetivo de identificar e determinar se existe alguma relação entre as amostras que passam pelo mesmo tipo de análise. Tais técnicas podem ser classificadas em dois grupos: não supervisionadas ou supervisionadas (BARROS NETO, SCARMINIO; BRUNS, 2006).

A técnica reconhecimento de padrões supervisionada ocorre com a construção de modelos de um conjunto de amostras conhecidas, denominadas conjunto de treinamento. Assim, o modelo construído é utilizado posteriormente para identificar amostras de classes desconhecidas. Na literatura, as técnicas supervisionadas estão bem concretizadas e com um elevado número de aplicações, entre as principais se encontram a análise discriminante por mínimos quadrados parciais (PLS-DA) e análise discriminante linear (LDA) (VILAR, 2018). Caso não esteja disponível informação prévia sobre a classificação da amostra, procede-se ao reconhecimento de padrões de forma não supervisionada, buscando-se identificar padrões de agrupamentos. A técnica não supervisionada mais conhecida e utilizada é a análise de componentes principais (PCA).

#### 2.5.2.1. Análise por componentes principais (PCA)

A Análise de Componentes Principais é amplamente reconhecida como uma das principais ferramentas para a análise exploratória de dados. Ela é frequentemente aplicada na etapa inicial da análise para identificar padrões nos dados coletados. A PCA é um método que projeta dados multivariados em um espaço de menor dimensão, reduzindo a dimensionalidade do conjunto de dados original sem alterar as relações entre as amostras. Essa técnica visa modelar a variância total do conjunto de dados original utilizando componentes principais que são não correlacionados entre si (DUNTEMAN, 1989). É possível, por meio da PCA,



identificar, visualizar e interpretar diferenças entre as variáveis, além de examinar as possíveis relações entre as amostras (FERREIRA, 2015).

Em termos matemáticos, o procedimento de análises de componentes principais é descrito na Equação II.3. Em que, a matriz que contém todos os *scores* das componentes é apresentada como T, a matriz que contém os *loadings* é chamada de P, e a matriz que representa a variância não explicada pelo modelo é conhecida como matriz de resíduos E (SOARES, 2010).

$$X = TP^t + E \quad (\text{Equação II-3})$$

#### 2.5.2.2. Análise discriminante pelos mínimos quadrados parciais (PLS-DA)

O PLS-DA é projetado para classificar observações em diferentes grupos ou classes, maximizando a variância entre as classes enquanto minimiza a variância dentro das classes.

O método combina características do modelo de regressão PLS (*Partial Least Squares*) com a análise discriminante, que relacionam as informações instrumentais com as propriedades de interesse que agora é obtida, utilizando as informações das classes armazenadas em um vetor. A modelagem preditiva PLS-DA envolve essencialmente dois procedimentos principais: (a) a criação de Variáveis Latentes (VL) (ou seja, a redução da dimensionalidade); e (b) o desenvolvimento de um modelo de previsão (ou seja, a análise discriminante) (LEE, et al., 2018).

Ele projeta os dados em um espaço de menor dimensão, no qual as variáveis predictoras são combinadas para formar componentes latentes, que capturam a maior parte da variação nos dados. Esses componentes são usados para prever a classe das observações (LEE, et al., 2018; FERREIRA, 2015).

Após a construção do modelo, novas observações podem ser classificadas com base em sua proximidade aos grupos definidos pelos componentes latentes. O modelo atribui uma classe a cada nova observação com base nas características que ela apresenta (LEE, et al., 2018).

#### 2.5.2.3. Análise discriminante linear (LDA)

A análise discriminante linear (LDA) é um método de análise de dados, originalmente proposto por R. Fisher para diferenciar diferentes tipos de flores (XANTHOPOULOS et al., 2013). Este método se baseia na criação de funções discriminantes (FDs) que maximizam a



distância entre as classes e minimizam a variância das amostras dentro de uma mesma classe. Para alcançar isso, as FDs são obtidas a partir de combinações lineares das variáveis originais, que são mais eficazes na discriminação entre as classes ou grupos de objetos (PONTES et al., 2005).

Ao determinar os centros das classes dos objetos a serem classificados, obtém-se uma fronteira denominada função discriminante linear, de distância igual entre os pontos e perpendicular à linha que une os centros. A função discriminante determina as regiões correspondentes de cada classe. Deste modo, se um objeto estiver localizado em uma região específica, ele será classificado como pertencente à classe associada a essa região; caso contrário, será classificado na outra classe. No modelo LDA, as amostras são sempre classificadas como pertencentes a uma ou outra classe (VILAR, 2018).

A análise discriminante linear apresenta limitação quando o número de variáveis é maior que o número de amostras, tornando a eficiência do modelo comprometida, levando a problemas de colinearidade. Em casos que apresentem um elevado número de variáveis, como dados espectroscópicos e imagens, é recomendado o uso de algoritmos de seleção de variáveis (FERREIRA, 2015).

### 2.5.3. Seleção de variáveis

#### 2.5.3.1. Algoritmo genético (GA)

O algoritmo genético é fundamentado nos princípios da seleção natural das espécies de Charles Darwin e da genética. No contexto de otimização esse algoritmo codifica uma solução potencial para um problema específico em uma estrutura de dados binarizados aplicam operadores de recombinação a essas estruturas para otimizar tais informações (MATHEW, 2012).

O algoritmo inicia com um conjunto de cromossomos formando uma população, no qual cada indivíduo é representado por um cromossomo. O tamanho da população pode ser mantido constante ao longo das gerações. A cada nova geração, cada cromossomo é avaliado por meio de um processo baseado em uma aptidão, pelo qual os indivíduos mais aptos (conforme medido por uma função custo) são tipicamente mais propensos a serem selecionados para reprodução e assim contribuirão para formação da próxima geração, contudo, indivíduos menos aptos também devem ter chances de serem selecionados (SOARES, 2010; MATHEW, 2012).



A população seguinte pode ser gerada por ações de competição, cruzamento e mutação. Na competição, cromossomos menos aptos são excluídos para dar lugar a novos cromossomos. Usando os métodos de cruzamento e mutação um cromossomo “filho” é criado, que normalmente compartilha muitas das características de seus “pais”. Novos pais são selecionados para cada novo filho, e o processo continua até que uma nova população de soluções de tamanho apropriado seja gerada (MATHEW, 2012). Já a mutação ocorre aleatoriamente com uma pequena percentagem, trocando de 0 e 1 um dos genes do cromossomo.

Em seleção de variáveis os genes são codificados de forma que o valor 1 indica que a variável será incluída no modelo, enquanto genes com valor 0 sinalizam que a variável não será avaliada (SOARES, 2010).

As etapas de cruzamento, mutação e avaliação são repetidas até que a satisfação do critério de parada seja atingida. Em muitos casos, tal critério pode ser um número máximo de gerações.

#### 2.5.3.2. Algoritmo das Projeções Sucessivas (SPA)

O SPA tem como finalidade buscar subconjuntos representativos de variáveis com menor multicolinearidade. Inicialmente, o parâmetro utilizado para a escolha do melhor subconjunto de variáveis é através do modelo MLR que produziu o menor erro RMSEV ou RMSECV (ARAÚJO et al., 2001). Quando é utilizado para fins de modelagem de classificação, outra função de custo precisa ser empregada para orientar a seleção das variáveis no SPA. Neste caso, tem sido utilizado como parâmetro a função custo  $G$  (Equação II.4), relacionada ao risco médio de classificação incorreta no conjunto de validação (PONTES et al., 2005).

$$G = \sum_{k=1}^N \frac{r^2(\mathbf{x}_k, \mu_{ik})}{\min_{ij \neq ik} r^2(\mathbf{x}_k, \mu_{ij})} \quad (\text{Equação II-4})$$

Na Equação II-4, o numerador  $r^2(\mathbf{x}_k, \mu_{ik})$  representa o quadrado da distância de Mahalanobis entre o objeto  $\mathbf{x}_k$  e a média de sua classe  $\mu_{ik}$ . Já o denominador,  $\min_{ij \neq ik} r^2(\mathbf{x}_k, \mu_{ij})$ , corresponde ao quadrado da distância de Mahalanobis entre o  $\mathbf{x}_k$  e o centro da classe incorreta mais próxima.



## CAPÍTULO III : METODOLOGIA

---

### 3.1. AQUISIÇÃO E PREPARO DAS AMOSTRAS

Neste trabalho foram adquiridas 40 amostras comerciais de paprica doce de diferentes fabricantes e lotes. Todas as marcas que foram coletadas para este trabalho, tem registro ativo e sao grandes fabricantes da industria de alimentos. Com o objetivo de evitar amostras de feiras e a granel na etapa de calibracao e validacao, cuja exposicao a poeira e contaminantes  direta e possveis adulteracao por parte dos produtores. Tais amostras foram adquiridas no estado da Paraba no perodo de agosto de 2022 a setembro de 2023. Apos a coleta as amostras foram armazenadas em dessecador, contendo slica e envolto por papel alumnio, para evitar a exposicao  umidade e a luz, e conseqente degradacao do material. O amido de milho e o urucum tambm foram adquiridos em estabelecimentos comerciais local.

Foram estabelecidos subconjuntos de amostras com 4 nveis de adulteracao para cada adulterante, sendo mantida a massa final padro em 3 g, como sendo a massa previamente avaliada como suficiente para o preenchimento de uma placa de Petri (40 x 10 mm) em vidro utilizada como suporte da amostra durante as aquisicoes dos espectros. As amostras foram acondicionadas em sacos *zip lock* e agitadas manualmente para promover a homogeneizacao das amostras. Os nveis dos adulterantes estabelecidos foram: 1%, 5%, 10% e 20%, todos em porcentagem massa/massa.

### 3.2. REGISTRO DOS ESPECTROS NIR

Um espectrmetro porttil (NIR-S-G1, InnoSpectra), com duas lmpadas de filamento de tungstnio de 0,7 W como fontes de radiacao e um detector InGaAs de 1 mm, foi utilizado para obter os espectros NIR operando na faixa de 900 a 1700 nm, com preciso do comprimento de onda de  $\pm 2$  nm. Cada amostra foi colocada na placa de Petri por vez e posicionada para aquisicao dos espectros. Na aquisicao, um feixe de luz foi emitido pelo instrumento, que interagiu com a amostra e o sinal foi registrado por meio de refletncia difusa (Figura III-1). Como espectro de referncia, foi utilizado teflon em p compactado. Para cada amostra, trs medioes foram realizadas utilizando o software ISC-NIRScan-GUI atravs de um computador.



Figura III-1 – Ilustração da aquisição dos espectros NIR



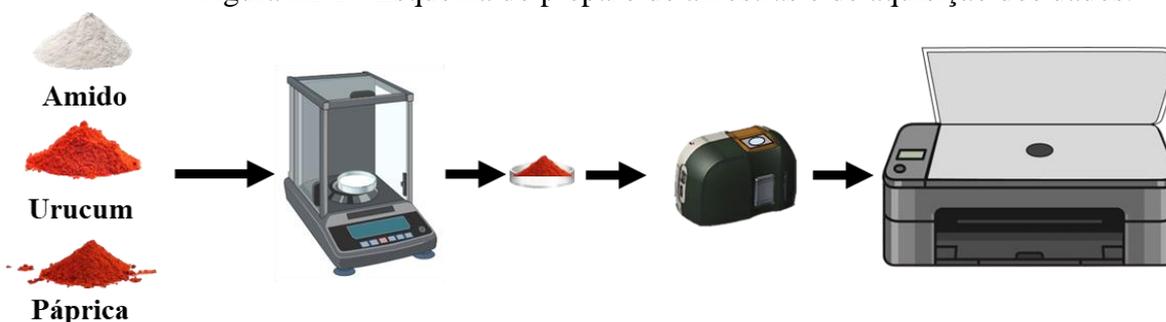
Fonte: Autor, 2024.

### 3.3. REGISTRO DAS IMAGENS DIGITAIS

As imagens das amostras acondicionadas em uma placa de Petri (40 x 10 mm) foram registradas utilizando um scanner da Impressora multifuncional da marca Epson, modelo L455. Com o objetivo de limitar a interferência da luz, foi elaborado para impressora um suporte de plástico, como pode ser observado na Figura III-2, com tampa nas mesmas dimensões da área de aquisição da imagem, deixando apenas o espaço para inserir a placa Petri disponível para o scanner.

Com o objetivo de tornar a análise mais eficiente, após as quantidades de páprica e do adulterante (amido e urucum) serem pesadas, as amostras tiveram os espectros registrados utilizando o nanoNIR, e em seguida, as imagens no scanner. Um esquema do aparato utilizado pode ser visto na Figura III-2.

Figura III-2 - Esquema do preparo de amostras e de aquisição dos dados.



Fonte: Autor, 2024.



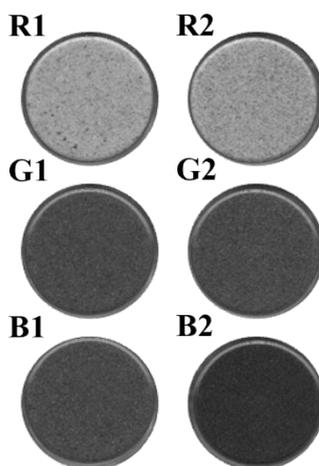
### 3.4. SOFTWARE E PROCEDIMENTO QUIMIOMÉTRICO

Foram realizados pré-processamentos nos espectros dos dados NIR com objetivo de minimizar as variações da linha de base, os efeitos do espalhamento e do ruído associado. Foram avaliados vários pré-processamentos, sendo apresentados apenas os dados que forneceram modelos com melhores resultados, como a correção de linha de base por *offset*, primeira derivada com filtro Savitzky–Golay e com janelas de 7 e 11 e, segunda derivada com Savitzky–Golay e janelas de 11 pontos.

Os histogramas das imagens foram extraídos usando uma interface gráfica desenvolvida em Matlab<sup>®</sup>. O formato e o tamanho da região de interesse, para obtenção dos dados, foram definidos de modo a evitar as bordas da placa Petri e maximizar a quantidade de pixels utilizados em cada amostra. Tais procedimentos foram padronizados na aquisição dos histogramas para todas as imagens.

Na Figura III-3, para exemplificar a decomposição das componentes RGB, são apresentadas a componente vermelha (R), verde (G) e azul (B), decompostas a partir de uma amostra com 1% de adulteração com amido (1) e uma amostra com 20% de adulteração com amido (2).

Figura III-3 - Amostras decompostas nas componentes vermelho (R), verde (G) e azul (B)



Fonte: Autor, 2024.

Após a extração dos histogramas de cores, foram construídos modelos quimiométricos de classificação PLS-DA, GA-LDA e SPA-LDA, utilizando o histograma completo, os canais individuais, bem como combinações binárias entre os sistemas RGB, HSI e escala de cinza. Com a finalidade de garantir a melhor representatividade dos modelos de classificação, as amostras



foram divididas em dois subconjuntos, treinamento e teste, de acordo com algoritmo KS (KENNARD; STONE, 1969), sendo 75% do conjunto total para treinamento e 25% para teste (Tabela III:1). Este critério de divisão foi adotado para os conjuntos de dados registrados no NIR e para as imagens digitais.

Tabela III:1 - Número de amostras de treinamento e teste usadas na construção dos modelos

| Amostra         | Número de amostras | Conjuntos   |       |
|-----------------|--------------------|-------------|-------|
|                 |                    | Treinamento | Teste |
| Não Adulteradas | 40                 | 30          | 10    |
| Adulteradas     | 160                | 120         | 40    |

Fonte: Autor, 2024.

A classificação dos subconjuntos de treinamento e teste foram avaliadas a partir dos parâmetros de desempenho: Taxa de Classificação Correta (TCC) (Equação III.1), Precisão (Equação III.2), Sensibilidade (Equação III.3) e Especificidade (Equação III.4). As equações utilizadas para os cálculos são apresentadas a seguir.

$$\text{Taxa de Classificação Correta} = \frac{VP + VN}{VP + FP + VN + FN} \times 100 \quad (\text{Equação III-1})$$

$$\text{Precisão} = \frac{VP}{VP + FP} \times 100 \quad (\text{Equação III-2})$$

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \times 100 \quad (\text{Equação III-3})$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \times 100 \quad (\text{Equação III-4})$$

Em que:

TP (Verdadeiro Positivo) = Número de amostras não adulteradas classificadas corretamente;



TN (Verdadeiro Negativo) = Número de amostras adulteradas classificadas corretamente;

FP (Falso Positivo) = Número de amostras adulteradas classificadas como não adulteradas. FP também é conhecido como erro do Tipo I;

FN (Falso Negativo) = Número de amostras não adulteradas classificadas como adulteradas. FN também é conhecido como erro do Tipo II.

Para este trabalho, as amostras de páprica doce não adulteradas foram tratadas como o caso positivo e as amostras adulteradas como negativo na matriz confusão. Ainda nessa matriz, as amostras não adulteradas são apresentadas como PD, amostras adulteradas com amido como PA e as amostras adulteradas com urucum PU.

Para escolher o número de variáveis latentes nos modelos PLS-DA foram empregadas a menor taxa de erro obtida no conjunto de treinamento usando o processo de validação cruzada.



## **CAPÍTULO IV : RESULTADOS E DISCUSSÃO**

---

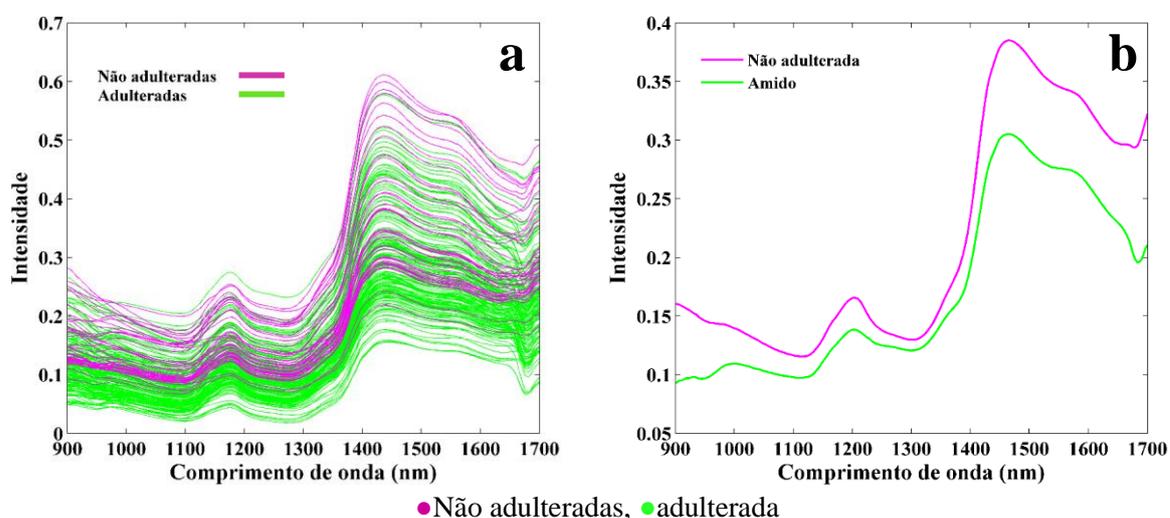


## 4.1.AVALIAÇÃO DOS ESPECTROS NIR

### 4.1.1. Adulteração da Páprica Doce com Amido

Os espectros NIR de 40 amostras de páprica doce não adulteradas e de 160 amostras adulteradas com amido são apresentados na Figura IV-1a e o espectro médio da páprica doce e do amido puro Figura IV-1b.

Figura IV-1 – a) Espectros NIR de páprica doce e adulterada com amido sem nenhum pré-processamento na faixa de 900 a 1700 nm e (b) espectro médio da páprica doce e do amido puro.



Ao analisar os espectros apresentado na Figura IV-1 é possível observar intensidades mais elevadas na região entre 1400 e 1600 nm, correspondentes à presença de bandas de combinação de C-H e ao primeiro sobretom de O-H (BURNS; CIURCZAK, 2007). Na faixa de 1125 a 1250 nm, também são detectadas bandas que podem estar associadas ao segundo sobretom de C-H e ao quarto sobretom de estiramento do grupo C=O. Essas bandas podem estar relacionadas à composição do pimentão vermelho, matéria-prima usada na produção da páprica, que contém água, proteínas, gorduras totais, carboidratos e fibras (LAGUNA et al., 2011; LIMA et al., 2011). Outros componentes presentes em menores quantidades, como a vitamina C (ácido ascórbico) e vitaminas do complexo B, também estão associados a absorções características no NIR.

Como pode-se perceber, todos os compostos relatados no parágrafo acima apresentam ligações químicas com absorção na região do infravermelho próximo (NIR). Além disso, percebe-se observando o espectro do adulterante Figura IV-1(b), a ocorrência de sobreposição

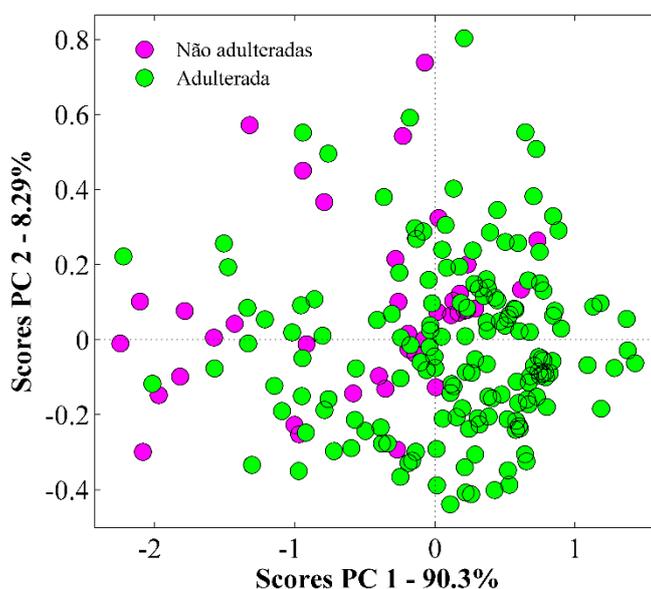


ao longo de toda faixa espectral estudada. Isso ocorre, pela existência de grupos funcionais na composição do amido que também absorvem nesta região.

#### 4.1.1.1. Análise das Componentes Principais (PCA) para os espectros NIR das amostras sem adulteração e adulterada com amido

Com o objetivo de realizar uma avaliação exploratória do comportamento dos dados, uma análise por componentes principais foi realizada nos espectros NIR para avaliar possíveis agrupamento e amostras anômalas (*outliers*) nos dados sem pré-processamentos. Na Figura IV-2 pode ser visualizado o gráfico dos scores de PC1 versus PC2.

Figura IV-2 - Gráfico de PC1 versus PC2 aplicada aos espectros sem pré-processamento NIR para Páprica doce e das amostras adulteradas com amido.

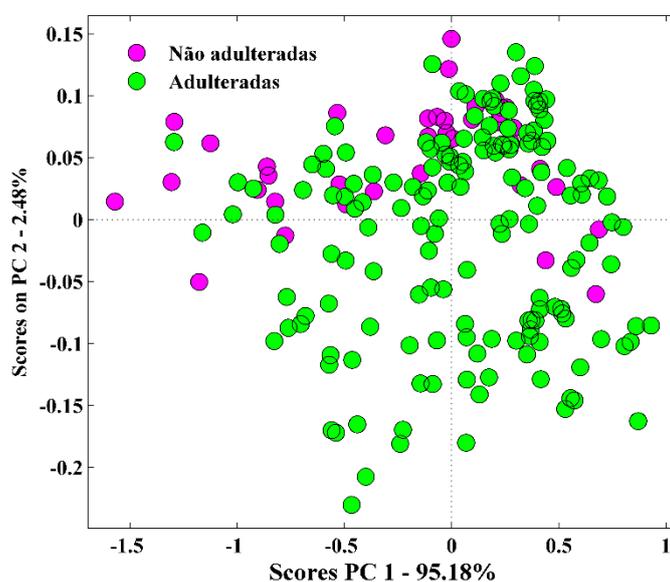


A PC1 e PC2 explicam 90,3% e 8,29% da variância dos dados, respectivamente. Muitas amostras adulteradas se encontram localizadas nos scores positivos para PC1 e negativo para PC2. A maioria das amostras de páprica doce não adulteradas estão do lado dos valores negativos de PC1, e as amostras adulteradas que se sobrepõem com as não adulteradas nesta região, são em sua maioria adulteradas com 1% de amido (Figura IV-2). Mesmo assim, é possível notar que há uma grande sobreposição entre as amostras, sendo necessário a utilização de ferramentas quimiométricas para auxiliar na separação das classes estudadas.

Foi realizada uma análise exploratória dos dados pré-processados com correção de linha de base *offset*, sendo apresentados os scores da PC1 versus PC2 na Figura IV-3.



Figura IV-3 - Gráfico de PC1 versus PC2 aplicada aos espectros NIR com correção de linha de base *offset* para Páprica adulterada com amido



Após a correção de linha de base *offset* (Figura IV-3), os scores da PC1 passaram a explicar 95.18% da variância dos dados e os da PC2 2.48%. As amostras não adulteradas passaram a formar um agrupamento nos scores positivos da PC1 (Figura IV-3) do quando comparadas aos da PC1 (Figura IV-2) dos dados sem pré-processamento, no qual as amostras estavam mais dispersas. Mesmo que seja possível notar o agrupamento, as duas classes ainda apresentam sobreposição, sendo necessária a aplicação de outras ferramentas quimiométricas para separação entre as classes.

#### 4.1.1.2. Métodos de Reconhecimento de Padrões Supervisionados aplicados as amostras não adulteradas e adulteradas com amido

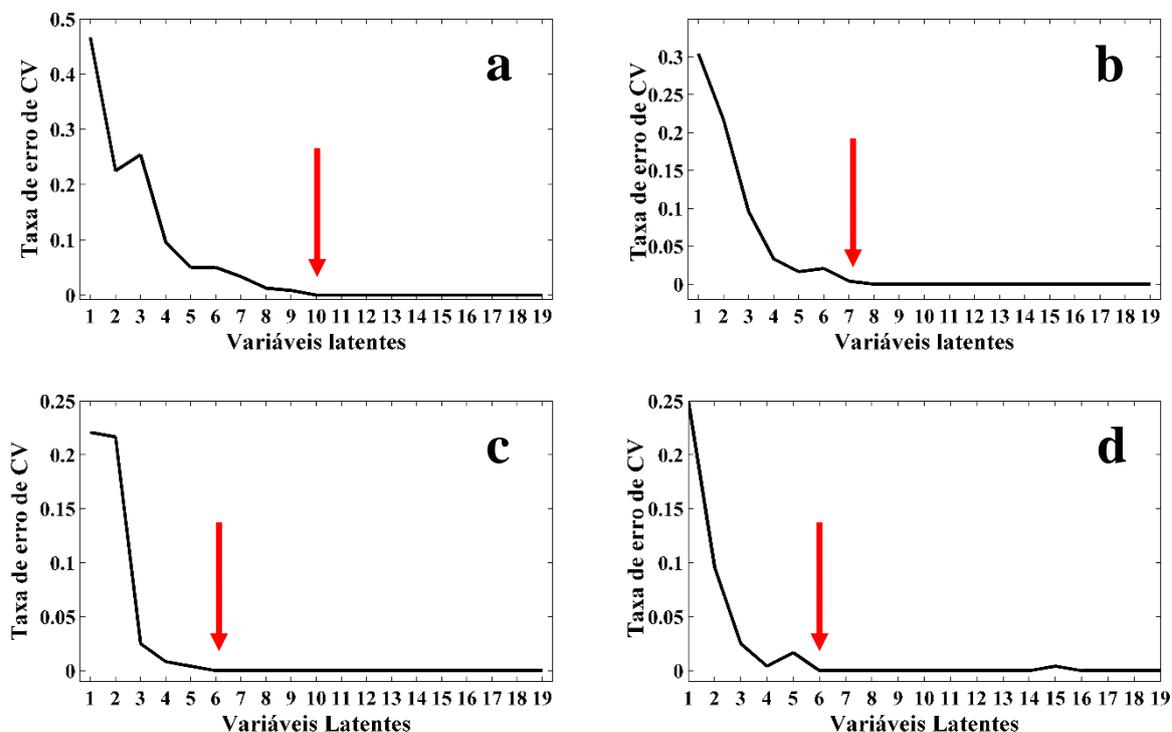
Os modelos PLS-DA foram obtidos utilizando os dados sem nenhum pré-processamento e com os tratamentos que foram apresentados na seção experimental. O número ótimo de variáveis latentes utilizado na construção de cada modelo PLS-DA foi obtido de acordo com o menor erro de classificação durante o processo de validação cruzada.

Na Figura IV-4a, a seta vermelha destaca as dez variáveis latentes selecionadas para a construção do modelo com os dados corrigidos pela linha de base *offset*. Nas Figuras IV4b, IV4c e IV4d, foram utilizadas, respectivamente, sete, seis e seis variáveis latentes. Esses dados foram processados com a primeira derivada e filtro de Savitzky-Golay (S.G) com janela de 11



pontos (Figura IV-4b), e a segunda derivada com janelas de 7 pontos (Figura IV-4c) e 11 pontos (Figura IV-4d).

Figura IV-4 - Gráfico da taxa de erro de validação cruzada *versus* o número de variáveis latentes incluídas no modelo para os modelos PLS-DA usando diferentes pré-processamentos.

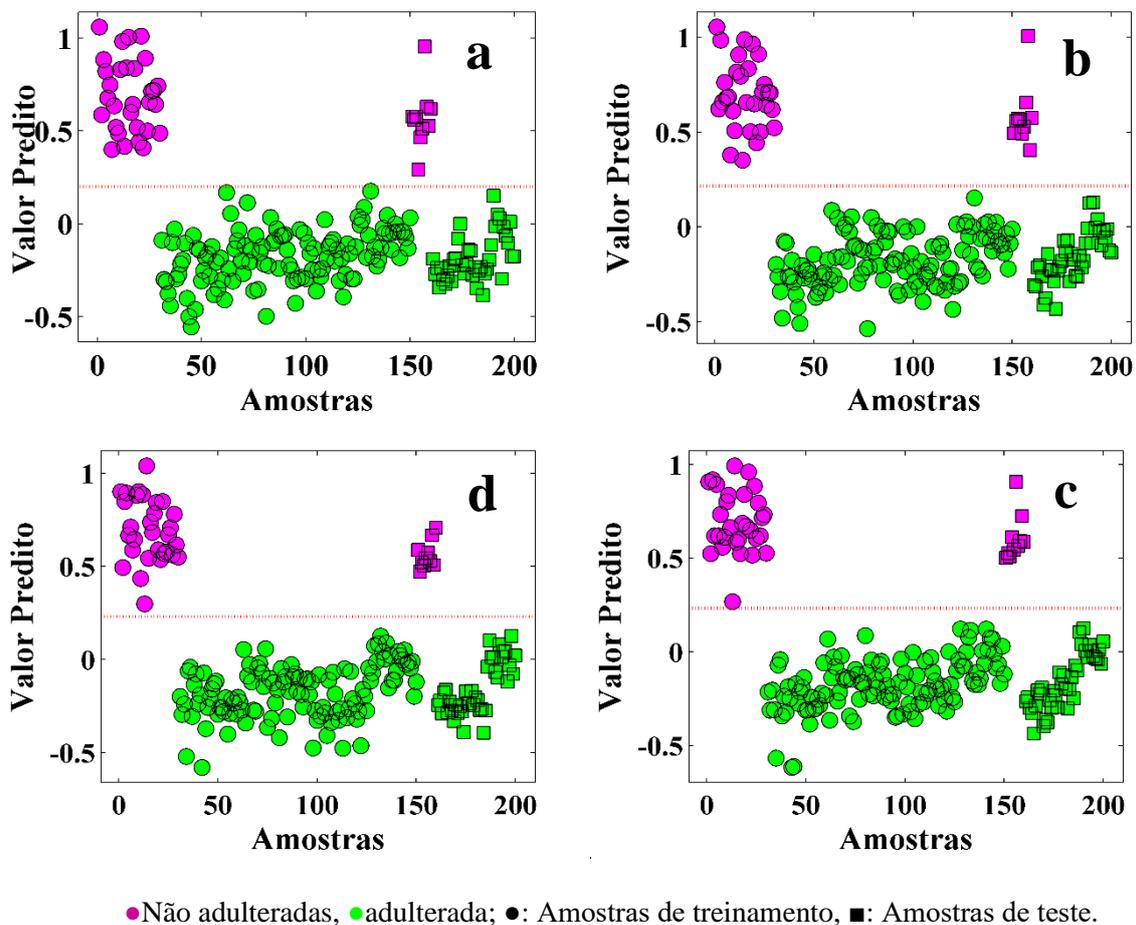


Os valores preditos das amostras dos conjuntos de treinamento e teste para os modelos PLS-DA desenvolvidos para a classificação das classes não adulteradas e adulteradas são apresentadas nas Figura IV-5. Embora tenha sido avaliados diferentes pré-processamentos, nesta figura são apresentados apenas os modelos construídos com o pré-processamentos que proporcionaram os resultados mais satisfatórios, considerando a TCC e menor número de variáveis latentes obtidos no conjunto de treinamento. Os modelos escolhidos para serem apresentados, foram obtidos com correção linha de base *offset* Figura IV-5a), 1ª Der. S.G 11\* Figura IV-5b), para 2º Der. S.G 7\* Figura IV-5c) para 2º Der. S.G 11\* Figura IV-5d).

Na Figura IV-5 as amostras não adulteradas estão acima do limiar de classificação e as adulteradas abaixo, evidenciando que nenhuma das amostras dos modelos apresentados, foram classificadas incorretamente, tanto para o conjunto de treinamento, quanto para o conjunto teste.



Figura IV-5 - Gráfico dos valores estimados pelo PLS-DA versus o índice de cada amostra não adulteradas e adulterada com amido utilizada nos conjuntos de treinamento e teste.



Na Tabela IV-1 são apresentados os resultados de classificação detalhados para os modelos PLS-DA utilizando o conjunto de teste. Entre parênteses estão o número de variáveis latentes escolhidos para cada modelo obtido.

Tabela IV-1 - Matriz de confusão obtida para os modelos PLS-DA para o conjunto de teste. Entre parênteses são apresentados o número de variáveis latentes utilizados.

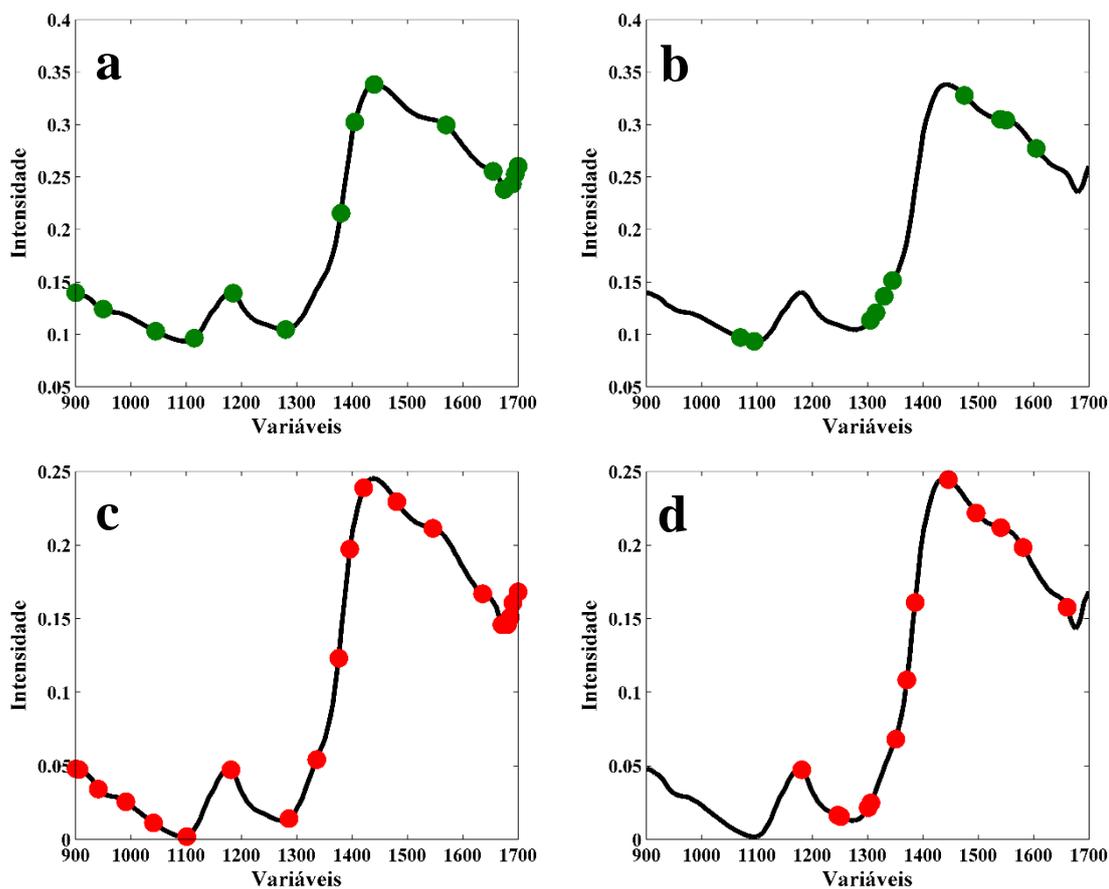
|                    |    | Correção linha de base <i>offset</i><br>(10) |    | 1º Der. S.G 7*<br>(7) |    | 2º Der. S.G 7*<br>(6) |    | 2º Der. S.G 11*<br>(6) |    |
|--------------------|----|--|----|-----------------------|----|-----------------------|----|------------------------|----|
|                    |    | PD   | PA | PD                    | PA | PD                    | PA | PD                     | PA |
| Matriz de confusão | PD | 10   | -  | 10                    | -  | 10                    | -  | 10                     | -  |
|                    | PA | -  | 40 | -                     | 40 | -                     | 40 | -                      | 40 |
| Precisão           |    | 1  |    | 1                     |    | 1                     |    | 1                      |    |
| Sensibilidade      |    | 1  |    | 1                     |    | 1                     |    | 1                      |    |
| Especificidade     |    | 1  |    | 1                     |    | 1                     |    | 1                      |    |
| TCC                |    | 100  |    | 100                   |    | 100                   |    | 100                    |    |

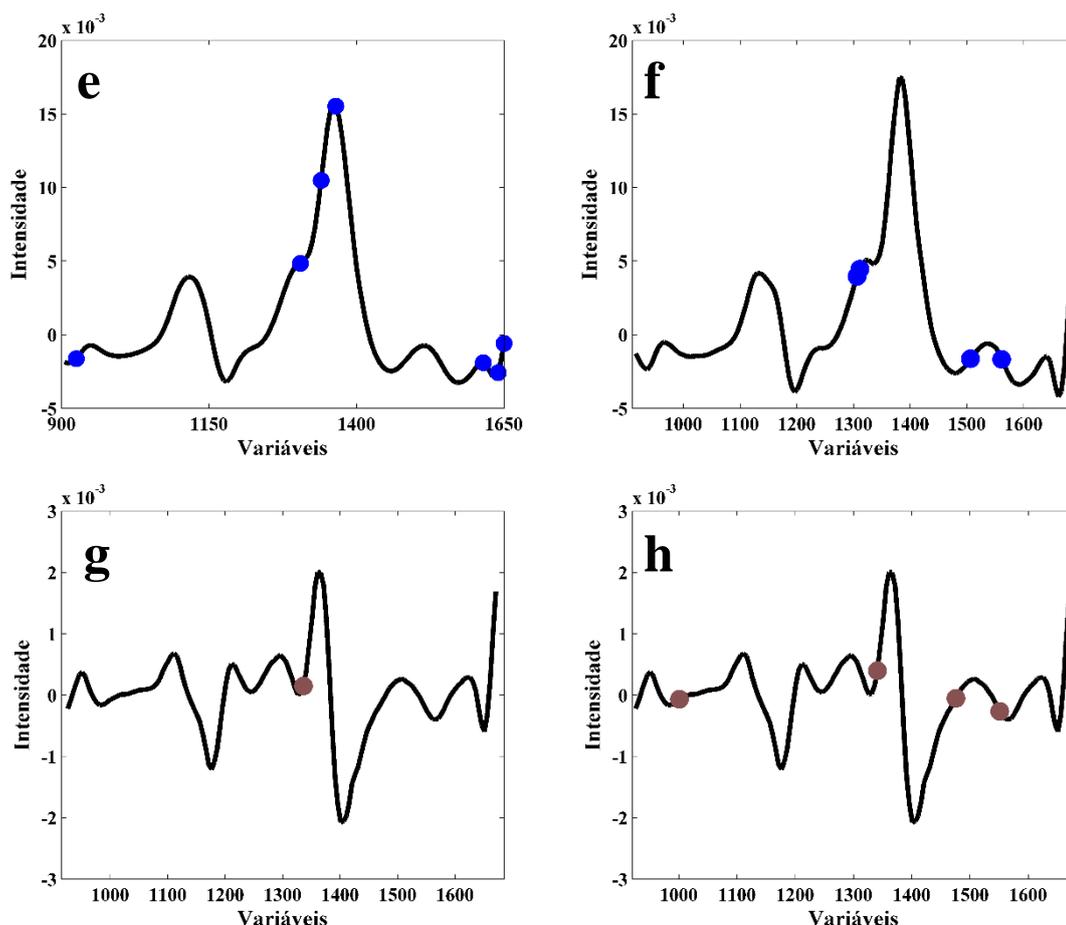


Nesta Tabela IV-1, PD representa a classe de amostras não adulteradas e PA a classe de amostras adulteradas com amido. Como pode ser visto na tabela, todos os modelos produziram taxa de classificação e parâmetros de desempenho iguais a 100%.

Modelos SPA-LDA e GA-LDA foram obtidos utilizando os dados sem pré-processamento e com diferentes pré-processamentos. Nas Figura IV-6 (a, c, e e g) são apresentadas a variáveis selecionadas pelos modelos SPA-LDA e nas Figura IV-6 (b, d, f e h) as variáveis selecionadas pelo GA-LDA. Para as duas metodologias de classificação utilizadas, foram apresentados apenas os pré-processamentos que proporcionaram os resultados mais satisfatórios para a classificação em amostras não adulteradas e adulteradas com amido.

Figura IV-6 - Espectros médios do subconjunto de treinamento com as variáveis selecionadas pelo SPA-LDA e GA-LDA.





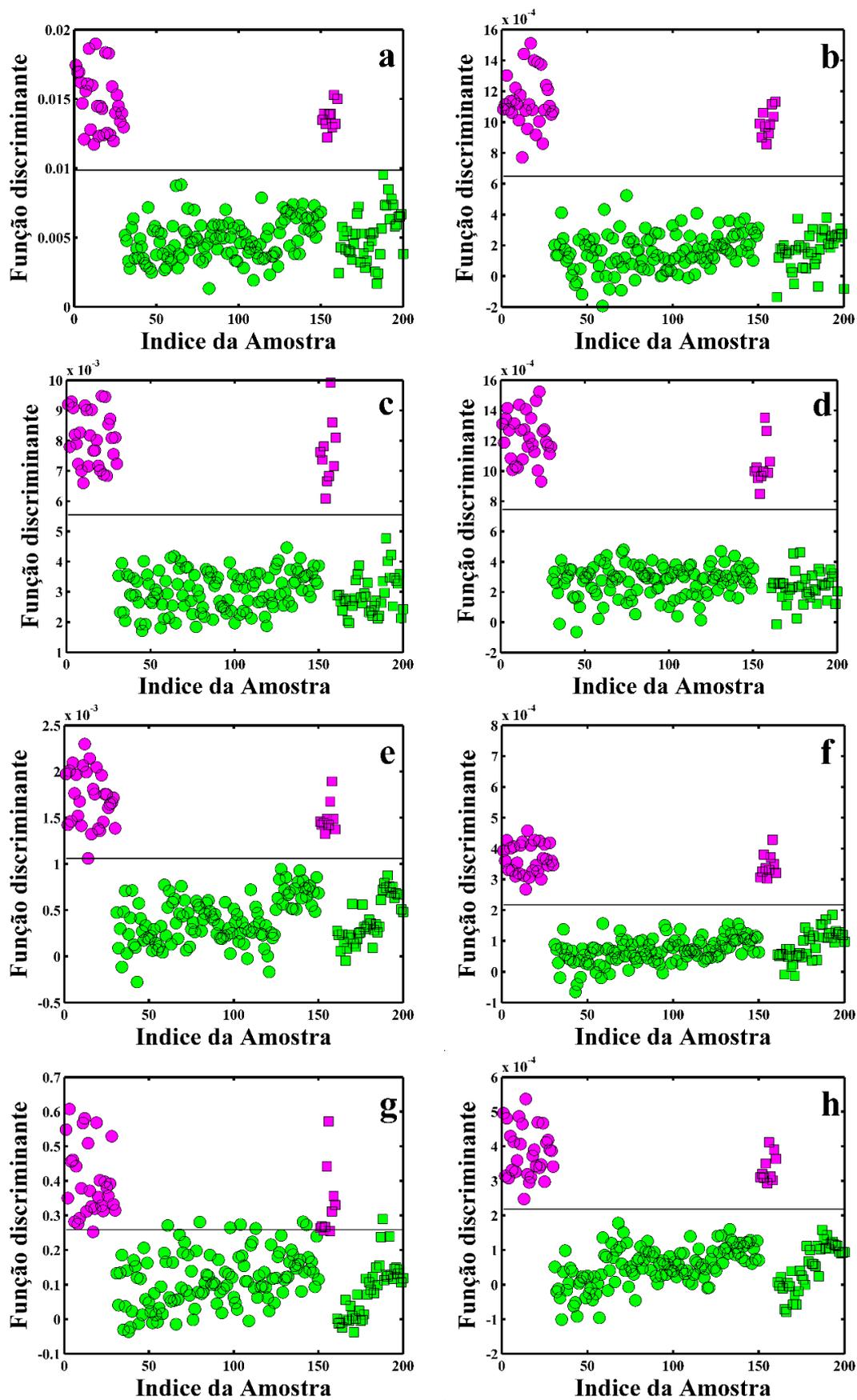
Dados sem pré-processamento (●), linha de base offset (●), 1<sup>o</sup> Der. S.G 11\* (●) e 2<sup>o</sup> Der. S.G 11\* (●)

As variáveis selecionadas estão distribuídas ao longo de todo o espectro quando o SPA foi empregado aos espectros NIR que utilizaram os dados sem pré-processamento e correção de linha de base *offset* (Figura IV-6a e Figura IV-6c, respectivamente), já o GA selecionou menos variáveis para esses tratamentos (Figura IV-6b e Figura IV-6d, respectivamente). Possivelmente, o SPA-LDA precisou de mais variáveis para modelar a linha de base que não foi removida dos dados. O SPA e o GA selecionaram variáveis entre 1400 e 1600 nm, região que apresenta o estiramento da ligação O-H que estão presentes nas moléculas do amido (BURNS; CIURCZAK, 2007).

Na Figura IV-7 são apresentados os valores dos scores de Fisher para o modelo SPA-LDA (Figura IV-7a, c, e e g) GA-LDA (Figura IV-7b, d, f e h). Apenas o modelo SPA-LDA construído utilizando a 1<sup>a</sup> Der. S.G com janela de 11 pontos (Figura IV-7g) apresentou erro de classificação, tanto para o conjunto de treinamento, quanto para o conjunto teste.



Figura IV-7 - Gráfico dos scores da função discriminante de Fisher (FD1) versus o número de amostras.



● Não adulteradas, ● adulterada; ●: Amostras de treinamento, ■: Amostras de teste. a) e b) dados sem pré-processamento, c) e d) correção de linha de base *offset*, e) e f) 1ª Der. S.G com janela de 11 e g) e h) 2ª Der. S.G com janela de 11 pontos.

Na Tabela IV-2 são apresentados a matriz de confusão para os modelos LDA no conjunto de teste e o número de variáveis selecionadas. Todos os modelos GA apresentados na Tabela IV-2 classificaram corretamente 100% das amostras de teste. É possível observar que o SPA-LDA com 2º Der. S.G com janela de 11 pontos classificou duas amostras como falso negativo e uma amostra como falso positivo.

Tabela IV-2 - Matriz de confusão obtida para os modelos SPA-LDA e GA-LDA construídos com os espectros NIR das amostras de pprica e pprica com amido

| Pr-processamento                      |    | SPA-LDA |    | GA-LDA |    |
|--|----|---------|----|--------|----|
|  |    | PD      | PA | PD     | PA |
| Sem pr-<br>processamento              | PD | 10      | -  | 10     | -  |
|  | PA | -       | 40 | -      | 40 |
| Variveis Sel.                         |    | 15      |    | 10     |    |
| Correo linha<br>de base <i>offset</i> |    | PD      | PA | PD     | PA |
|  | PD | 10      | -  | 10     | -  |
|  | PA | -       | 40 | -      | 40 |
| Variveis Sel.                         |    | 20      |    | 13     |    |
| 1 Der. S.G 11*                        |    | PD      | PA | PD     | PA |
|  | PD | 10      | -  | 10     | -  |
|  | PA | -       | 40 | -      | 40 |
| Variveis Sel.                         |    | 7       |    | 4      |    |
| 2 Der. S.G 11*                        |    | PD      | PA | PD     | PA |
|  | PD | 8       | 2  | 10     | -  |
|  | PA | 1       | 39 | -      | 40 |
| Variveis Sel.                         |    | 1       |    | 4      |    |

Na Tabela IV-3 so apresentados os parmetros de desempenho do conjunto teste dos modelos SPA-LDA e GA-LDA, para os diferentes pr-processamentos.  possvel observar que todos os modelos GA-LDA construidos, classificaram corretamente 100% das amostras de teste. O SPA-LDA apresentou resultados satisfatorios com taxa de classificao correta de 100%, no entanto, para o modelo construido utilizando 2 Der. S.G com janela de 11 pontos a classe no adulterada obteve preciso de 0.8889, pois uma amostra adulterada foi classificada incorretamente (Tabela IV-2).



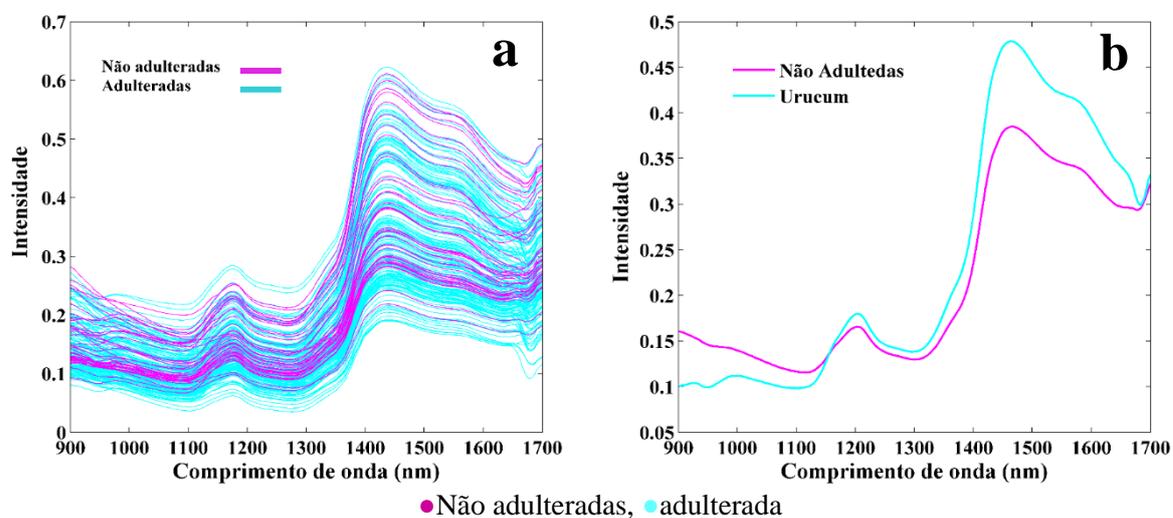
Tabela IV-3 - Parâmetros de desempenho dos resultados dos modelos SPA-LDA e GA-LDA a classe das amostras não adulteradas

|   |         | Precisão | Sensibilidade | Especificidade | TCC (%) |
|---|---------|----------|---------------|----------------|---------|
| Sem pré-processamento                   | SPA-LDA | 1        | 1             | 1              | 100     |
|   | GA-LDA  | 1        | 1             | 1              | 100     |
| Correção linha de base<br><i>offset</i> | SPA-LDA | 1        | 1             | 1              | 100     |
|   | GA-LDA  | 1        | 1             | 1              | 100     |
| 1º Der. S.G<br>11*                      | SPA-LDA | 1        | 1             | 1              | 100     |
|   | GA-LDA  | 1        | 1             | 1              | 100     |
| 2º Der. S.G<br>11*                      | SPA-LDA | 0.8889   | 0.8000        | 0.9750         | 94      |
|   | GA-LDA  | 1        | 1             | 1              | 100     |

#### 4.1.2. Adulteração da Páprica Doce com Urucum

Na Figura IV-8 são apresentados os espectros NIR das amostras de páprica não adulteradas e adulteradas com urucum.

Figura IV-8 - Espectros NIR da páprica sem adulteração e adulterada com urucum sem nenhum pré-processamento na faixa de 900 a 1700 nm



Ao analisar a Figura IV-8(a e b), é possível notar uma faixa de maior intensidade entre 1400 e 1600 nm, região que corresponde as bandas de combinação da ligação C-H do grupo CH<sub>3</sub> e CH<sub>2</sub>, bandas de primeiro sobretom de estiramento da ligação N-H em proteínas e O-H da celulose. A faixa de 1100 a 1250 nm também apresenta intensidade elevada, bandas que são associadas ao segundo sobretom de C-H e ao quarto sobretom de estiramento do grupo C=O (BURNS; CIURCZAK, 2007). O urucum contém em sua composição uma quantidade elevada

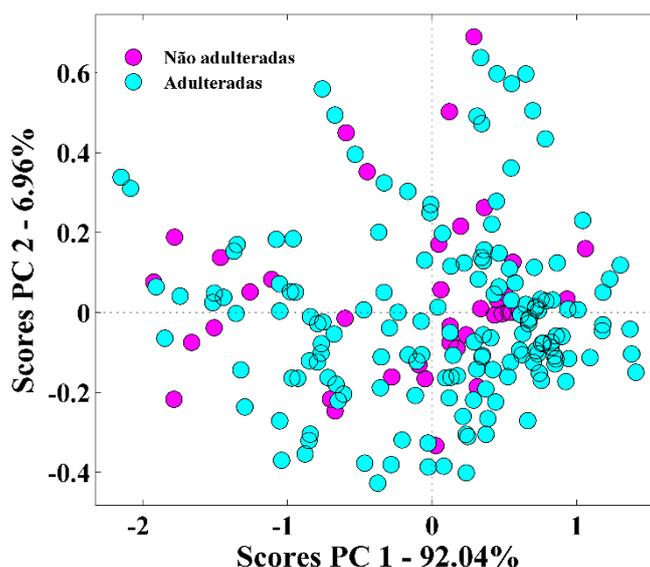


de celulose e proteínas, o que pode justificar as regiões como maior intensidade (LEMOS, 2008; VALÉRIO, et al., 2015).

#### 4.1.2.1. Análise das Componentes Principais (PCA) para os espectros NIR das amostras sem adulteração e adulterada com urucum

Uma análise de componentes principais foi realizada nos dados sem pré-processamento. Na Figura IV-9 é encontrado o gráfico de PC1 versus PC2. A PC1 explicou 92,04% da variância dos dados e a PC2 8,29%. Um comportamento muito parecido com o já observado nas amostras adulteradas com amido é encontrado nesta figura. No qual, os scores das amostras não adulteradas são observados dispersos em valores negativos de PC1. Neste gráfico (Figura IV-9), nota-se uma grande sobreposição dos dados, sendo necessária a aplicação de ferramentas quimiométricas para melhor interpretação dos dados.

Figura IV-9 - Gráfico de PC1 versus PC2 aplicada aos espectros sem pré-processamento NIR para Páprica adulterada com urucum

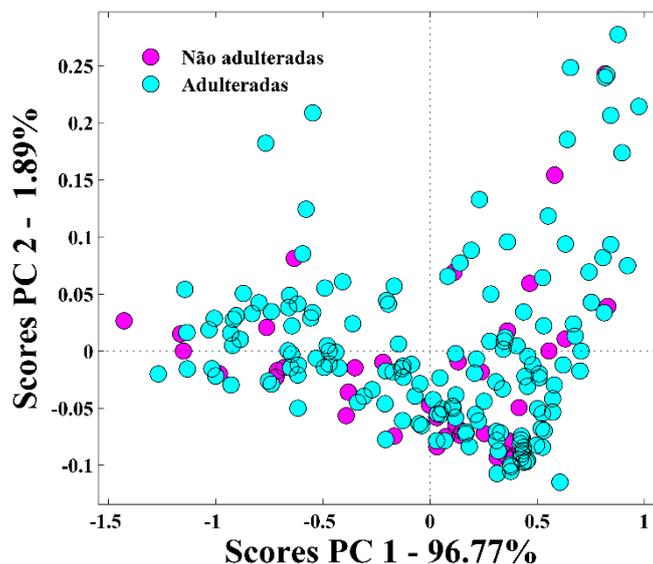


Na Figura IV-10 são apresentados os scores da PC1 versus PC2 dos espectros NIR das amostras de páprica doce e páprica doce com urucum após a correção de linha de base *offset*. É possível notar que após a correção de linha de base *offset*, os scores da PC1 passaram a explicar 96,77% da variância explicada dos dados e os da PC2 1,89%. As amostras não adulteradas passaram a agrupar mais nos scores negativos da PC1 (Figura IV-10) do quando comparadas aos da PC1 (Figura IV-9) dos dados sem pré-processamento. No entanto, as duas classes ainda



apresentam sobreposição, sendo necessária a aplicação de ferramentas quimiométricas para que haja separação entre as classes.

Figura IV-10 - Gráfico de PC1 versus PC2 aplicada aos espectros NIR com correção de linha de base *offset* para Páprica adulterada com urucum



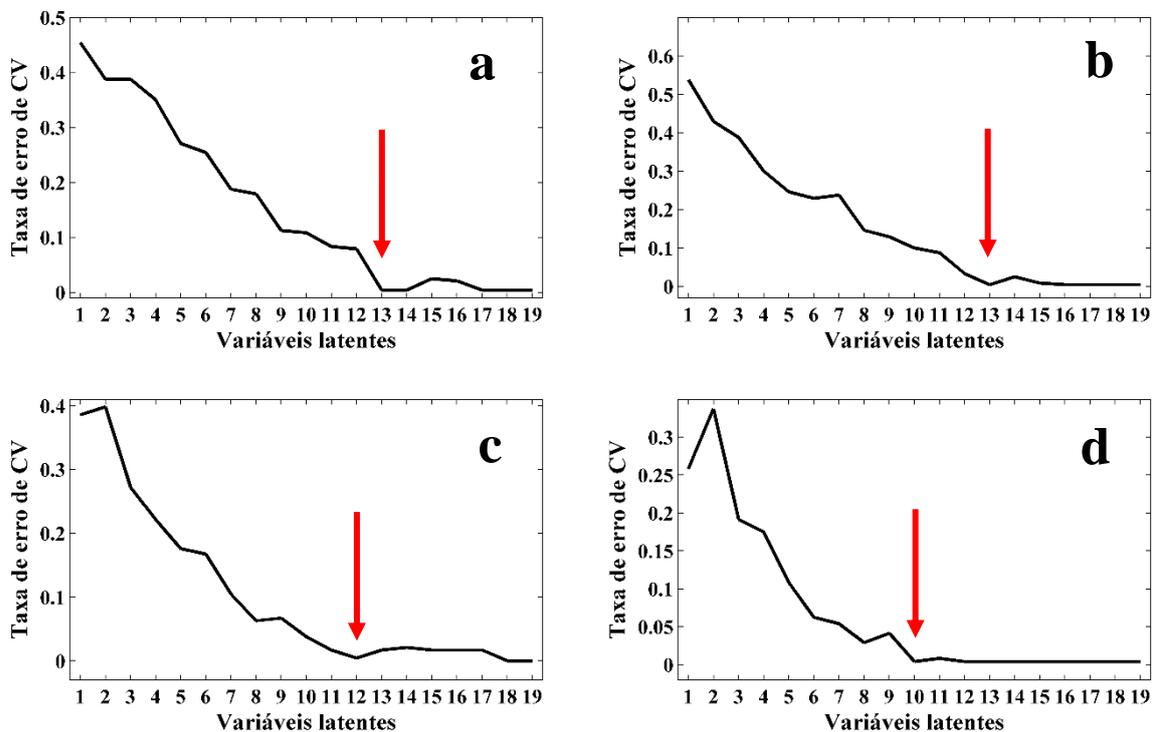
#### 4.1.2.2. Métodos de Reconhecimento de Padrões Supervisionados aplicados as amostras não adulteradas e adulteradas com urucum

Os modelos de classificação PLS-DA entre amostras não adulteradas e adulteradas com urucum foram obtidos para os dados sem pré-processamento e com diferentes pré-processamentos. O número ótimo de variáveis latentes utilizado na construção de cada modelo PLS-DA foi obtido de acordo com o menor erro de classificação durante o processo de validação cruzada.

Na Figura IV-11, as setas em vermelho destacam o número de variáveis latentes escolhido para a construção do modelo com os dados sem pré-processamento (Figura IV-11a), correção linha de base *offset* (Figura IV-11b), 1ª Der. S.G com janelas de 7 pontos (Figura IV-11c) e 2ª Der. S.G com janela de 11 pontos (Figura IV-11d), os números indicados foram 13, 13, 12 e 10, respectivamente.

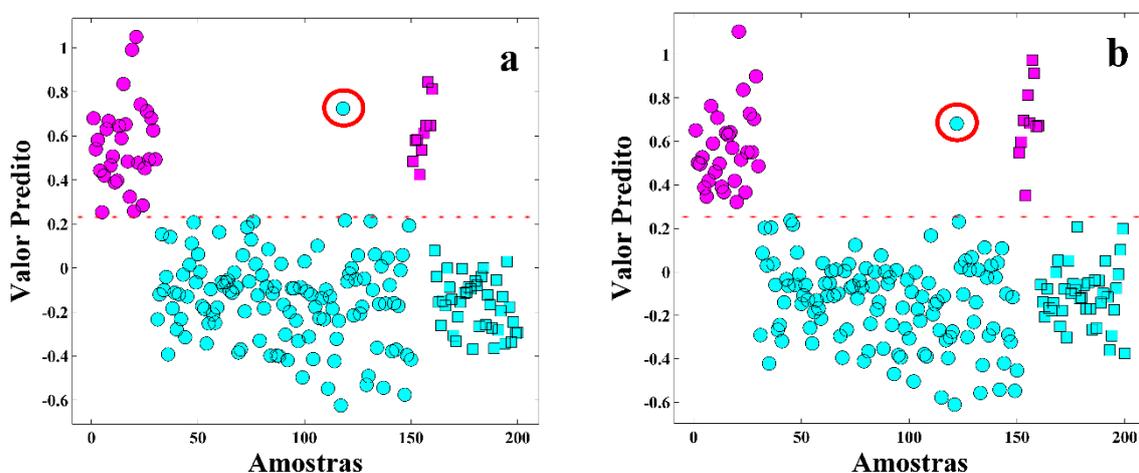


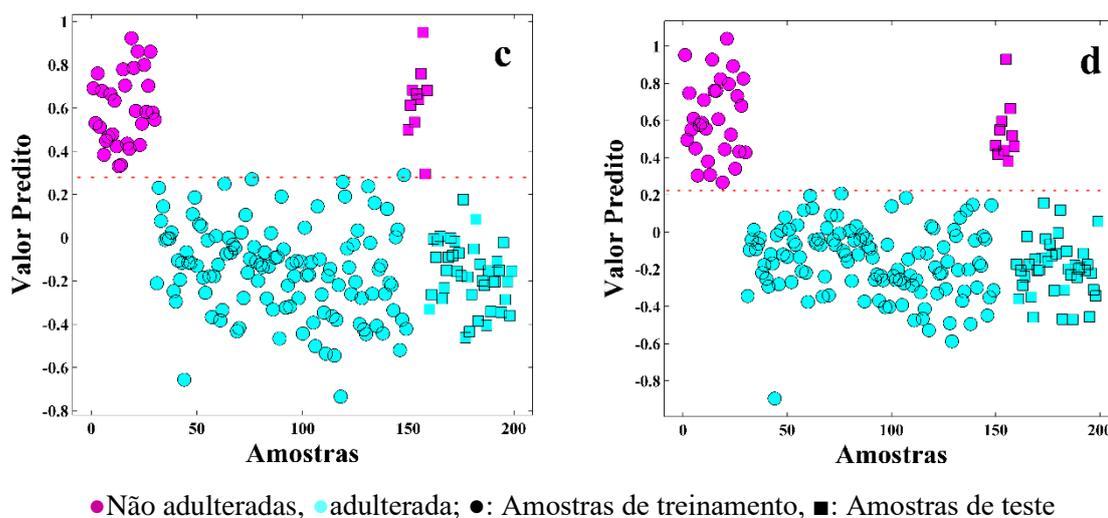
Figura IV-11 - Taxa de erro de validação cruzada versus o número de variáveis latentes incluídas nos modelos PLS-DA



Na Figura IV-12 são apresentados valores estimados para os modelos PLS-DA desenvolvido para as classes não adulteradas e adulteradas com urucum, levando em consideração os pré-processamentos que proporcionaram resultados mais satisfatórios, considerando a TCC no conjunto de treinamento e o menor número de variáveis latentes.

Figura IV-12 - Valores de classe estimados pelo PLS-DA versus o gráfico de amostras de treinamento e teste das amostras não adulteradas e adulteradas com urucum.





É apresentado o modelo com valor limiar de 0,2311 para os dados sem pré-processamento (Figura IV-12a), 0,2311 para correção linha de base *offset* (Figura IV-12b), 0,2785 para 1º Der. S.G 11\* (Figura IV-12c) e 0,2225 para 2º Der. S.G 11\* (Figura IV-12d). É possível notar que uma amostra foi classificada incorretamente nos modelos construídos com os dados sem pré-processamento e correção linha de base *offset*, destacada em vermelho (Figura IV-12a e Figura IV-12b).

Na Tabela IV-4 são apresentados os resultados detalhados para os modelos PLS-DA utilizando o conjunto de treinamento. O modelo PLS-DA construído utilizando 1ª Der. S.G 7\* e 2ª Der. S.G 11\* apresentou 100% de classificação correta. Uma amostra foi classificada como falso positivo pelos modelos dos dados sem pré-processamento e correção linha de base *offset*, no entanto, foi uma amostra diferente em cada modelo, levando a uma TCC de 99 % para os dois modelos.

Tabela IV-4 - Matriz de confusão obtida para os modelos PLS-DA para o conjunto de treinamento. Entre parênteses são apresentados o número de variáveis latentes utilizados.

|                 |    | Sem pré-processamento |     | Correção linha de base <i>offset</i> |     | 1ª Der. S.G 11* |     | 2ª Der. S.G 11* |     |
|-----------------|----|-----------------------|-----|--------------------------------------|-----|-----------------|-----|-----------------|-----|
|                 |    | PD                    | PU  | PD                                   | PU  | PD              | PU  | PD              | PU  |
| Matriz confusão | PD | 30                    | -   | 30                                   | -   | 30              | -   | 30              | -   |
|                 | PU | 1                     | 119 | 1                                    | 119 | -               | 120 | -               | 120 |
| Precisão        |    | 0.9677                |     | 0.9677                               |     | 1               |     | 1               |     |
| Sensibilidade   |    | 1.0000                |     | 1.0000                               |     | 1               |     | 1               |     |
| Especificidade  |    | 0.9917                |     | 0.9917                               |     | 1               |     | 1               |     |
| TCC (%)         |    | 99                    |     | 99                                   |     | 100             |     | 100             |     |



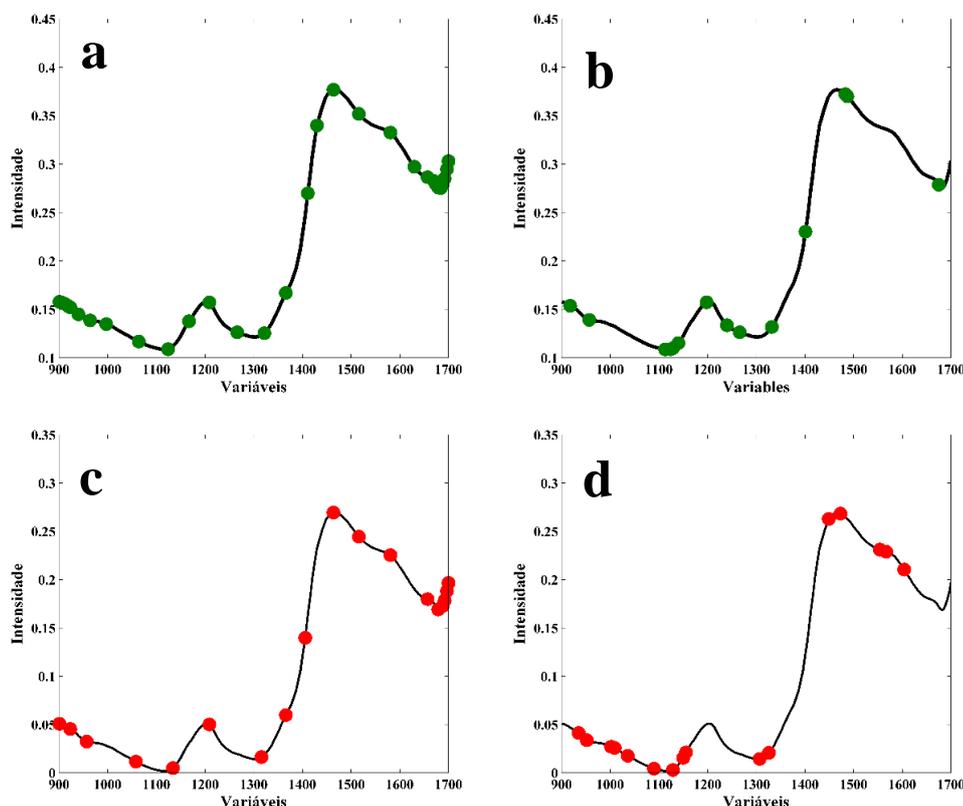
A matriz confusão obtida para os modelos PLS-DA no conjunto de teste pode ser vista na Tabela IV-5. Todos os modelos classificaram corretamente as amostras do conjunto de teste. Proporcionando parâmetros de desempenho iguais a 100%.

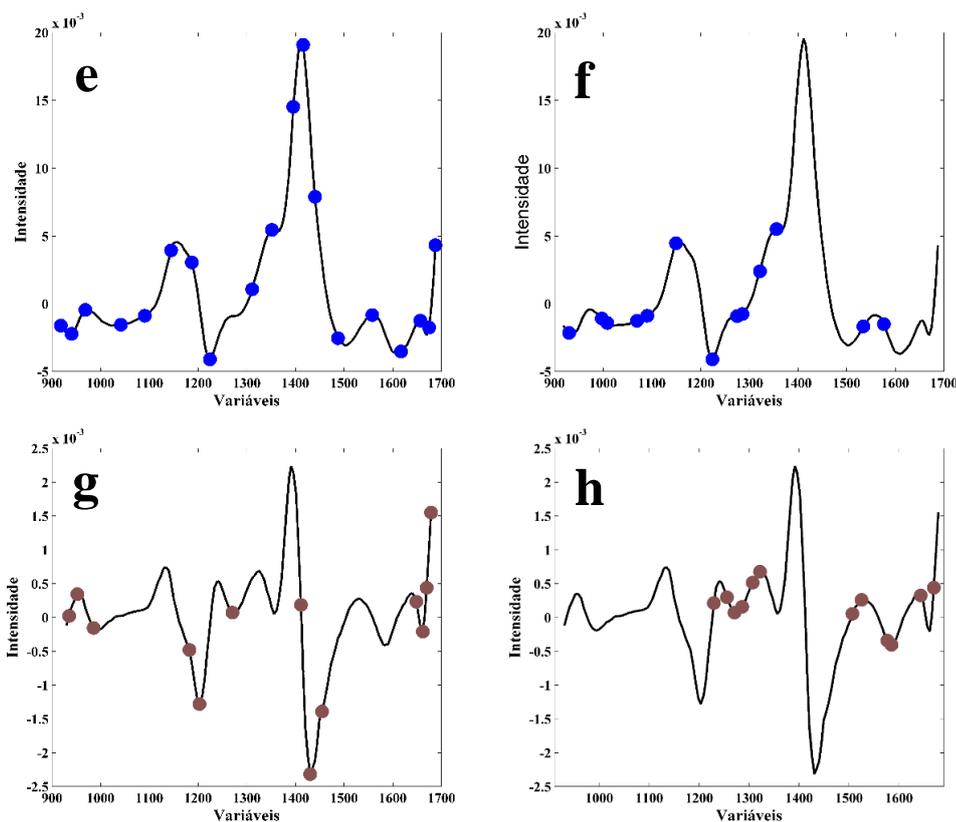
Tabela IV-5 - Matriz de confusão obtida para os modelos PLS-DA para o conjunto de teste. Entre parênteses são apresentados o número de variáveis latentes utilizados.

| Matriz confusão | PD | Sem pré-processamento |    | Correção linha de base <i>offset</i> |    | 1° Der. S.G 7* |    | 2° Der. S.G 11* |    |
|-----------------|----|-----------------------|----|--------------------------------------|----|----------------|----|-----------------|----|
|                 |    | PD                    | PU | PD                                   | PU | PD             | PU | PD              | PU |
|                 | PD | 10                    | -  | 10                                   | -  | 10             | -  | 10              | -  |
|                 | PU | -                     | 40 | -                                    | 40 | -              | 40 | -               | 40 |
| Precisão        |    | 1                     |    | 1                                    |    | 1              |    | 1               |    |
| Sensibilidade   |    | 1                     |    | 1                                    |    | 1              |    | 1               |    |
| Especificidade  |    | 1                     |    | 1                                    |    | 1              |    | 1               |    |
| TCC (%)         |    | 100                   |    | 100                                  |    | 100            |    | 100             |    |

Nas Figura IV-13 (a, c, e e g) são apresentadas a variáveis selecionadas pelos modelos SPA-LDA e nas Figura IV-13 (b, d, f e h) as variáveis selecionadas pelo GA-LDA.

Figura IV-13 - Espectros médios do subconjunto de treinamento com as variáveis selecionadas pelo SPA-LDA e GA-LDA.





Dados sem pré-processamento (●), linha de base *offset* (●), 1° Der. S.G 7\* (●) e 2° Der. S.G 11\* (●)

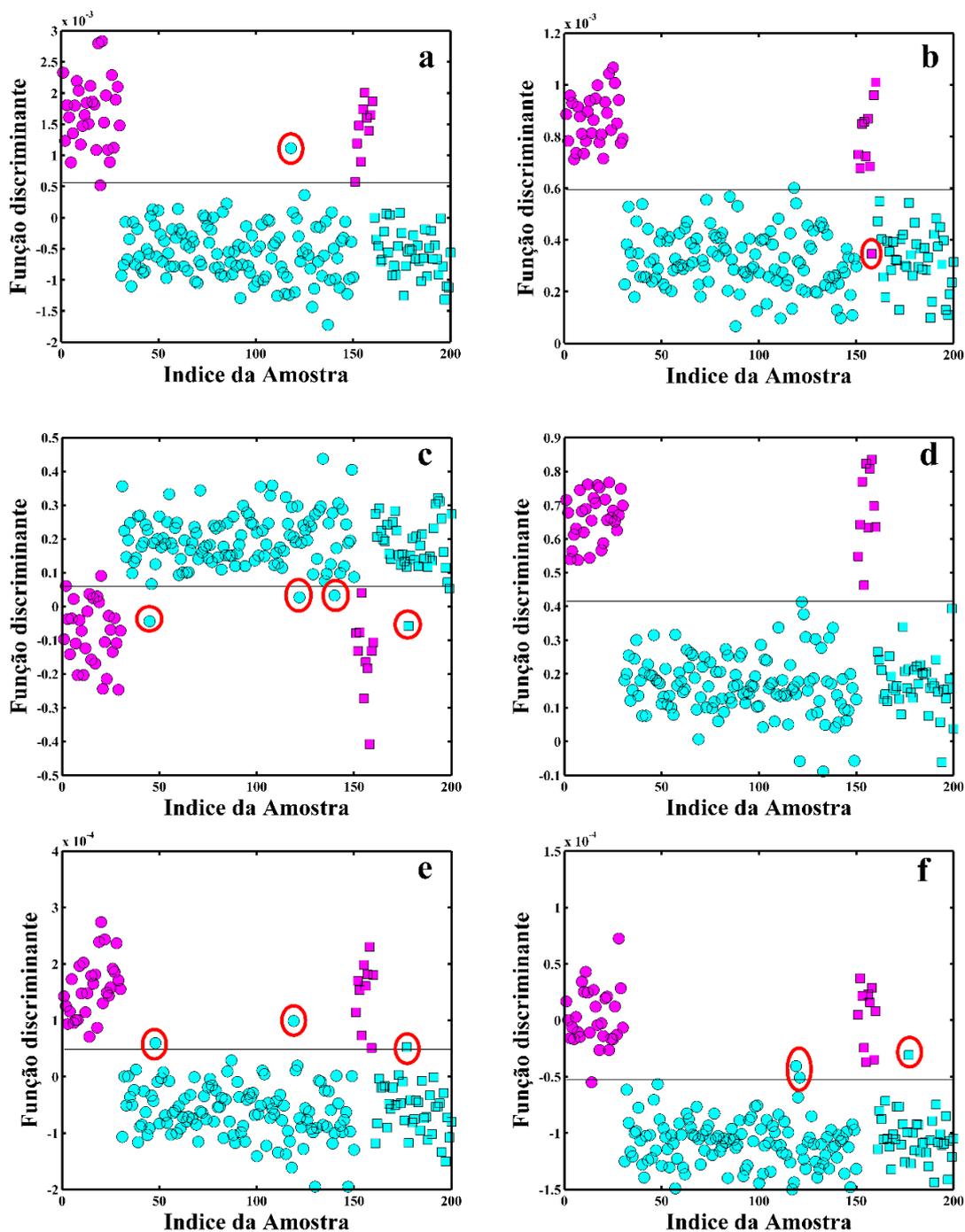
São apresentadas as variáveis selecionadas pelos modelos SPA-LDA e GA-LDA construídos apenas com os pré-processamentos que proporcionaram os resultados mais satisfatórios para a classificação em amostras não adulteradas e adulteradas com urucum. Quando o SPA e GA foi utilizado na seleção de variáveis para os dados sem pré-processamento (Figura IV-13 a e b), 30 variáveis foram selecionadas em ambos os modelos LDA que estão distribuídas ao longo de todo espectro.

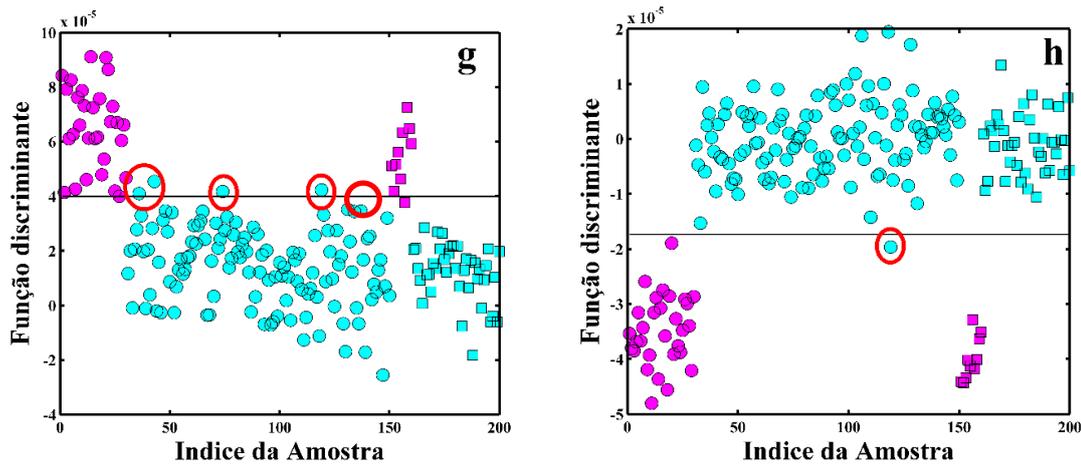
Todos os modelos construídos selecionaram variáveis na região entre 1100 e 1300 nm. Essa região é responsável pelas bandas de segundo sobretom de estiramento de C-H dos grupos CH<sub>3</sub>, CH<sub>2</sub> e CH. Quando o GA-LDA foi utilizado na construção do modelo dos dados da 2° Der. S.G com janela de 11 pontos, variáveis na faixa de 1490 a 1700 foram selecionadas, essa região é responsável pelas bandas de primeiro sobretom de estiramento da ligação N-H presente nas proteínas e primeiro sobretom da ligação O-H da celulose (BURNS; CIURCZAK, 2007). O urucum tem em sua composição quantidade elevada de celulose, proteínas e fibras (LEMOS, 2008; VALÉRIO, et al., 2015), sendo as variáveis selecionadas presentes nas ligações química da sua composição.



Na Figura IV-14 são apresentados os valores dos scores de Fisher para o modelo LDA com as variáveis selecionadas pelo SPA (Figura IV-14a, c, e e g) e pelo GA (Figura IV-14b, d, f e h). Levando em consideração o conjunto de treinamento, apenas os modelos GA-LDA com os dados sem pré-processamento (Figura IV-14b) e com correção de linha de base *offset* (Figura IV-14d) não apresentaram classificação incorreta.

Figura IV-14 - Gráfico dos scores da função discriminante de Fisher (FD1) versus o número de amostras.





● Não adulteradas, ● adulterada; ●: Amostras de treinamento, ■: Amostras de teste. a) e b) dados sem pré-processamento, c) e d) correção de linha de base *offset*, e) e f) 1ª Der. S.G com janela de 7 e g) e h) 2ª Der. S.G com janela de 11 pontos.

Na Tabela IV-6 é apresentada a matriz confusão dos modelos SPA-LDA e GA-LDA das amostras do conjunto de teste.

Tabela IV-6 - Matriz de confusão obtida para os modelos SPA-LDA e GA-LDA construídos com os espectros NIR das amostras de páprica e páprica com Urucum

| Pré-processamento                    |    | SPA-LDA |    | GA-LDA |    |
|--------------------------------------|----|---------|----|--------|----|
|                                      |    | PD      | PU | PD     | PU |
| Sem pré-processamento                | PD | 10      | -  | 9      | 1  |
|                                      | PU | -       | 40 | -      | 40 |
| Variáveis Sel.                       |    | 30      |    | 30     |    |
| Correção linha de base <i>offset</i> | PD | 10      | -  | 10     | -  |
|                                      | PU | 2       | 38 | -      | 40 |
| Variáveis Sel.                       |    | 18      |    | 16     |    |
| 1º Der. S.G 7*                       | PD | 10      | -  | 10     | -  |
|                                      | PU | 1       | 39 | 1      | 39 |
| Variáveis Sel.                       |    | 19      |    | 13     |    |
| 2º Der. S.G 11*                      | PD | 9       | 1  | 10     | -  |
|                                      | PU | -       | 40 | -      | 40 |
| Variáveis Sel.                       |    | 13      |    | 12     |    |

Ainda confirme a Tabela IV-6, o SPA-LDA classificou todas as amostras corretamente apenas quando o modelo foi construído utilizando os dados sem pré-processamento. Já os modelos GA-LDA classificaram corretamente todas as amostras teste, com exceção do modelo construído com dados sem pré-processamento e 1º Der. S.G com janela de 7 pontos, que classificou uma amostra como falso negativo e uma como falso positivo, respectivamente



Os parâmetros de desempenho dos modelos SPA-LDA e GA-LDA para as amostras de teste são apresentados na Tabela IV-7.

Tabela IV-7 - Parâmetros de desempenho dos resultados dos modelos SPA-LDA e GA-LDA

|                                      | Modelo  | Precisão | Sensibilidade | Especificidade | TCC (%) |
|--------------------------------------|---------|----------|---------------|----------------|---------|
| Sem pré-processamento                | SPA-LDA | 1        | 1             | 1              | 100     |
|                                      | GA-LDA  | 0.9677   | 0.9000        | 0.9917         | 99      |
| Correção linha de base <i>offset</i> | SPA-LDA | 0.8333   | 1             | 0.9500         | 96      |
|                                      | GA-LDA  | 1        | 1             | 1              | 100     |
| 1º Der. S.G 11*                      | SPA-LDA | 0.9091   | 1             | 0.9750         | 98      |
|                                      | GA-LDA  | 0.9091   | 1             | 0.9750         | 98      |
| 2º Der. S.G 11*                      | SPA-LDA | 0.9677   | 0.9000        | 0.9917         | 99      |
|                                      | GA-LDA  | 1        | 1             | 1              | 100     |

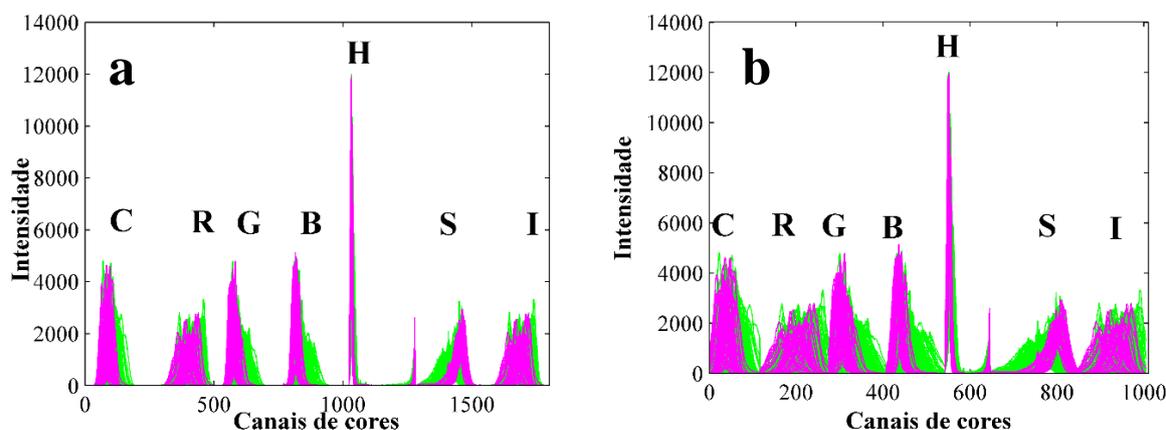
Na Tabela IV-7 é possível notar que quando os dados sem pré-processamentos foram utilizados, apenas o SPA-LDA alcançou a taxa de classificação correta de 100%. O GA-LDA obteve melhores valores de precisão, sensibilidade e especificidade no conjunto treino, quando aplicado aos dados tratados com correção de linha de base *offset* e 2ª Der. S.G com janela de 11 pontos.

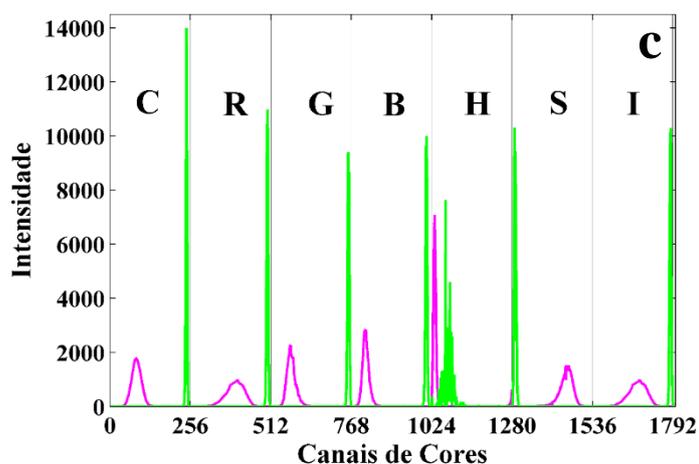
## 4.2. AVALIAÇÃO DAS IMAGENS DIGITAIS

### 4.2.1. Adulteração da Páprica Doce com Amido

O histograma contendo todos os canais de cores pode ser encontrado na Figura IV-15 das imagens digitais da páprica doce não adulteradas e adulteradas com amido.

Figura IV-15 - Histograma com todas as escalas de cores da Páprica doce não adulterada e adulterada com amido





● Não adulteradas, ● Adulteradas (c amido); C: escala de cinzas, R: vermelho; G: verde; B: azul; H: matriz; S: saturação; I: intensidade.

Na Figura IV-15b são apresentados os histogramas após a eliminação de variáveis não informativas (variáveis com frequência de ocorrência iguais a zero). Já o histograma médio da páprica doce e do amido são encontrados na Figura IV-15c.

É possível notar que os histogramas nos canais Cinza (C), Vermelho (R), Verde (G) e azul (B) apresentados nas Figura IV-15a e Figura IV-15b, para as amostras adulteradas, um deslocamento para a direita em relação as não adulteradas. Isto ocorre porque a adulteração com amido adiciona na imagem mais pixels com valores próximos a 255 (Figura IV-15c). Como a escala de cinzas converte cores em tons de cinza a partir da média do R, G e B, este canal também sofre um efeito similar.

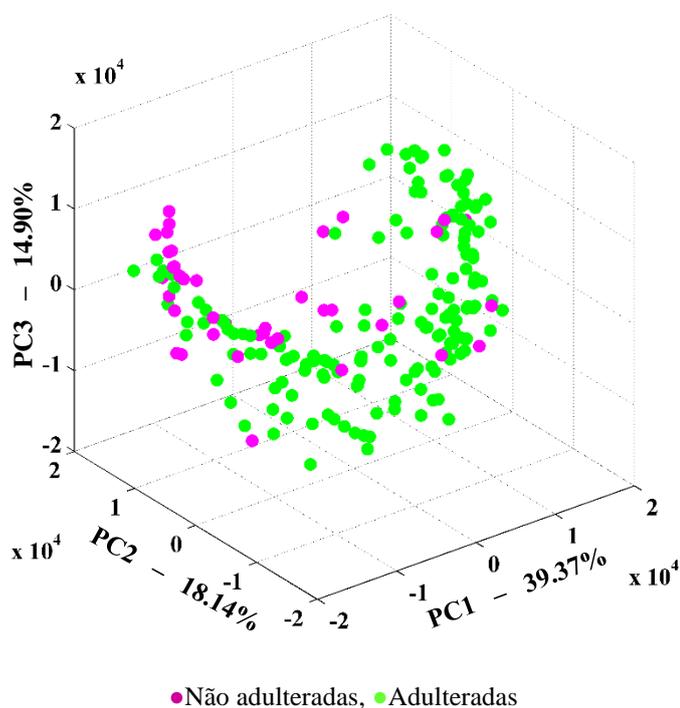
O canal R, que representa o vermelho, se mantém mais alargado no histograma, devido a variação que ocorrer na cor de cada pixel da amostra de páprica doce sem adulteração (Figura IV-15c). Quando o canal S é observado, nota-se que as amostras adulteradas são deslocadas para a esquerda, mostrando que a adulteração torna a cor da amostra mais impura, sendo mais diluída no branco.

#### 4.2.1.1. Análise das Componentes Principais (PCA) do histograma de cores da páprica e páprica com amido

Uma análise por componentes principais foi realizada nos histogramas para todos os canais de cores. Na Figura IV-16 os scores das três primeiras PCs são apresentados. A PC1 explicou 39,37% da variância dos dados, a PC2 18,14% e a PC3 14,90%.



Figura IV-16 - Gráfico dos scores da PC1 x PC2 x PC3 aplicada ao histograma completo para as amostras de páprica doce não adulteradas e adulterada com amido



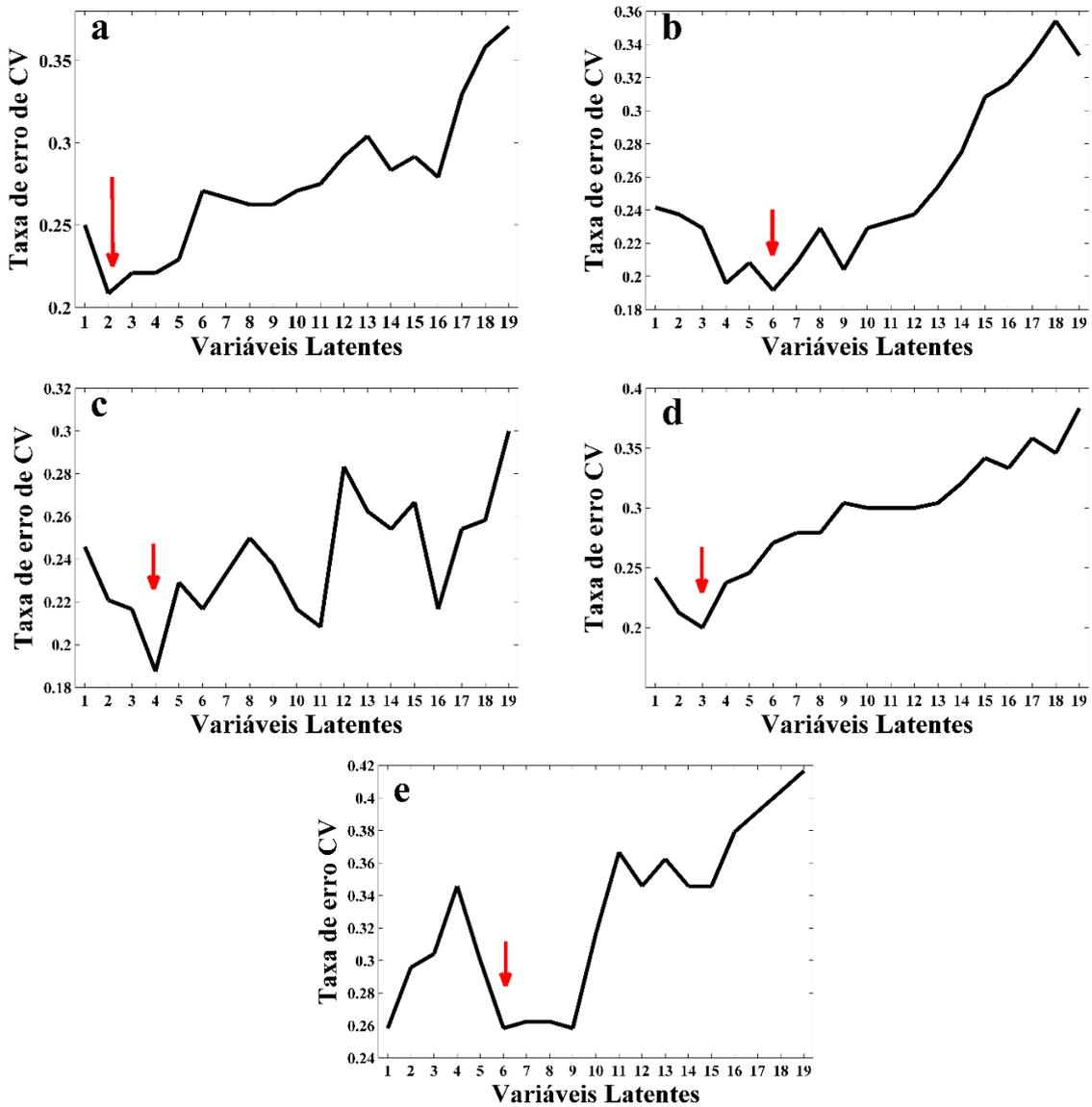
Na Figura IV-16, observa-se que as amostras adulteradas estão distribuídas tanto nos scores negativos quanto nos positivos de PC1 e PC3, enquanto a maioria dos dados em PC2 se concentra nos scores positivos. No entanto, há uma considerável sobreposição entre as amostras adulteradas e não adulteradas. Portanto, é necessária a aplicação de outras ferramentas quimiométricas para aprimorar a separação entre as classes.

#### 4.2.1.2. Métodos de Reconhecimento de Padrões Supervisionados

A Figura IV-17 apresenta o gráfico da taxa de erro pelo número de variáveis latentes para os modelos PLS-DA construídos com o histograma completo, canal RGB, canal HSI, combinação do canal RGB com HSI e G. Esses canais foram escolhidos para que os resultados fossem apresentados com base nos melhores resultados para precisão, sensibilidade, especificidade e TCC para o conjunto de treinamento. Com base na menor taxa de erro foi escolhido o número de variáveis latentes.



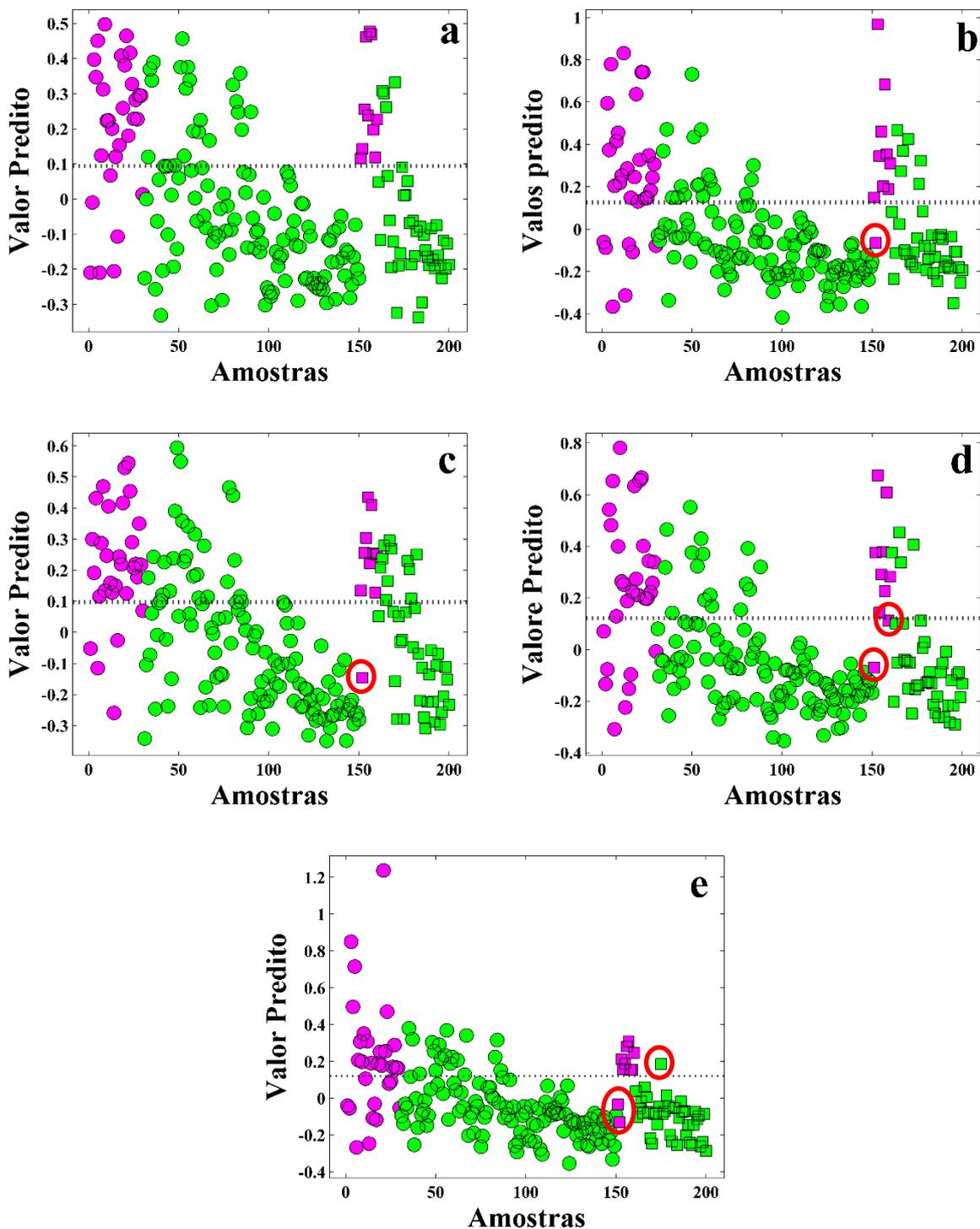
Figura IV-17 - Taxa de erro de validação cruzada versus o número de variáveis latentes



Para o modelo PLS-DA construído com o histograma completo (Figura IV-17a) foram escolhidas três VL, para o canal RGB (Figura IV-17b) foram seis VL, para o canal HSI (Figura IV-17c) foram quatro VL, para a combinação do canal RGB com HSI (Figura IV-17d) foram três VL e para o canal G (Figura IV-17e) foram seis VL.

O valor predito pelos modelos PLS-DA para as amostras páprica doce não adulteradas e adulteradas com amido podem ser vistos na Figura IV-18.

Figura IV-18 - Gráfico dos valores estimados pelo PLS-DA versus o índice de cada amostra não adulteradas e adulterada com amido utilizada nos conjuntos de treinamento e teste



● Não adulteradas, ● Adulteradas, ●: Amostras de treinamento, ■: Amostras de teste, a) Histograma completo, b) RGB, c) HSI, d) RGB+HSI, e) G.

Todos os modelos PLS-DA construídos (Figura IV-18 a, b, c e d), com exceção do canal G, classificaram mais amostras da classe adulterada como sendo da classe não adulterada



do que o contrário, tanto para o conjunto de treinamento, quanto para o conjunto de teste. Já o canal verde (G) (Figura IV-18e) proporcionou apenas uma amostra da classe adulterada como pertencente a não adulterada, e duas amostra da classe não adulterada como sendo adulterada, para o conjunto de teste.

A matriz confusão e os parâmetros de desempenho dos modelos construídos podem ser vistos na Tabela IV-8.

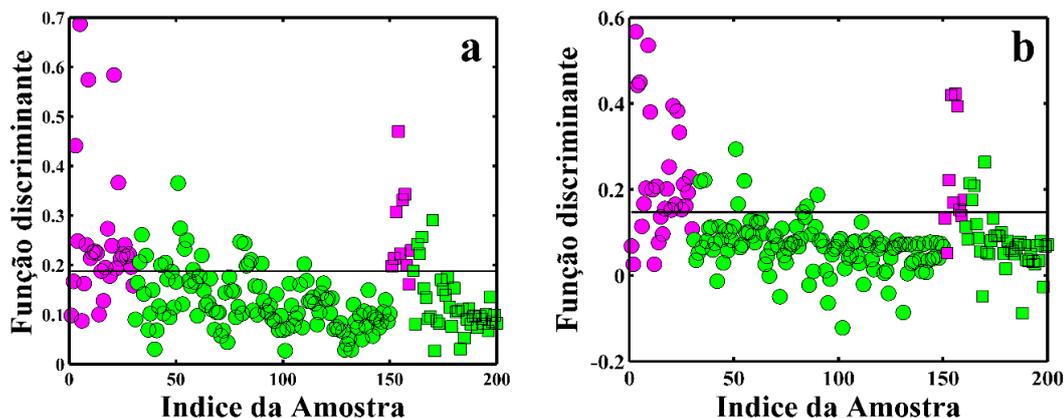
Tabela IV-8 - Matriz de confusão obtida para os modelos PLS-DA para o conjunto de teste do histograma de cor. Entre parênteses são apresentados o número de variáveis latentes utilizados

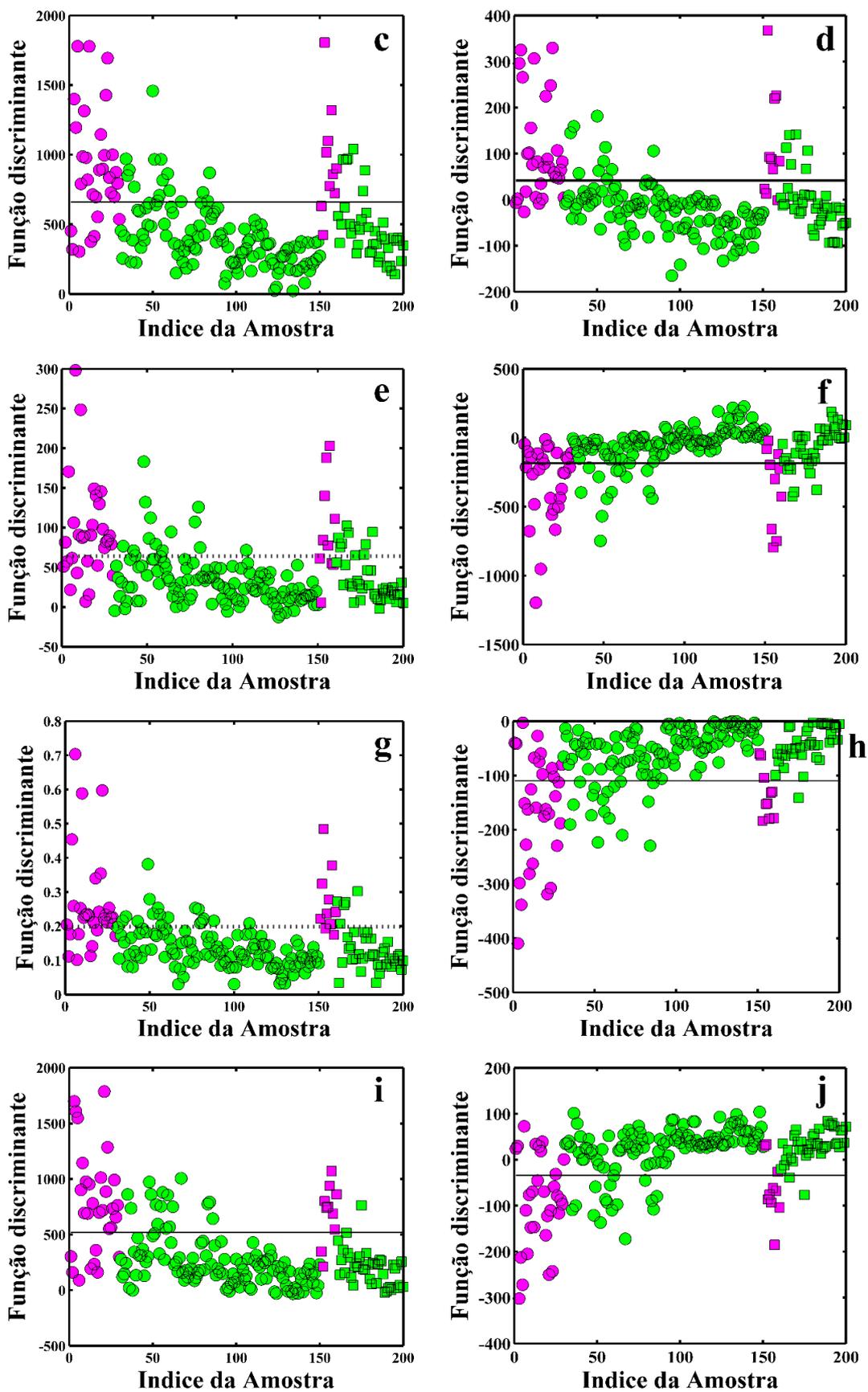
|                 |    | Histograma Completo |    | RGB    |    | HSI    |    | RGB + HSI |    | G      |    |
|-----------------|----|---------------------|----|--------|----|--------|----|-----------|----|--------|----|
|                 |    | PD                  | PA | PD     | PA | PD     | PA | PD        | PA | PD     | PA |
| Matriz confusão | PD | 7                   | 3  | 9      | 1  | 9      | 1  | 8         | 2  | 8      | 2  |
|                 | PA | 4                   | 36 | 6      | 34 | 11     | 29 | 4         | 36 | 1      | 39 |
| Precisão        |    | 0.6364              |    | 0.6000 |    | 0.4500 |    | 0.6667    |    | 0.8889 |    |
| Sensibilidade   |    | 0.7000              |    | 0.9000 |    | 0.9000 |    | 0.8000    |    | 0.8000 |    |
| Especificidade  |    | 0.9000              |    | 0.8500 |    | 0.7250 |    | 0.9000    |    | 0.9750 |    |
| TCC (%)         |    | 86                  |    | 86     |    | 76     |    | 88        |    | 94     |    |

O modelo construído utilizando o canal G apresentou o melhor resultado (Tabela IV-8), com taxa de classificação correta de 94% e precisão de 0.8889, sendo o maior valor quando comparado aos outros modelos.

Os gráficos dos scores de Fisher obtidos para os modelos de classificação SPA-LDA (figuras à esquerda) e GA-LDA (figuras à direita) são apresentados na Figura IV-19.

Figura IV-19 - Gráfico de scores obtido por SPA-LDA e GA-LDA para as amostras de páprica doce não adulteradas e adulterada com amido





● Não adulteradas, ● Adulteradas; ●: Amostras de treinamento, ■: Amostras de teste



Os modelos SPA-LDA e GA-LDA foram construídos utilizando o histograma completo (Figura IV-19a e b), RGB (Figura IV-19c e d), HSI (Figura IV-19e e f), combinação do RGB + HSI (Figura IV-19g e h) e para o canal G (Figura IV-19i e j), respectivamente.

Na Figura IV-19 maioria dos erros obtidos foram amostras da classe de adulteradas com menores concentrações, que foram classificadas como amostras não adulteradas. A exemplo, Figura IV-19e, 16 amostras da classe adulteradas foram classificadas erroneamente como não adulteradas. Entre estas amostras, 12 foram adulteradas com 1% de amido e as outras 5 com 5% de amido.

No conjunto de teste, o modelo SPA-LDA construído utilizando o histograma com todos os canais (Tabela IV-9) classificou apenas uma amostra do conjunto de teste como falso negativo e 5 amostras como falso positivo. Já para o GA-LDA, todos os modelos classificaram 3 amostras como falso negativo.

Tabela IV-9 - Matriz de confusão obtida para os modelos SPA-LDA e GA-LDA construídos com histograma de cores das amostras de páprica e páprica com amido.

| Canal de Cor        |    | SPA-LDA |    | GA-LDA |    |
|---------------------|----|---------|----|--------|----|
|                     |    | PD      | PA | PD     | PA |
| Histograma completo | PD | 9       | 1  | 7      | 3  |
|                     | PA | 5       | 35 | 4      | 36 |
| Variáveis Sel.      |    | 3       |    | 5      |    |
| RGB                 | PD | 8       | 2  | 7      | 3  |
|                     | PA | 6       | 34 | 6      | 34 |
| Variáveis Sel.      |    | 6       |    | 5      |    |
| HSI                 | PD | 6       | 4  | 7      | 3  |
|                     | PA | 9       | 31 | 8      | 32 |
| Variáveis Sel.      |    | 4       |    | 7      |    |
| RGB + HSI           | PD | 8       | 2  | 7      | 3  |
|                     | PA | 5       | 35 | 1      | 39 |
| Variáveis Sel.      |    | 3       |    | 7      |    |
| G                   | PD | 8       | 2  | 7      | 3  |
|                     | PA | 1       | 39 | 2      | 38 |
| Variáveis Sel.      |    | 7       |    | 7      |    |



O melhor resultado em termos de TCC, para o modelo SPA-LDA, foi obtido para o canal G, no qual, classificou duas amostras de pprica doce no adulterada como adulterada e apenas uma amostra adulterada como no adulterada. Para o GA-LDA, o melhor resultado foi encontrado na combinao RGB + HSI, com trs amostras no adulteradas classificadas como adulterada e apenas uma amostra adulterada sendo classificada como no adulterada. As amostras classificadas incorretamente como pertencentes a classe PD, foram identificadas como sendo com concentrao de 1% (maioria) e 5% de adulterante.

Na Tabela IV-10 so fornecidos os parmetros de desempenho dos melhores modelos SPA-LDA e GA-LDA, para a classe no adulterada, construdos para a classificao de pprica doce adulterada com amido. Como pode ser observado Tabela IV-10, modelo SPA-LDA com o canal G apresentou o melhor resultado de preciso, sensibilidade e especificidade, com valores de 0,8889, 0,8000 e 0,9750, respectivamente, e uma taxa de classificao correta de 94% das amostras do conjunto de teste. J para o GA-LDA, o melhor resultado foi encontrado quando a combinao RGB + HSI foi utilizada, com taxa de classificao correta de 92% e valores de preciso, sensibilidade e especificidade iguais a 0,8750, 0,7000 e 0,9750, respectivamente.

Tabela IV-10 - Parmetros de desempenho dos modelos SPA-LDA e GA-LDA para o conjunto de teste

|                            | Modelos        | Preciso | Sensibilidade | Especificidade | TCC (%) |
|----------------------------|----------------|----------|---------------|----------------|---------|
| <b>Histograma completo</b> | <b>SPA-LDA</b> | 0.6429   | 0.9000        | 0.8750         | 88      |
|                            | <b>GA-LDA</b>  | 0.6363   | 0.7000        | 0.9000         | 86      |
| <b>RGB</b>                 | <b>SPA-LDA</b> | 0.5714   | 0.8000        | 0.8500         | 84      |
|                            | <b>GA-LDA</b>  | 0.5384   | 0.7000        | 0.8500         | 82      |
| <b>HSI</b>                 | <b>SPA-LDA</b> | 0.4000   | 0.6000        | 0.7750         | 74      |
|                            | <b>GA-LDA</b>  | 0.4667   | 0.7000        | 0.8000         | 78      |
| <b>RGB + HSI</b>           | <b>SPA-LDA</b> | 0.6154   | 0.8000        | 0.8750         | 86      |
|                            | <b>GA-LDA</b>  | 0.8750   | 0.7000        | 0.9750         | 92      |
| <b>G</b>                   | <b>SPA-LDA</b> | 0.8889   | 0.8000        | 0.9750         | 94      |
|                            | <b>GA-LDA</b>  | 0.7778   | 0.7000        | 0.9500         | 90      |

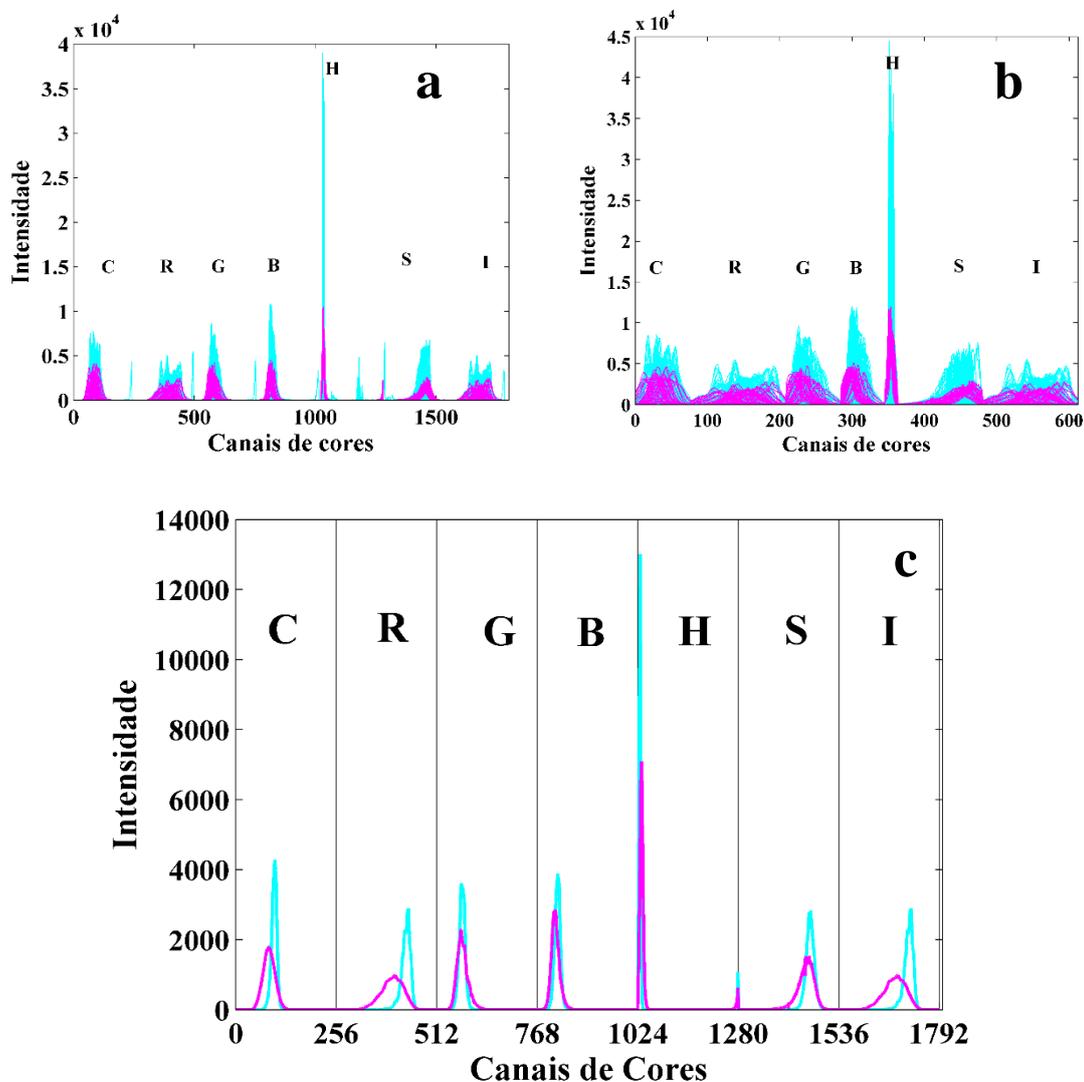
#### 4.2.2. Adulterao da Pprica Doce com Urucum

Na Figura IV-20a e Figura IV-20b so apresentados os histogramas com todas as variveis e aps a eliminao de variveis no significativas (remoo das variveis com valores iguais a zero para todas as amostras), respectivamente. Na Figura IV-20 o histograma



médio para todas as amostras de páprica doce e do urucum, podem ser observados na Figura IV-20c.

Figura IV-20 - Histograma com todas as escalas de cores da Páprica doce não adulterada e adulterada com Urucum



● Não adulteradas, ● Adulteradas (c urucum); C: escala de cinzas, R: vermelho; G: verde; B: azul; H: matriz; S: saturação; I: intensidade.

Analisando os histogramas de cores apresentados nas Figura IV-20a, b e c. É perceptível que as amostras adulteradas com urucum apresentam um aumento significativo na intensidade em todos os canais de cores. Essa elevação de intensidade tem relação direta com as cores adicionadas pelo urucum, que neste caso é mais intensa devido a menor variação de cor (Figura IV-20c), e com máximos mais deslocados para 255 quando comparado a páprica doce sem adulteração.

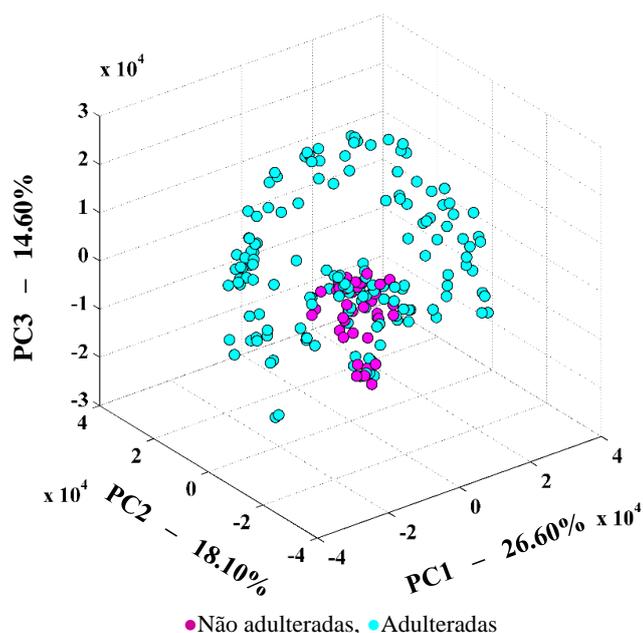


#### 4.2.2.1. Análise das Componentes Principais (PCA) do histograma de cores da páprica doce e adulterada com Urucum

Os scores das três primeiras PCs na análise de componentes principais realizada no histograma das imagens de páprica doce sem adulteração e adulterada com urucum, podem ser vistos na Figura IV-21. A PC1 teve uma variância explicada dos dados de 26.60%, PC2 18,10% e PC3 14.60%.

Nesta figura, é possível notar (Figura IV-21) que as amostras não adulteradas formam um agrupamento central em PC1 e PC2 e com valores negativos nos scores de PC3. As amostras adulteradas se encontram, em sua maioria, distribuídas ao longo de PC1 e PC2 e com uma grande parte das amostras localizadas nos scores positivos de PC3.

Figura IV-21 - Gráfico dos scores da PC1 x PC2 x PC3 aplicada ao histograma completo para as amostras de páprica doce não adulteradas e adulterada com Urucum



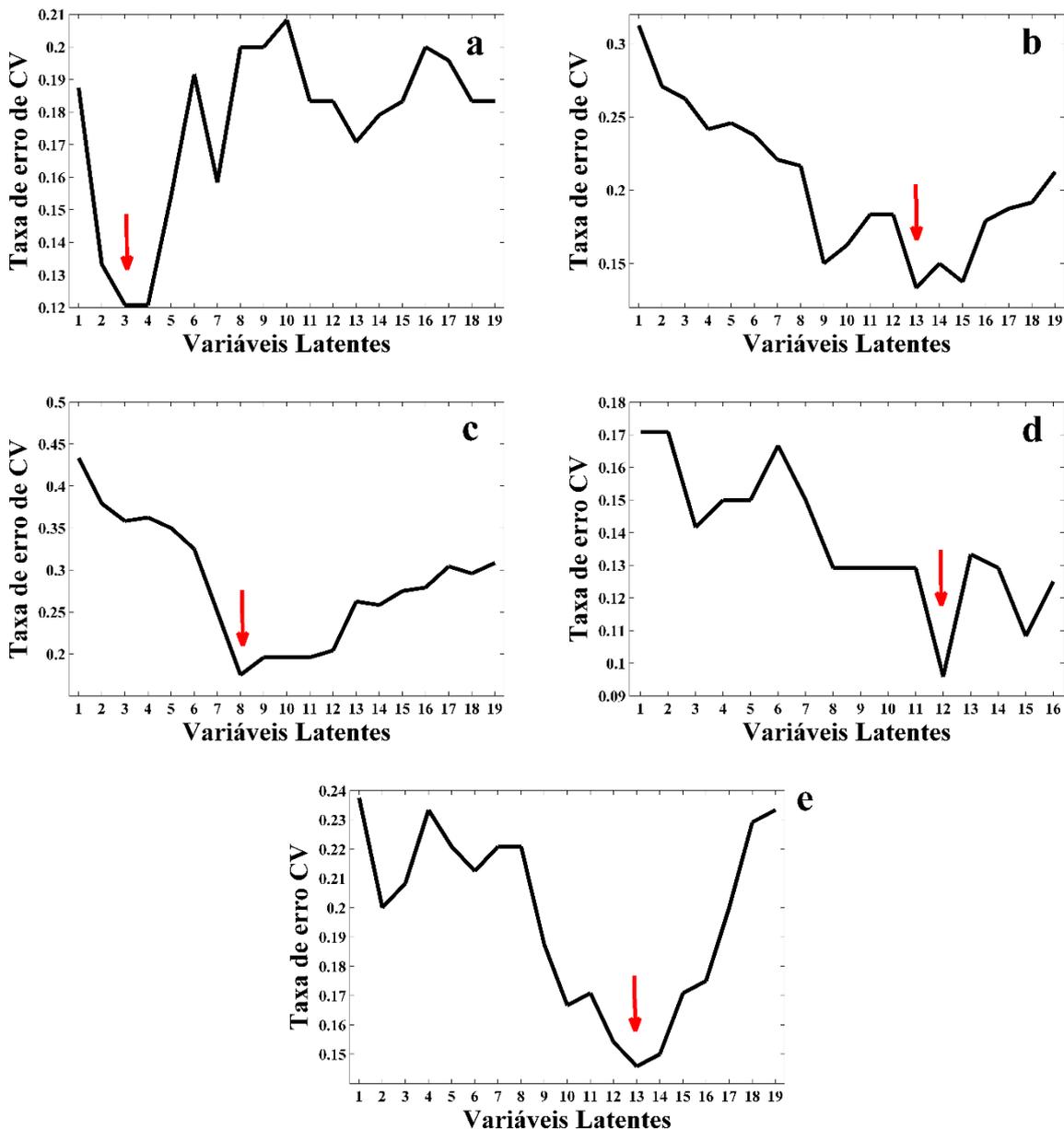
Quando identificadas as concentrações das amostras, as de menor concentração (1% e 5%), apresentaram sobreposição com as amostras não adulteradas. Mostrando a necessidade da utilização de ferramentas quimiométricas para a separação das classes.



4.2.2.2. Métodos de Reconhecimento de Padrões Supervisionados

A Figura IV-22 apresenta o gráfico da taxa de erro de classificação durante o processo de validação cruzada versus o número de variáveis latentes incluídas nos modelos PLS-DA. Com base no menor valor foi escolhido o número ótimo de variáveis latentes.

Figura IV-22 - Taxa de erro de validação cruzada versus o número de variáveis latentes



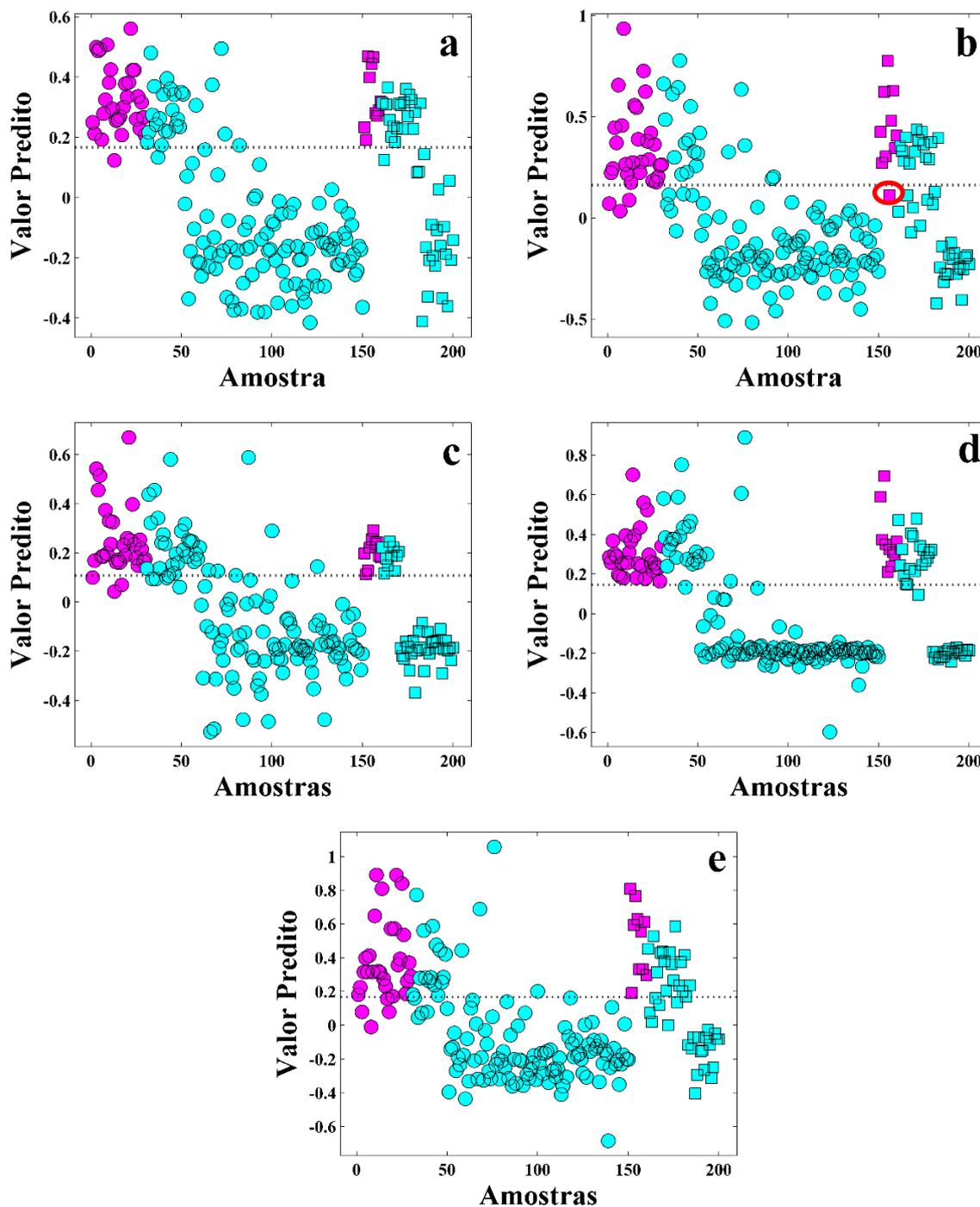
Conforme observado, para o modelo construído com o histograma completo (Figura IV-22a) foram selecionadas três variáveis latentes (VL). Para o canal RGB (Figura IV-22b),



foram escolhidas treze VL; para o canal G (Figura IV-22c), oito VL; para o canal H (Figura IV-22d), doze VL; e para o canal S (Figura IV-22e), treze VL.

O valor predito pelos modelos PLS-DA para as amostras de paprica doce e adulteradas com urucum podem ser vistos na Figura IV-23.

Figura IV-23 - Grafico dos valores estimados pelo PLS-DA versus o ndice de cada amostra no adulteradas e adulterada com Urucum utilizada nos conjuntos de treinamento e teste



● No adulteradas, ● Adulterada, ● Amostras de treinamento, ■ Amostras de teste, a) Histograma completo, b) RGB, c) G, d)H, e) G.



Todos os modelos classificaram um número maior de amostras adulteradas como não adulteradas para o conjunto de treinamento. No conjunto de teste, apenas o modelo construído com o canal RGB (Figura IV-23b) apresentou erro na classe não adulterada.

A matriz de confusão e os parâmetros de desempenho, para os modelos PLS-DA construídos, podem ser vistos na Tabela IV-11.

Tabela IV-11 - Matriz de confusão obtida para os modelos PLS-DA na classificação das amostras o conjunto de teste do histograma de cor. Entre parênteses são apresentados o número de variáveis latentes utilizados

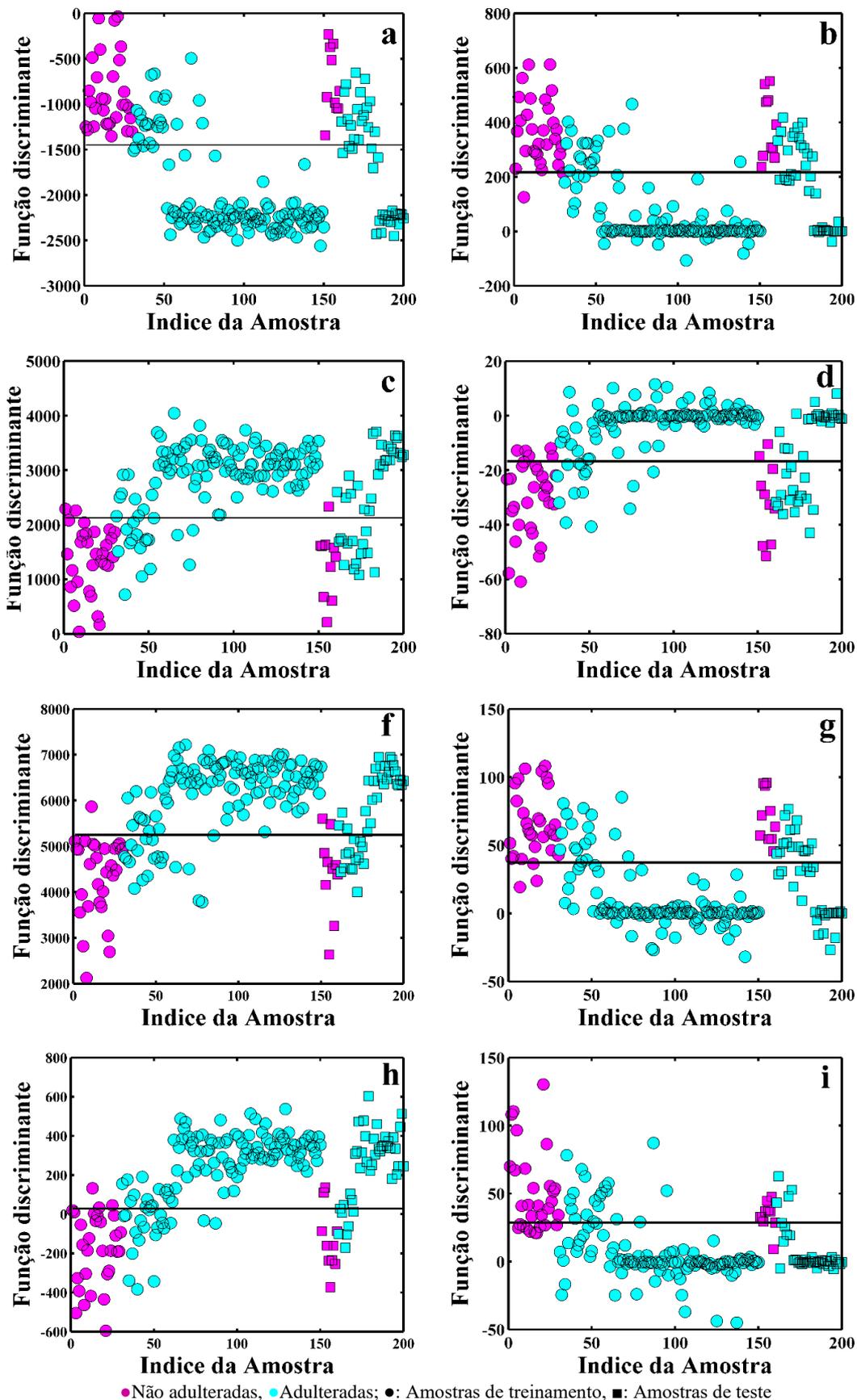
|                        |           | <b>Histograma Completo (3)</b> |           | <b>RGB (13)</b> |           | <b>G (8)</b> |           | <b>H (12)</b> |           | <b>S (13)</b> |           |
|------------------------|-----------|--------------------------------|-----------|-----------------|-----------|--------------|-----------|---------------|-----------|---------------|-----------|
|                        |           | <b>PD</b>                      | <b>PU</b> | <b>PD</b>       | <b>PU</b> | <b>PD</b>    | <b>PU</b> | <b>PD</b>     | <b>PU</b> | <b>PD</b>     | <b>PU</b> |
| <b>Matriz confusão</b> | <b>PD</b> | 10                             | -         | 9               | 1         | 10           | -         | 10            | -         | 10            | -         |
|                        | <b>PU</b> | 19                             | 21        | 14              | 26        | 10           | 30        | 18            | 22        | 17            | 23        |
| Precisão               |           | 0.3448                         |           | 0.3913          |           | 0.5000       |           | 0.3571        |           | 0.3704        |           |
| Sensibilidade          |           | 1.0000                         |           | 0.9000          |           | 1.0000       |           | 1.0000        |           | 1.0000        |           |
| Especificidade         |           | 0.5250                         |           | 0.6500          |           | 0.7500       |           | 0.5500        |           | 0.5750        |           |
| TCC (%)                |           | 62                             |           | 70              |           | 80           |           | 64            |           | 66            |           |

O modelo PLS-DA construído utilizando o canal G obteve o melhor resultado de taxa de classificação correta e especificidade, com valor de 80% e 0.7500, respectivamente. O melhor resultado de precisão foi 0.5000 para o canal G, que classificou menos amostras como falso positivo. Essas 10 amostras classificadas de forma errada, têm concentração entre 1 e 5% de adulterante (Tabela IV-11).

Na Figura IV-24 (a e b) são apresentados os scores de Fisher obtidos pelos modelos SPA-LDA (figuras à esquerda) e GA-LDA (figuras à direita) construídos com histograma completo, Figura IV-24 (c e d) canal RGB, Figura IV-24 (e e f) a junção dos canais RGB com C e na Figura IV-24 (g e h) o canal G.



Figura IV-24 - Gráfico de scores obtido por SPA-LDA e GA-LDA para as amostras de páprica doce não adulteradas e adulterada com Urucum



● Não adulteradas, ● Adulteradas; ●: Amostras de treinamento, ■: Amostras de teste



Quando o modelo SPA-LDA foi construído com o histograma completo (Figura IV-24a), nenhuma amostra da classe não adulteradas foi classificada como adulterada tanto no conjunto de treinamento, quanto no conjunto teste. No entanto, 17 amostras do conjunto de teste foram classificadas falso positivo, levando o modelo a uma precisão de 0,3704 (Tabela IV-13).

O canal RGB (Figura IV-24c) utilizando SPA-LDA apresentou uma precisão de 0.3913 (Tabela IV-13) para a classe PD no conjunto teste, em que 14 amostras foram classificadas como falso positivo. O canal G classificou duas amostras falso positivo e sete amostras como falso negativo. A matriz confusão do conjunto de teste dos modelos SPA-LDA e GA-LDA são apresentados na Tabela IV-12, mostrando mais detalhadamente a quantidade de amostras classificadas incorretamente pelos modelos.

Tabela IV-12 - Matriz de confusão obtida para os modelos SPA-LDA e GA-LDA construídos com histograma de cores das amostras de páprica e páprica com urucum

| Canal de Cor               |           | SPA-LDA   |           | GA-LDA    |           |
|----------------------------|-----------|-----------|-----------|-----------|-----------|
|                            |           | PD        | PU        | PD        | PU        |
| <b>Histograma completo</b> | <b>PD</b> | 10        | -         | 9         | 1         |
|                            | <b>PU</b> | 17        | 23        | 11        | 29        |
| Variáveis Sel.             |           | 13        |           | 11        |           |
|                            |           | <b>PD</b> | <b>PU</b> | <b>PD</b> | <b>PU</b> |
| <b>RGB</b>                 | <b>PD</b> | 9         | 1         | 9         | 1         |
|                            | <b>PU</b> | 14        | 26        | 13        | 27        |
| Variáveis Sel.             |           | 21        |           | 13        |           |
|                            |           | <b>PD</b> | <b>PU</b> | <b>PD</b> | <b>PU</b> |
| <b>C + RGB</b>             | <b>PD</b> | 8         | 2         | 10        | -         |
|                            | <b>PU</b> | 13        | 27        | 15        | 25        |
| Variáveis Sel.             |           | 20        |           | 9         |           |
|                            |           | <b>PD</b> | <b>PU</b> | <b>PD</b> | <b>PU</b> |
| <b>G</b>                   | <b>PD</b> | 8         | 2         | 8         | 2         |
|                            | <b>PU</b> | 7         | 33        | 4         | 36        |
| Variáveis Sel.             |           | 15        |           | 10        |           |

Os parâmetros de desempenho dos modelos SPA-LDA e GA-LDA construídos com o histograma de cores obtidos das imagens de páprica doce não adulteradas e adulterada com urucum podem ser encontrados na Tabela IV-13.



Tabela IV-13 - Parâmetros de desempenho dos modelos SPA-LDA e GA-LDA para o conjunto de teste

|                            |                | <b>Precisão</b> | <b>Sensibilidade</b> | <b>Especificidade</b> | <b>TCC (%)</b> |
|----------------------------|----------------|-----------------|----------------------|-----------------------|----------------|
| <b>Histograma completo</b> | <b>SPA-LDA</b> | 0.3704          | 1.0000               | 0.5750                | 66             |
|                            | <b>GA-LDA</b>  | 0.4500          | 0.9000               | 0.7250                | 76             |
| <b>RGB</b>                 | <b>SPA-LDA</b> | 0.3913          | 0.9000               | 0.6500                | 70             |
|                            | <b>GA-LDA</b>  | 0.4091          | 0.9000               | 0.6750                | 72             |
| <b>C + RGB</b>             | <b>SPA-LDA</b> | 0.3810          | 0.8000               | 0.6750                | 70             |
|                            | <b>GA-LDA</b>  | 0.4000          | 1.0000               | 0.6250                | 72             |
| <b>G</b>                   | <b>SPA-LDA</b> | 0.5333          | 0.8000               | 0.8250                | 82             |
|                            | <b>GA-LDA</b>  | 0.6667          | 0.8000               | 0.9000                | 88             |

Os resultados evidenciaram que o modelo GA-LDA construído com o canal G apresenta a maior precisão (0.6667). Esse canal se destaca com uma especificidade de 0.9000, indicando um bom desempenho em identificar casos negativos. A precisão da classe PD para os modelos SPA-LDA e GA-LDA construídos com o histograma completo, RGB e C + RGB foi inferior a 0.5000. De forma geral, o canal verde (G) apresentou os melhores resultados, com uma taxa de classificação correta de 82% para os modelos SPA e 88% para o modelo GA.

Vale destacar que para todos os modelos construídos, tanto para o conjunto de treinamento, quanto para o conjunto de teste, as amostras da classe PU classificadas como sendo da classe PD, apresentam concentração de 1 e 5%, sendo 1% a maioria. Isso se refletiu em menores valores de especificidade para quase todos os modelos construídos.



## CAPÍTULO V : CONCLUSÃO

---



Neste presente estudo, foram desenvolvidas metodologias baseadas em espectrometria no infravermelho NIR e de imagens digitais para a detecção de adulteração em pprica doce com adulterantes, amido ou com urucum.

Foram construdos modelos de classificao SPA-LDA, GA-LDA e PLS-DA utilizando os espectros NIR das amostras no adulteradas e das adulteradas, sendo duas classes para cada tipo de adulterantes. Em que a classe PD foram as amostras no adulteradas e a classe PA e PU as amostras adulteradas com amido e urucum, respectivamente.

Os modelos de classificao construdos com os espectros NIR para a identificao de pprica doce e pprica doce adulterada com amido apresentaram timos resultados, demonstrando o elevado potencial desta tcnica para detectar esse tipo de fraude. Todos os modelos construdos apresentaram 100% de taxa de classificao correta, tanto para o conjunto de treinamento, quanto para o conjunto de teste.

Para os modelos construdos para deteco do urucum como adulterante, os dados NIR apresentaram resultados satisfatrios quando o SPA-LDA foi aplicada aos dados sem pr-processamento e com primeira derivada de Savitzky–Golay com janela de 11 pontos. J quando o GA-LDA foi utilizado, os melhores resultados foram com correo de linha de base *offset*, primeira e segunda derivada de Savitzky–Golay com janela de 11 pontos. Todos com TCC de 100%.

Quando o processamento de imagens digitais foi utilizado para identificar a adulterao de pprica doce com amido, o melhor resultado entre os modelos PLS-DA foi apresentado pelo canal verde (G). Para este canal, o conjunto teste obteve preciso de 0.8889, tendo classificado corretamente 94% das amostras.

O modelo SPA-LDA com o canal G apresentou o melhor resultado de preciso e especificidade, com valores de 0,8889 e 0,9750, respectivamente, e taxa de classificao correta de 94% das amostras do conjunto de teste. J para o GA-LDA, o melhor resultado foi encontrado quando a combinao RGB + HSI foi utilizada, com taxa de classificao correta de 92%.

Quando o modelo PLS-DA foi construdo para a identificao da adulterao com urucum, utilizando o processamento de imagens digitais, o canal G obteve o melhor resultado de TCC e especificidade, com valor de 0.8000 e 0.7500, respectivamente. De forma geral, para os modelos SPA-LDA e GA-LDA, o canal verde (G) apresentou os melhores resultados, com uma taxa de classificao correta de 82% e 88%, respectivamente.



Por fim, mediante os resultados encontrados nessa pesquisa, é possível concluir que a aplicação da técnica NIR, associada as ferramentas quimiométricas de classificação, é uma alternativa promissora para identificação de adulteração de páprica doce com amido ou com urucum. Para os dados de imagens digitais, foi notado que para o amido como adulterantes, as amostras adulteradas classificadas erradas tinham concentração de 1 e 5% (m/ m). Já quando o urucum foi utilizado como adulterante, a maioria das amostras adulteradas que foram classificadas como não adulteradas tinha concentração de 1% (m/ m).



## PERSPECTIVAS

Durante o desenvolvimento deste trabalho, identificaram-se algumas possibilidades de estudos que podem ser considerados para futuras pesquisas. Nesse sentido, ressaltamos as seguintes propostas:

- Construir e disponibilizar um banco de dados;
- Aplicar o método NIR proposto para avaliar amostras comerciais;
- Utilizar o método de imagens digitais proposto para analisar amostras comerciais;
- Desenvolver modelos de classificação com base em imagens digitais, excluindo amostras com concentração de 1% (m/ m);
- Aplicar os novos modelos construídos para avaliar amostras comerciais.



## REFERÊNCIAS

- ACHARYA, T., RAY, A. K. Image Processing- Principles and Applications. **John Wiley & Sons, Inc.** 2005.
- ARAÚJO, A.; MARINHO, W.; DE ARAÚJO GOMES, A. A fast and inexpensive chemometric-assisted method to identify adulteration in acai (*Euterpe oleracea*) using digital images. **Food Analytical Methods**, v. 11, p. 1920-1926, 2018.
- ARAÚJO, M. C. U., et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. **Chemometrics and Intelligent Laboratory Systems**. 57: 65, 2001.
- BANTI, M. Food adulteration and some methods of detection, review. **International Journal of Nutrition and Food Sciences**, v. 9, n. 3, p. 86-94, 2020.
- BARROS NETO, B; SCARMINIO, I. S.; BRUNS, R. E. 25 anos de quimiometria no Brasil. **Química Nova**, v. 29, p. 1401-1406, 2006.
- BEĆ, K. B.; GRABSKA, J.; HUCK, C. W. Principles and applications of miniaturized near-infrared (NIR) spectrometers. **Chemistry—A European Journal**, v. 27, n. 5, p. 1514-1532, 2021.
- BLACK, C. et al. A comprehensive strategy to detect the fraudulent adulteration of herbs: The oregano approach. **Food Chemistry**, v. 210, p. 551-557, 2016.
- BRASIL, RDC nº 276, de setembro de 2005. **REGULAMENTO TÉCNICO PARA ESPECIARIAS, TEMPEROS E MOLHOS. Diário oficial República Federativa do Brasil.** Brasília, DF, 23 set. 2005.
- BRASIL. Lei nº 7.209, de 07 de dezembro de 1940. Código Penal. **Diário Oficial da União**, Rio de Janeiro, 31 dez. 1940.
- BRERETON, R. G. Chemometrics: data analysis for the laboratory and chemical plant. **John Wiley & Sons**, 2003.
- BURNS, D. A.; CIURCZAK, Emil W. (Ed.). Handbook of near-infrared analysis. **CRC press**, ISBN-3: 978-0-8493-7393-0, p. 12, 2007.
- CARDIN, L. M. A. G. et al. Spices and their importance in gastronomy. **Brazilian Journal of Development**, v. 7, n. 10, p. 97282-97291, 2021.
- DE LIMA, A. B. S. et al. Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling. **Food Control**, v. 107, p. 106802, 2020.
- DE OLIVEIRA, J. M. F.; LUZ, F. J. F. Orientações para o cultivo do pimentão em Roraima. **Empresa**. 1998.
- DE QUEIROZ, J. E. R.; GOMES, H. M. Introdução ao processamento digital de imagens. **Rita**, v. 13, n. 2, p. 11-42, 2006.



DEINEKO, Z., ZELENIIY O., LYASHENKO V., TABAKOVA, I. Color space image as a factor in the choice of its processing technology. In: **Problems Of Modern Science and Practice**. 2021. p. 389-394.

DUNTEMAN, G. H. *Principal Components Analysis*; Sage: Newbury Park, CA, 1989.

FOOD AND DRUG ADMINISTRATION, **Economically Motivated Adulteration** (Food Fraud), 2024. <https://www.fda.gov/food/compliance-enforcement-food/economically-motivated-adulteration-food-fraud>. Acesso em: 10 ago 2024.

FERREIRA, M. M. C. *Quimiometria: conceitos, métodos e aplicações*. **Editora da UNICAMP**, 2015.

FORSYTH, J E. et al. Turmeric means “yellow” in Bengali: Lead chromate pigments added to turmeric threaten public health across Bangladesh. **Environmental research**, v. 179, p. 108722, 2019.

FRANÇA, V. F. **Aplicação de imagens digitais e técnicas quimiométricas para detecção de adulteração em mel**. 2020. TCC (Trabalho de Conclusão de Curso) – Universidade Federal da Paraíba. 2020.

GALVIN-KING, P. et al. Herb and spice fraud; the drivers, challenges and detection. **Food Control**, v. 88, p. 85-97, 2018.

GONZALEZ, R. C., WOODS, R. E. **Processamento digital de imagem**. Pearson, ISBN-10: 8576054019, v. 10, p. 11-27, 2010.

JOHNSON, R. Food fraud and economically motivated adulteration of food and food ingredients. In: **Congressional Research Service**. 2014.

KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**, v. 11, n. 1, p. 137-148, 1969.

KUCHARSKA-AMBROŻEJ, K; KARPINSKA, J. The application of spectroscopic techniques in combination with chemometrics for detection adulteration of some herbs and spices. **Microchemical Journal**, v. 153, p. 104278, 2020.

KUMAR, N. et al. Chemometrics tools used in analytical chemistry: An overview. **Talanta**, v. 123, p. 186-199, 2014.

LAGUNA, L. E. et al. Queijo de cabra maturado adicionado de pasta de pimentão. **Embrapa**. 2011.

LEE, L. C.; LIONG, C. Y.; JEMAIN, A. A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. **Analyst**, v. 143, n. 15, p. 3526-3539, 2018.

LEMOS, A. R. **Caracterização físico-química, bioquímica e avaliação da atividade antioxidante em genótipos de Urucueiros (Bixa orellana L.)**. Universidade Estadual do Sudoeste da Bahia, Itapetinga, 2008.



LIMA, D. M. et al. Tabela brasileira de composição de alimentos - TACO: 4. ed. Campinas: **NEPA-UNICAMP**, 2011. 113 p.

LIMA, K. M. G. et al. Sensores ópticos com detecção no infravermelho próximo e médio. **Química Nova**, v. 32, n. 6, p. 1635-1643, 2009.

MACHADO, A. W., SOUKI, B. Q. Simplificando a obtenção e a utilização de imagens digitais: scanners e câmeras digitais. **Revista Dental Press de Ortodontia e Ortopedia Facial**, v. 9, p. 133-156, 2004.

MASSART, D. L.; VANDEGINSTE, B. G.; BUYDENS, L. M.; LEWI P. J.; JONG, S. D.; SMEYERS-VERBEKE, J. Handbook of chemometrics and qualimetrics: Part A. **Elsevier Science Amsterdam**, 1998.

MATHEW, T. V. Genetic algorithm. **Report submitted at IIT Bombay**, v. 53, 2012.

MELO, M. F. et al. Pimentão: fonte de vitaminas. **Embrapa**. 2009.

MENESES, M. P. "Especiarias", Dicionário Alice, 2019. Disponível em: [https://alice.ces.uc.pt/dictionary/?id=23838&pag=23918&id\\_lingua=1&entry=24279](https://alice.ces.uc.pt/dictionary/?id=23838&pag=23918&id_lingua=1&entry=24279). Acesso em: 25 mai. 2023.

MOORE J. C.; SPINK, J.; LIPP, M. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. **J Food Sci**. 2012, v. 77, n. 4, p. 118-126.

MARQUES FILHO, O.; VIEIRA NETO, H. Processamento Digital de Imagens, Rio de Janeiro: **Brasport**, 1999. ISBN 8574520098.

OSMAN, A. G. et al. Overview of analytical tools for the identification of adulterants in commonly traded herbs and spices. **Journal of AOAC International**, v. 102, n. 2, p. 376-385, 2019.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. **Analytica chimica acta**, v. 1026, p. 8-36, 2018.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian chemical society**, v. 14, p. 198-219, 2003.

PAULI-YAMADA, L. F. et al. **Estudo microscópico de páprica (*Capsicum annum* L.): detecção de fraudes e matérias estranhas**. Vigilância Sanitária em Debate: Sociedade, Ciência & Tecnologia (Health Surveillance under Debate: Society, Science & Technology) – Visa em Debate, v. 9, n. 1, p. 123-128, 2021.

PAVIA, D. L. et al. Introdução à espectroscopia. **Cengage Learning**, 2021.

PEREIRA, E. V. et al. Goat milk authentication by one-class classification of digital image-based fingerprint signatures: Detection of adulteration with cow milk. **Microchemical Journal**, v. 180, p. 107640, 2022.

**PL 7.664/2017**. Projeto de Lei Dispõe sobre o agravamento de penas dos crimes de fraude, falsificação e adulteração de alimentos e bebidas destinados a consumo humano, mediante



alterações no Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal). 2017. Disponível em: [https://www.camara.leg.br/proposicoesWeb/prop\\_mostrarintegra?codteor=1564518&filename=Avulso%20PL%207664/2017#:~:text=Disp%C3%B5e%20sobre%20o%20agravamento%20de,\(Lei%20dos%20Crimes%20Hediondos\)](https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1564518&filename=Avulso%20PL%207664/2017#:~:text=Disp%C3%B5e%20sobre%20o%20agravamento%20de,(Lei%20dos%20Crimes%20Hediondos)). Acesso em: 04 de outubro de 2024.

PONTES, M. J. C. **Algoritmo das Projeções Sucessivas para seleção de variáveis espectrais em problemas de classificação**. João Pessoa, Programa de Pós-Graduação em Química, UFPB, 2009, Tese de doutorado, 144p.

PONTES, M. J. C., et al. The successive projections algorithm for spectral variable selection in classification problems. **Chemometrics and Intelligent Laboratory Systems**. 78: 11, 2005.

RING, E. F. J. The discovery of infrared radiation in 1800. **The Imaging Science Journal**, v. 48, n. 1, p. 1-8, 2000.

RIOS, C. A., BREITKREITZ, M. C. Determinação do teor de azeite de oliva extravirgem presente em misturas adulteradas com óleo de soja usando imagens digitais e ferramentas quimiométricas. In: **Simpósio Latino-Americano de Ciências de Alimentos**, XIV., 2021. Proceedings of 14 SLACA - Latin American Symposium on Food Science. 2021.

RODRIGUES, D. O. B. F. **Amplificação da fraude alimentar por média impressos nacionais: Estudo de caso**. Universidade do Porto (Portugal), Departamento de Geociências, Ambiente e Ordenamento do Território. 2017. Dissertação de Mestrado, 96p.

SERGYÁN, S. Color content-based image classification. In: **5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics**. 2007. p. 427-434.

SILVA, L. M. M. P. **Fraude alimentar: Reconhecer a sua existência através da aplicação de critérios operacionais**. 2018. Dissertação (Mestrado em Tecnologias de Produção e Transformação Agro-Industrial) - Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal.

SKOOG, D. A.; HOLLER J. F.; NIEMAN, A. T. **Princípios da Análise Instrumental**. 5ed. Bookman, 2002.

SOARES, S. F. C. **Um novo critério para seleção de variáveis usando o Algoritmo das Projeções Sucessivas**. João Pessoa, Programa de Pós-Graduação em Química, UFPB, 2010, Dissertação de mestrado, 108p.

STEVEN D. B., TAULER R., WALCZAK, B.: Comprehensive chemometrics. Chemical and biochemical data analysis. **Analytical and Bioanalytical Chemistry**, v. 396, n. 2, p. 121-230, 2010.

TORRENT, J. O. S. E.; BARRÓN, V. Diffuse reflectance spectroscopy. **Methods of Soil Analysis Part 5—Mineralogical Methods**, v. 5, p. 367-385, 2008.

VALÉRIO, M. A. et al. Annatto seed residue (*Bixa orellana* L.): nutritional quality. **Food Science and Technology** (Campinas), v. 35, n. 2, p. 326-330, 2015.



VILAR, W. T. S. **Desenvolvimento de metodologias analíticas baseadas em HPLC-DAD, espectroscopia no infravermelho e quimiometria para o controle de qualidade de repelentes de insetos.** João Pessoa, Programa de Pós-Graduação em Química, UFPB, 2018., Tese de doutorado, 185p.

WIELOGORSKA, E. et al. Development of a comprehensive analytical platform for the detection and quantitation of food fraud using a biomarker approach. The oregano adulteration case study. **Food chemistry**, v. 239, p. 32-39, 2018.

XANTHOPOULOS, P., PARDALOS, P. M., TRAFALIS, T. B. Linear discriminant analysis. In: **Robust data mining**, Springer, New York, 2013. p. 27-33.

