

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

DISSERTAÇÃO DE MESTRADO

EMPREGO DE ESPECTROSCOPIA NO INFRAVERMELHO
MÉDIO E TÉCNICAS QUIMIOMÉTRICAS EM ESTUDOS DE
DOCUMENTOSCOPIA E DATILOSCOPIA PARA FINS FORENSES

VICTÓRIA ZAGNA DOS SANTOS

Porto Alegre, Abril de 2024.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

VICTÓRIA ZAGNA DOS SANTOS

**EMPREGO DE ESPECTROSCOPIA NO INFRAVERMELHO MÉDIO E
TÉCNICAS QUIMIOMÉTRICAS EM ESTUDOS DE DOCUMENTOSCOPIA E
DATILOSCOPIA PARA FINS FORENSES**

Dissertação apresentada como
requisito parcial para a obtenção
do grau de Mestre em Química

Prof. Dr. Marco Flôres Ferrão
Orientador

Porto Alegre, Abril de 2024.

CIP - Catalogação na Publicação

Santos, Victória Zagna dos
EMPREGO DE ESPECTROSCOPIA NO INFRAVERMELHO MÉDIO E
TÉCNICAS QUIMIOMÉTRICAS EM ESTUDOS DE DOCUMENTOSCOPIA
E DATILOSCOPIA PARA FINS FORENSES / Victória Zagna dos
Santos. -- 2024.
85 f.
Orientador: Marco Flôres Ferrão.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Instituto de Química, Programa de
Pós-Graduação em Química, Porto Alegre, BR-RS, 2024.

1. Quimiometria. 2. Espectroscopia no
infravermelho. 3. ATR-FTIR. 4. Química Forense. 5.
Seleção de variáveis. I. Ferrão, Marco Flôres, orient.
II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com
os dados fornecidos pelo(a) autor(a).

Agradecimentos

Ao meu orientador, Prof. Ferrão, pelo acolhimento, pela paciência e pelos ensinamentos.

Ao PPGQ, IQ, UFRGS e Polícia Federal-RS pela oportunidade de estudo e de desenvolvimento desse trabalho.

À CAPES pelo fomento ao PPGQ.

Ao Marco Antônio, pela colaboração.

À Kris, que foi grande parceira nessa jornada, sempre alegre e atenciosa.

À minha mãe, pelo amor sempre a mim dedicado e pelo grande exemplo de mulher guerreira que é. Te amo!

À minha família, pelo apoio e compreensão nos momentos em que estive ausente.

Ao meu amor, Rafa, pela parceria, carinho e todo suporte emocional que foi fundamental nessa jornada.

À Amorzinha, por estar sempre ao meu lado (literalmente), sendo uma grande 'cãopanheira'.

Meu muito obrigada!

“Os homens fazem sua própria história, mas não a fazem como querem; não a fazem sob circunstâncias de sua escolha, e sim sob aquelas com que se defrontam diretamente, legadas e transmitidas pelo passado”.

Karl Marx

Resumo

A constante evolução das ciências forenses conta com o desenvolvimento e aprimoramento de métodos investigativos, bem como com profissionais especializados em diversas áreas de aplicação. As amostras a serem investigadas podem ser macro ou microscópicas, tendo como característica em comum a preferência por metodologias que sejam rápidas e não-destrutivas. O presente trabalho teve como objetivo estudar o emprego de espectroscopia no infravermelho médio com transformada de Fourier por reflexão total atenuada (ATR-FTIR) associado a quimiometria em diferentes aplicações forenses. Num primeiro estudo, espectros de traços de 3 canetas esferográficas (sendo 2 de mesma marca e cor diferente e uma de marca diferente) aplicados em 4 papéis (que diferenciaram em gramatura, nacionalidade e matriz), foram pré-processados e diferenciados por PCA, LDA e PLS-DA, sendo na LDA empregado dois métodos de seleção de variáveis, GA e SPA. Os modelos não-supervisionados e supervisionados separaram com sucesso as 3 canetas nos papéis brancos e demonstraram elevada contribuição do papel nos espectros. Para os modelos supervisionados as figuras de mérito ficaram entre 80 e 100%. Um segundo estudo foi feito com amostras de impressões digitais com o objetivo de diferenciar tempo e condição de armazenamento. Foram obtidos espectros de impressões digitais latentes de 19 pessoas do sexo masculino, aos quais foi aplicado HCA para selecionar os mais semelhantes entre as triplicatas e após foram pré-processados. A PCA aplicada aos dados demonstrou que há significativa preservação nas amostras armazenadas ao abrigo da luz entre 0 e 7 dias, mas que há uma tendência em atingir similar padrão de envelhecimento em 30 dias. Na modelagem por LDA foi avaliado o desempenho de quatro métodos de seleção de variáveis, GA, AOC, SW e SPA, resultando todos em excelentes figuras de mérito, assim como modelo PLS-DA. Portanto, ATR-FTIR associada a técnicas quimiométricas demonstra grande potencial para aplicação em casos forenses, sendo uma metodologia eficaz, rápida, de baixo custo e com preservação da integridade amostral.

Palavras-chave: quimiometria, espectroscopia no infravermelho, ATR-FTIR, química forense, seleção de variáveis.

Abstract

The constant evolution of forensic sciences relies on the development and improvement of investigative methods, as well as professionals specialized in different areas of application. The samples to be investigated can be macro or microscopic, with the common characteristic being the preference for methodologies that are fast and non-destructive. The present work aimed to study the use of mid-infrared spectroscopy with attenuated total reflection Fourier transform (ATR-FTIR) associated with chemometrics applied in different forensic areas. In a first study, trace spectra of 3 ballpoint pens (2 of the same brand and different color and one of a different brand) applied to 4 papers (which differed in weight, nationality and matrix), were pre-processed and differentiated by PCA, LDA and PLS-DA, with LDA using two variable selection methods, GA and SPA. The unsupervised and supervised models successfully separated the 3 pens on white paper and demonstrated a high contribution of the paper to the spectra. For the supervised models, the figures of merit were between 80 and 100%. A second study was carried out with fingerprint samples with the aim of differentiating storage time and condition. Latent fingerprint spectra were obtained from 19 males, to which HCA was applied to select the most similar among the triplicates and then pre-processed. PCA applied to the data demonstrated that there is significant preservation in samples stored away from light between 0 and 7 days, but that there is a tendency to reach a similar aging pattern within 30 days. In LDA modeling, the performance of four variable selection methods, GA, AOC, SW and SPA, was evaluated, all resulting in excellent figures of merit, as well as the PLS-DA model. Therefore, ATR-FTIR associated with chemometric techniques demonstrates great potential for application in forensic cases, being an effective, fast, low-cost methodology that preserves sample integrity.

Keywords: chemometrics, infrared spectroscopy, ATR-FTIR, forensic chemistry, variable selection.

Lista de Figuras

Figura 1. Regiões do espectro eletromagnético com destaque para a faixa do infravermelho. Fonte: John, D.K. 2020 ³²	19
Figura 2. Modos de vibrações moleculares. Fonte: Santos, C.S. 2017 ⁶ . . 20	
Figura 3. Ilustração representativa da análise por ATR. Fonte: John, D.K. 2020 ³²	21
Figura 4. Exemplo de reagrupamento de objetos com LDA. Fonte: Junges, C.H. 2023 ³⁹	26
Figura 5. Esquema da etapa de treinamento do PLS-DA.	27
Figura 6. Fluxograma das etapas do algoritmo genético para seleção de variáveis. Fonte: Junges, C.H. 2023 ³⁹	29
Figura 7. Exemplo do comportamento das formigas antes (a) e depois (b) de otimizarem o caminho entre o ninho e o alimento. Fonte: Sanca, H.S. 2013 ⁵⁶	30
Figura 8. Matriz de confusão de duas classes. V=verdadeiro, F=falso, P=positivo e N=negativo	32
Figura 9. Espectros sem tratamento de ATR-FTIR das amostras de traço de caneta.	37
Figura 10. Espectros de ATR-FTIR após os pré-processamentos. a) 75Br b) 63Br c) 75Arg e d) 75Rec.	38
Figura 11. Gráfico dos escores PC2xPC3 da PCA com os dados obtidos sob o papel 75Br.	39
Figura 12. Gráfico dos escores PC1xPC3 da PCA com os dados obtidos sob o papel 63Br.	40
Figura 13. Gráfico dos escores PC1xPC2 da PCA com os dados obtidos sob o papel 75Arg.	40
Figura 14. Gráfico dos escores PC2xPC3 da PCA com os dados obtidos sob o papel 75Rec.	41
Figura 15. Gráfico dos escores PC1xPC2 da PCA com os dados obtidos sob os papéis brancos.	42
Figura 16. Gráfico dos escores PC3xPC4 da PCA com os dados obtidos sob os papéis brancos.	42

Figura 17. Gráfico dos escores PC1xPC3 da PCA com os dados obtidos sob os papéis nacionais.	43
Figura 18. Gráfico dos escores PC2xPC4 da PCA com os dados obtidos sob os papéis nacionais.	44
Figura 19. Gráfico dos escores PC1xPC4 da PCA com os dados obtidos sob os papéis de mesma gramatura.	45
Figura 20. Gráfico dos escores PC2xPC3 da PCA com os dados obtidos sob os papéis de mesma gramatura.	45
Figura 21. Gráfico dos escores FD1xFD2 da SPA-LDA com os dados classificados pelas tintas de caneta. ● corresponde as amostras de treinamento e × as amostras de teste.	48
Figura 22. Dados classificados pelas tintas de caneta. a) Seleção das variáveis mais relevantes do modelo SPA-LDA; b) Gráfico de pesos das funções discriminantes 1 (preto) e 2 (azul) da SPA-LDA.	48
Figura 23. Gráfico dos escores FD1xFD2 do modelo GA-LDA com os dados classificados pelas tintas de caneta. ● corresponde as amostras de treinamento e × as amostras de teste.	49
Figura 24. Dados classificados pelas tintas de caneta. a) Seleção das variáveis mais relevantes do modelo GA-LDA; b) Gráfico de pesos das funções discriminantes 1 (preto) e 2 (azul) da GA-LDA.	50
Figura 25. Gráfico dos escores FD1xFD3 do modelo SPA-LDA com os dados classificados pelos papéis. ● corresponde as amostras de treinamento e × as amostras de teste.	52
Figura 26. Dados classificados pelos papéis. a) Seleção das variáveis mais relevantes do modelo SPA-LDA; b) Gráfico de pesos das funções discriminantes 1 (preto) e 3 (azul) da SPA-LDA.	52
Figura 27. Gráfico dos escores FD1xFD3 do modelo GA-LDA com os dados classificados pelos papéis. ● corresponde as amostras de treinamento e × as amostras de teste.	53
Figura 28. Gráfico dos escores FD1xFD2 do modelo GA-LDA com os dados classificados pelos papéis. ● corresponde as amostras de treinamento e × as amostras de teste.	54

Figura 29. Dados classificados pelos papéis. a) Seleção das variáveis mais relevantes do modelo GA-LDA; b) Gráfico de pesos das funções discriminantes 1 (preto), 2 (azul) e 3 (rosa) da GA-LDA.	54
Figura 30. Classificação das amostras por PLS-DA considerando as canetas como classes. ● corresponde as amostras de treinamento e ✕ as amostras de teste.....	56
Figura 31. Classificação das amostras por PLS-DA considerando os papéis como classes. ● corresponde as amostras de treinamento e ✕ as amostras de teste.....	57
Figura 32. Espectro FTIR das amostras sem pré-processamento. C=Claro, amostras expostas à luz; E=Escuro, amostras armazenadas ao abrigo da luz; D0=Dia da coleta; D7=7 dias após coleta; D30=30 dias após coleta. ..	59
Figura 33. Dendrograma das amostras pertencentes ao grupo CD0.....	60
Figura 34. Espectro FTIR dos dados após recorte, suavização, normalização e derivada.	61
Figura 35. Gráfico dos escores PC1xPC2 da PCA.	62
Figura 36. Gráfico de pesos da segunda componente principal da PCA.	63
Figura 37. Variáveis selecionadas pelo modelo SPA-LDA.	65
Figura 38. Espectro de uma amostra do tempo 0 sem tratamento, recortado na região de interesse e com indicativo das bandas com informações químicas mais relevantes para o estudo.	66
Figura 39. Gráfico dos escores FD1XFD2 da SPA-LDA. ● corresponde as amostras de treinamento e ✕ as amostras de teste.....	67
Figura 40. Gráfico dos escores FD3XFD4 da SPA-LDA. ● corresponde as amostras de treinamento e ✕ as amostras de teste.....	68
Figura 41. Gráfico de pesos das funções discriminantes 1 (preto), 2 (azul), 3 (rosa) e 4 (verde) da SPA-LDA.	69
Figura 42. Respostas do modelo PLS-DA por classe. ● corresponde as amostras de treinamento e ✕ as amostras de teste.....	71
Figura 43. Variáveis selecionadas modelo GA-LDA.	83
Figura 44. Variáveis selecionadas modelo AOC-LDA.....	84
Figura 45. Variáveis selecionadas modelo SW-LDA.....	84

Lista de Tabelas

Tabela 1. Regiões espectrais do infravermelho	19
Tabela 2. Figuras de mérito LDA grupo canetas.....	46
Tabela 3. Figuras de mérito LDA grupo papéis.....	51
Tabela 4. Figuras de mérito PLS-DA	55
Tabela 5. Figuras de mérito dos modelos LDA	64
Tabela 6. Figuras de mérito PLS-DA	70

Lista de Siglas

ACO	Algoritmo de Otimização da Colônia de Formigas (do inglês, “ <i>Ant Colony Optimization</i> ”)
ATR	Reflexão Total Atenuada (do inglês, “ <i>Attenuated Total Reflectance</i> ”)
ATR-FTIR	Espectroscopia de Infravermelho Médio com Transformada de Fourier por Reflexão Total Atenuada (do inglês, “ <i>Fourier Transform Infrared Spectroscopy with Attenuated Total Reflectance</i> ”)
FD	Função Discriminante
FIR	Infravermelho Distante (do inglês “ <i>Far Infrared</i> ”)
FTIR	Espectroscopia no Infravermelho Médio com Transformada de Fourier (do inglês, “ <i>Fourier Transform Infrared Spectroscopy</i> ”)
GA	Algoritmo Genético (do inglês, “ <i>Genetic Algorithm</i> ”)
HCA	Análise de Agrupamentos por Métodos Hierárquicos (do inglês “ <i>Hierarchical Cluster Analysis</i> ”)
IR	Infravermelho (do inglês “ <i>Infrared</i> ”)
LDA	Análise Discriminante Linear (do inglês “ <i>Linear Discriminant Analysis</i> ”)
MIR	Infravermelho Médio (do inglês “ <i>Mid Infrared</i> ”)
NIR	Infravermelho Próximo (do inglês “ <i>Near Infrared</i> ”)
PC	Componente Principal (do inglês “ <i>Principal Component</i> ”)
PCA	Análise por Componentes Principais (do inglês “ <i>Principal Component Analysis</i> ”)
PLS-DA	Análise Discriminante por Mínimos Quadrados Parciais (do inglês “ <i>Partial Least Squares Discriminant Analysis</i> ”)
SPA	Algoritmo das Projeções Sucessivas (do inglês, “ <i>Successive Projections Algorithm</i> ”)
SW	Algoritmo <i>Stepwise</i>

Sumário

1. Introdução	14
2. Revisão bibliográfica	16
2.1 Ciências forenses	16
2.1.1 Documentoscopia.....	17
2.1.2 Datiloscopia.....	18
2.2 Espectroscopia no infravermelho.....	19
2.2.1 Refletância Total Atenuada (ATR).....	21
2.3 Quimiometria	21
2.3.1 Pré-processamento de dados	22
2.3.2 Análise de componentes principais (PCA)	23
2.3.3 Análise de agrupamentos por métodos hierárquicos (HCA)	24
2.3.4 Partição do grupo amostral em subconjuntos	25
2.3.5 Análise discriminante linear (LDA)	25
2.3.6 Análise discriminante por mínimos quadrados parciais (PLS-DA)	27
2.3.7 Técnicas de seleção de variáveis.....	28
2.3.8 Figuras de mérito	32
3. Objetivos	34
3.1 Objetivo geral	34
3.2 Objetivos específicos.....	34
3.2.1 Desenvolver e validar modelos de análise multivariada para identificação de tintas de canetas esferográficas em diferentes papéis: ...	34
3.2.2 Avaliar a preservação e degradação de impressões digitais ao longo do tempo:	34
3.2.3 Aplicar e avaliar técnicas de seleção de variáveis em modelagem quimiométrica:.....	35
3.2.4 Implementação de ATR-FTIR como metodologia padrão em análises forenses:	35
4. Diferenciação de canetas esferográficas em diferentes papéis	36
4.1 Materiais e métodos	36
4.1.1 Amostras	36
4.1.2 Espectroscopia de infravermelho médio	36
4.1.3 Pré-processamentos e análise multivariada.....	36

4.2	Resultados e discussão.....	37
4.2.1	Análise de componentes principais.....	38
4.2.2	Análise discriminante linear.....	46
4.2.3	Análise discriminante por mínimos quadrados parciais.....	55
5.	Envelhecimento de impressões digitais latentes.....	58
5.1	Materiais e métodos.....	58
5.1.1	Amostras.....	58
5.1.2	Espectroscopia de infravermelho médio.....	58
5.1.3	Pré-processamentos e análise multivariada.....	58
5.2	Resultados e discussão.....	59
5.2.1	Análise de componentes principais.....	61
5.2.2	Análise discriminante linear.....	63
5.2.3	Análise discriminante por mínimos quadrados parciais.....	69
	Conclusões.....	72
6.	Referências bibliográficas.....	73
7.	Anexos.....	83
7.1	Variáveis selecionadas nos modelos GA-LDA, AOC-LDA e SW-LDA.....	83

1. Introdução

As ciências forenses estão em constante desenvolvimento, buscando cada vez mais aprimorar os métodos de investigação e contando com profissionais especializados em diversas áreas de aplicação. As amostras a serem investigadas podem ser macro ou microscópicas, tendo como característica em comum a preferência por metodologias que sejam rápidas e mantenham a integridade da amostra¹.

O estudo de tintas de canetas nas investigações forenses sobre documentos questionados busca identificar possíveis alterações nesses documentos. Isso pode ser evidenciado pela identificação de diferentes corantes, ou outros compostos encontrados nas formulações das tintas, ao longo do documento, comprovando que diferentes canetas foram utilizadas². Também é de suma importância a investigação do papel do documento, que pode ser alterado entre as páginas, como no caso de contratos, ou também para identificação de falsificação de papel moeda.

Na investigação de crimes, muitas vezes são encontradas evidências microscópicas, como no caso das impressões digitais latentes, cujo objetivo principal do estudo costuma ser a identificação dos envolvidos e a dinâmica do crime^{3,4}. A impressão digital possui um desenho único para cada indivíduo, mas muitas vezes são encontrados apenas fragmentos dela, que quando analisados quanto à composição química, podem indicar sexo e faixa etária do indivíduo, assim como o tempo de deposição desse fragmento, elucidando quando o crime ocorreu⁴.

Diversas técnicas analíticas são empregadas nos estudos forenses e apresentam excelentes resultados, contudo a grande maioria causa destruição da amostra no preparo, o que não é desejável, sendo até mesmo proibido, como no caso de investigações de itens de elevado valor. A aplicação da espectroscopia no infravermelho é de grande valia pela rapidez na análise e por ser não destrutiva, podendo ser aplicada nos mais diversos tipos de amostra e tendo menor custo de análise⁵. A associação dessa técnica com ferramentas quimiométricas possibilita maior facilidade e agilidade na

interpretação dos resultados, além de destacar as informações químicas mais relevantes, proporcionando resultados mais acertivos⁶.

O presente trabalho estudou o emprego de espectroscopia no infravermelho médio com transformada de Fourier por reflexão total atenuada (ATR-FTIR) e ferramentas quimiométricas aplicadas em duas áreas forenses descritas a seguir, com o objetivo de desenvolver metodologias rápidas e não destrutivas para aplicações em casos policiais:

- i) para diferenciar traços de canetas esferográficas, na área de documentoscopia e,
- ii) para estudar o envelhecimento de impressões digitais latentes, na área de datiloscopia.

2. Revisão bibliográfica

2.1 Ciências forenses

As ciências forenses representam a aplicação do conhecimento das mais diversas áreas científicas, não só na elucidação de crimes, mas contribuindo com investigações judiciais em geral^{1,7}. Com o desenvolvimento das ciências forenses, cada vez mais profissionais especializados de diferentes áreas são necessitados, caracterizando essa disciplina por ampla interdisciplinaridade. Isso é corroborado pelo aumento das pesquisas e publicações na área nas duas últimas décadas, conforme demonstrado por Rodrigues *et al.* (2022)⁸.

Na área de química forense, as amostras oriundas da investigação são submetidas a análises físicas e químicas tendo como objetivo principal caracterizar uma amostra suspeita ou detectar alguma adulteração ou falsificação⁹. Muitas das amostras investigadas são evidências físicas macroscópicas, como documentos, roupas ou armas. No entanto, esta categoria também inclui quantidades microscópicas de matéria física comumente referidas como "evidências residuais", que podem incluir marcas de impressões digitais, sujeira, vidro, tinta, cabelo, fibras ou partículas explosivas^{9,10}.

Devido ao elevado campo de aplicação das ciências forenses, ainda encontramos subdivisões conforme o tipo de amostra a ser investigada, exemplos:

- Balística forense, relacionada às armas de fogo, projéteis e explosivos¹¹;
- Papiloscopia, que trabalha com identificação de pessoas, sendo ainda subdividida em podoscopia (impressões da sola do pé), quiroscopia (impressões da palma da mão) e datiloscopia (impressões digitais dos dedos)^{11,12};
- Toxicologia, aplicada ao uso de venenos, drogas de abuso e até mesmo elucidação de acidentes de trânsito¹³;
- Documentoscopia, estuda desde o papel, chamado de suporte do documento, como tintas de caneta e carimbos, impressões, documentos oficiais, entre outros².

2.1.1 Documentoscopia

A documentoscopia é o estudo da autenticidade ou autoria de documentos^{2,14}. A falsificação de um documento se dá pela modificação e/ou acréscimo de palavras e/ou números¹⁵ e os documentos suspeitos podem ser submetidos a análises físicas visuais e microscópicas, o que permite a identificação de recortes, manchas e alterações de alinhamento de texto, além de inspeções com utilização de luzes (infravermelho e ultravioleta) visando a identificação de elementos de segurança presentes em documentos oficiais¹⁶.

Na documentoscopia forense é muito estudado sobre tintas de canetas, cujo objetivo principal é a identificação de diferentes tintas de canetas existentes em um mesmo documento, buscando evidenciar possíveis alterações no documento questionado². As tintas de caneta possuem diversos componentes em suas formulações, sendo os corantes os mais importantes na distinção de tintas por permanecerem fixados ao papel¹⁷.

Na literatura diversas técnicas analíticas são empregadas em análises de tinta de canetas, como cromatografia gasosa com espectrometria de massas^{18,19}, eletroforese capilar²⁰, cromatografia em camada delgada^{17,21}, cromatografia líquida de alta eficiência²², entre outras. Apesar de eficientes, essas técnicas descritas envolvem a remoção da tinta do documento para posterior análise, o que não é desejável nas aplicações forenses por destruir a amostra. Técnicas não destrutivas são mais indicadas, reduzindo os danos ao documento questionado e ainda, algumas vezes, sendo mais rápidas⁵. Silva (2017)⁶ salienta o quanto técnicas espectroscópicas, entre elas o infravermelho, aliadas a quimiometria, têm capacidade de preservar as amostras, serem rápidas e ainda obter resultados objetivos e independentes da opinião do analista. Essa combinação de espectroscopia no infravermelho com quimiometria têm amplo campo de estudo, podendo ser aplicada para discriminação de tintas de canetas esferográfica e gel²³, tintas de carimbo²⁴, tintas de documentos impressos²⁵, como também pelo emprego de diferentes pré-processamentos e modelagens quimiométricas.

2.1.2 *Datilosopia*

A datilosopia é a subárea da papiloscopia mais utilizada e onde se estuda a respeito das impressões digitais existentes nas pontas dos dedos. Há muito tempo empregada nas resoluções de casos forenses, a obtenção e identificação das impressões digitais é uma ferramenta poderosa e confiável para identificação dos envolvidos, suspeitos e vítimas, e também da dinâmica do crime, como meios de acesso e fuga^{3,4}.

Devido ao desenho único que possui a impressão digital de cada indivíduo, o método de identificação convencional baseia-se em comparar os padrões de linhas encontrados na cena de um crime com um suspeito ou um banco de dados, entretanto, quando os vestígios de impressão digital são fragmentos distorcidos, de baixa qualidade ou sem padrão em banco de dados é desejável que outras informações possam ser extraídas desse vestígio⁴.

Um movimento natural das pessoas é tocar o rosto e os cabelos com as mãos, por esse motivo a maioria dos componentes presentes nas pontas dos dedos são provenientes de secreções das glândulas sebáceas e sudoríparas, e esses são depositados nas superfícies em que os dedos tocam, deixando a impressão digital latente²⁶.

Proveniente das glândulas sudoríparas, o suor é predominantemente composto por água (98%) e contém também aminoácidos, proteínas e sais. Já as secreções sebáceas são compostas por ácidos graxos, esqualeno, glicerídeos, colesterol e ésteres lipídicos³. Essa composição química existente nas impressões digitais latentes vem sendo amplamente estudada, como por exemplo, para fins de diferenciar vestígios provenientes de homens e mulheres³, recentes e envelhecidos^{27,28}, adultos e crianças²⁹.

Não diferente das outras áreas das ciências forenses, no estudo de impressões digitais latentes também é preferido métodos analíticos não destrutivos, rápidos e confiáveis, o que tem feito com que as técnicas espectroscópicas tenham bastante destaque, principalmente a espectroscopia no infravermelho, no modo de refletância total atenuada e aliada à quimiometria²⁷.

2.2 Espectroscopia no infravermelho

A espectroscopia no infravermelho (IR – *Infrared*) tem como base a absorção molecular e compreende a região do espectro eletromagnético entre o visível e as micro-ondas, conforme apresentado na Figura 1. Devido a ampla faixa de energia, de 18.000 a 10 cm^{-1} , e as diferentes aplicações, essa região é dividida em infravermelho próximo (NIR – *near infrared*), médio (MIR – *mid infrared*) e distante (FIR – *far infrared*), conforme as faixas de energia apresentadas na Tabela 1. A partir dos anos 1980 começaram a ser utilizados os espectrômetros de infravermelho com transformada de Fourier (FTIR – *Fourier Transform Infrared*), uma operação matemática complexa que tornou os equipamentos mais rápidos, sensíveis e menos ruidosos^{30,31}.

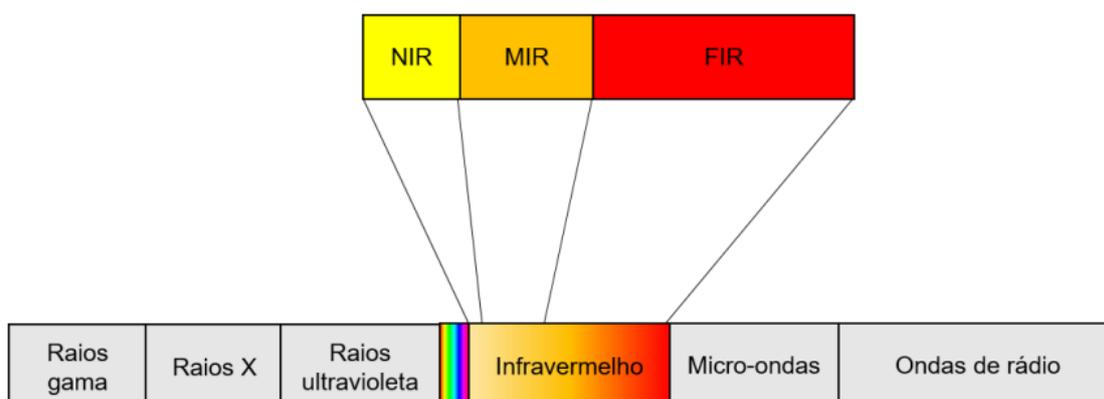


Figura 1. Regiões do espectro eletromagnético com destaque para a faixa do infravermelho. Fonte: John, D.K. 2020³².

Tabela 1. Regiões espectrais do infravermelho

Região	Faixa de Número de onda (cm^{-1})	Faixa de Comprimento de onda (nm)
NIR	12.800 a 4.000	780 a 2.500
MIR	4.000 a 200	2.500 a 50.000
FIR	200 a 10	50.000 a 100.000

A energia da radiação infravermelha está associada aos movimentos vibracionais e rotacionais das moléculas. As absorções para movimentos

unicamente rotacionais são restritas a região do FIR, devido à baixa energia necessária³³.

A absorção de energia referente aos movimentos vibracionais está associada a variação no momento dipolar de uma ligação e quando essa energia é absorvida aumenta a amplitude desses movimentos, que naturalmente já estão acontecendo. O momento dipolar de uma ligação é determinado pela dimensão da diferença de carga e distância entre os átomos envolvidos, varia conforme a combinação dos átomos nas moléculas e, por isso, a energia específica envolvida em cada absorção é relacionada ao grupo funcional presente na molécula. Além disso, a intensidade da absorção de determinada banda está diretamente relacionada com a concentração do analito, permitindo análises qualitativas e quantitativas. Os movimentos vibracionais são divididos em estiramentos e deformações angulares, conforme demonstrado na Figura 2^{31,34}.



Figura 2. Modos de vibrações moleculares. Fonte: Santos, C.S. 2017⁶.

Os espectrômetros de FTIR possuem acessórios de compartimento amostral que diferem na maneira como a radiação interage com a amostra e também no preparo da amostra para a análise. Um modo em destaque, que não requer preparo de amostra e pode ser utilizado para líquidos e sólidos é a refletância total atenuada (ATR – *Attenuated Total Reflectance*)³¹.

2.2.1 Refletância Total Atenuada (ATR)

No método de refletância total atenuada um cristal com alto índice de refração é pressionado sobre a amostra. A radiação infravermelha incide nesse cristal com um ângulo específico, de modo que o feixe seja totalmente refletido no interior desse cristal, conforme demonstrado na Figura 3. Nesse processo de reflexão parte da radiação penetra na amostra, sofrendo uma atenuação relativa à energia absorvida para estímulo vibracional molecular. A radiação que interage com a amostra é chamada de onda evanescente e a medida das variações dessa onda é apresentada num espectro de infravermelho³¹.

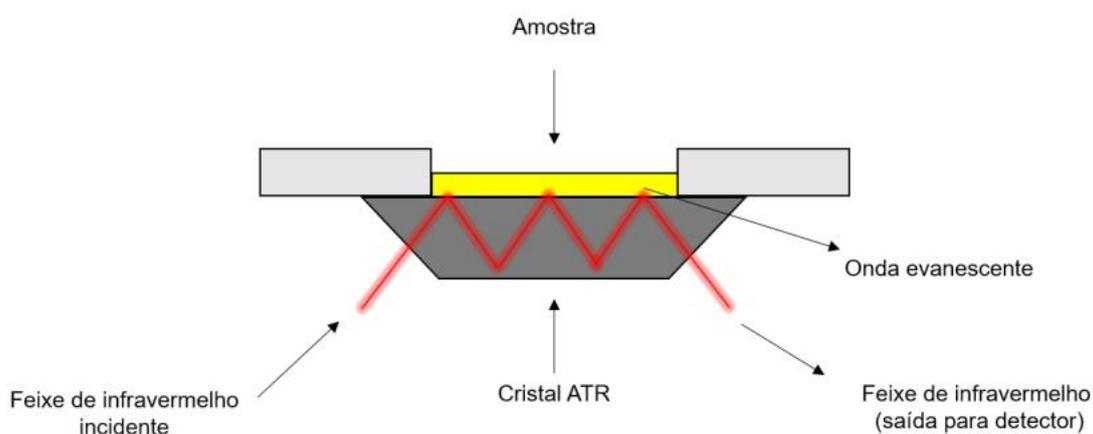


Figura 3. Ilustração representativa da análise por ATR. Fonte: John, D.K. 2020³².

No mercado há diferentes opções de cristais para ATR, que variam conforme a faixa espectral abrangida, a profundidade de penetração da onda evanescente, a resistência e o custo do material empregado. Mais comumente é utilizado seleneto de zinco (ZnSe) com faixa espectral de 15.000 a 650 cm^{-1} ³⁵.

2.3 Quimiometria

A quimiometria é uma área da química e é definida, pela União Internacional de Química Pura e Aplicada (IUPAC - *International Union of Pure and Applied Chemistry*), como a ciência que relaciona dados químicos

utilizando métodos estatísticos e matemáticos, de modo a se extrair ao máximo as informações químicas contidas nesses dados, ou seja, embora matemática e estatística sejam fundamentais para se aplicar a quimiometria, num geral os problemas são oriundos e de interesse da química³⁶.

Desde o surgimento, a quimiometria tem tido crescente aplicação, vista pelo aumento de artigos científicos publicados com esse tópico, sendo considerada tendência no mundo moderno, devido a agilidade que proporciona no tratamento de grandes volumes de dados complexos e multivariados³⁷.

Existem muitos métodos quimiométricos e esses são divididos conforme sua finalidade, como por exemplo reconhecimento de padrões, planejamento experimental, calibração, entre outros. Nos métodos de reconhecimento de padrões ainda há subdivisão entre métodos não supervisionados, onde não há informações de entrada sobre as amostras e essas são agrupadas conforme as semelhanças, diferenças e tendências naturais existentes, e os métodos supervisionados, em que é necessário criar um modelo de treinamento com um conjunto de amostras conhecidas e após testar esse modelo com outro conjunto de amostras, geralmente denominado de conjunto teste^{38,39}.

2.3.1 Pré-processamento de dados

Dados espectrais brutos podem conter informações que não dizem respeito exatamente a amostra, mas as condições de aquisição dos dados, carregando erros aleatórios e sistemáticos. Com o intuito de minimizar esses artefatos instrumentais, alguns pré-processamentos podem ser aplicados nos dados. A suavização busca minimizar o ruído dos espectros, elevando a razão sinal/ruído. A normalização entre 0 e 1 consiste em padronizar todas as amostras em uma mesma escala, removendo variações relacionadas ao tamanho da amostra. Outro pré-processamento utilizado é a derivada, que pode destacar bandas não evidentes e, quando aplicada a 1ª derivada, uma correção em relação aos deslocamentos do espectro é feita, enquanto a 2ª derivada corrige inclinação da linha de base^{38,40,41}. Centrar os dados na média é frequentemente aplicado à dados espectroscópicos e consiste em subtrair para cada comprimento de onda a média das intensidades dessa variável, alterando o sistema de coordenadas para o centro dos dados, não havendo

alteração de informação dos dados, mas reduzindo influências não relacionadas ao analito^{35,41,42}.

2.3.2 Análise de componentes principais (PCA)

A análise de componentes principais (PCA – *Principal Component Analysis*) pode ser utilizada como um método exploratório de análise de dados na quimiometria, sendo também a base de outros métodos de tratamento de dados multivariados⁴³.

A PCA consiste em reduzir a dimensão dos dados, mantendo as informações dos dados originais, selecionando um conjunto menor de novas variáveis, que salientam as informações mais relevantes, agrupando as amostras semelhantes, além de indicar amostras possivelmente fora do padrão (*outliers*). Essas novas variáveis criadas são as componentes principais (PCs – *Principal Components*) e, entre si, são ortogonais e não correlacionadas, acumulando na primeira componente principal (PC1) a maior quantidade de informações (percentual de variância explicada) sobre os dados, sendo ela a mais representativa. As demais PCs vão conter gradualmente menos informações^{32,40}.

Inicialmente, as amostras estão dispostas num espaço com dimensão igual ao número de variáveis e, após a matriz de dados ser processada pela PCA, passam a ser apresentadas em duas ou três dimensões. De forma matemática, na PCA a matriz de dados original, **X**, é decomposta no produto de duas novas matrizes, contendo uma delas os escores (**T**) e a outra os pesos (**P**), mais uma matriz de resíduos (**E**), conforme demonstrado na Equação 1. Os escores carregam a posição das amostras no espaço (bi ou tridimensional) formado pelas PCs, enquanto os pesos representam o quanto cada variável original contribui para determinada PC e, nos resíduos, ficam os erros experimentais e principalmente os ruídos^{30,41}.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{Equação 1}$$

É importante a avaliação conjunta dos escores e pesos, de modo a entender como as amostras são agrupadas e quais variáveis originais são responsáveis pelas diferenças entre os grupos⁴¹.

2.3.3 Análise de agrupamentos por métodos hierárquicos (HCA)

Similar à PCA, a análise de agrupamentos por métodos hierárquicos (HCA – *Hierarchical Clusters Analysis*) é também um método de reconhecimento de padrões não supervisionado, cujo objetivo é agrupar as amostras conforme suas similaridades naturais, através do cálculo das distâncias no espaço multidimensional³⁰.

Inicialmente, cada amostra é um ponto no espaço de n dimensões e a distância entre cada par de amostras é calculada por um parâmetro de agrupamento. As amostras mais similares, ou seja, as que apresentam maior proximidade entre si formam um grupo. Após todas as amostras terem formado grupos, esses grupos são considerados pontos no espaço de n dimensões e novamente o parâmetro de agrupamento é calculado. Essas etapas se repetem até restar apenas um grande grupo e o resultado é apresentado em forma de dendrograma^{40,44}.

Como exemplos de parâmetros de agrupamento, existe a distância de Manhattan, a correlação de Pearson e, o mais comumente utilizado, a distância Euclidiana. A distância Euclidiana é calculada pela aplicação do teorema de Pitágoras para o espaço n dimensional, conforme a Equação 2, em que D_{ab} consiste na distância entre as amostras a e b, obtida pela raiz quadrada do somatório dos quadrados das distâncias entre as n coordenadas de a e b^{32,43}.

$$D_{ab} = \sqrt{\sum_{i=1}^n (x_{ai} - x_{bi})^2} \quad \text{Equação 2}$$

Para a ligação entre os grupos pode-se utilizar: a menor distância entre duas amostras pertencentes à grupos diferentes; o oposto, a maior distância entre duas amostras pertencentes à grupos diferentes; a distância entre os pontos médios de cada grupo; a distância entre os pontos centrais de cada grupo; ou ainda, pelo método de Ward que, bem diferente dos anteriores, junta dois agrupamentos buscando a variância mínima dentro do novo grupo formado^{32,44,45}.

2.3.4 *Partição do grupo amostral em subconjuntos*

Na aplicação dos métodos supervisionados de reconhecimento de padrões, inicialmente é criado um modelo com as amostras de treinamento e depois faz-se o teste desse modelo com um conjunto de amostras teste. Para isso é importante a divisão do conjunto amostral em subconjuntos representativos. O algoritmo mais conhecido e empregado na seleção de subgrupo representativo é o Kennard-Stone⁴⁶. Basicamente, este método seleciona a amostra mais próxima da média do conjunto amostral, e após, sucessivamente, as amostras mais distantes das já selecionadas, até atingir o quantitativo desejado. Dessa forma, esse algoritmo forma primeiramente um subconjunto com as amostras mais representativas do conjunto amostral, que é utilizado para o modelo empregando as amostras de treinamento. As amostras restantes são empregadas como conjunto teste, com as quais calcula-se o desempenho do modelo^{42,47}.

2.3.5 *Análise discriminante linear (LDA)*

A análise discriminante linear (LDA – *Linear Discriminant Analysis*) foi apresentada por Fisher em 1936 e se tornou bastante popular e bem estabelecida devido à sua simplicidade e robustez⁴⁸. De forma análoga à PCA, esse método busca uma redução de dimensionalidade removendo informações redundantes, porém resultando numa direção que atinja a máxima separação entre as classes^{39,49}. A Figura 4 ilustra a distribuição espacial das amostras antes e após uma LDA.

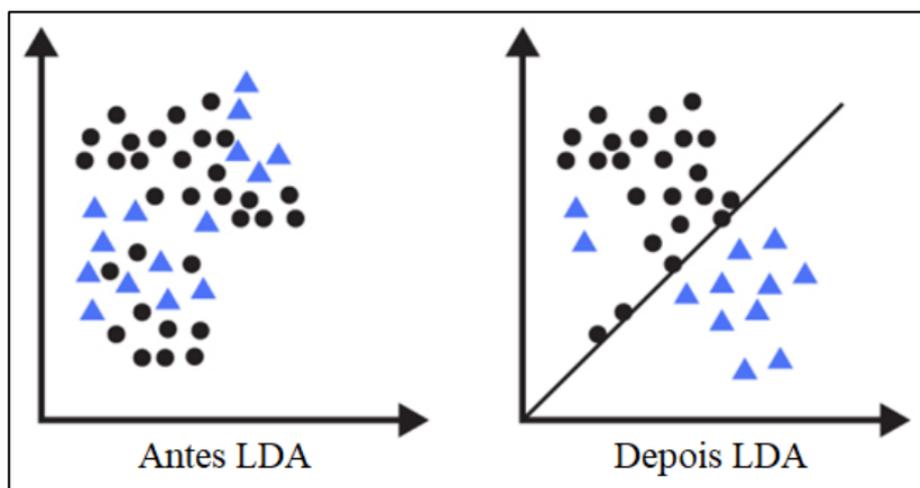


Figura 4. Exemplo de reagrupamento de objetos com LDA. Fonte: Junges, C.H. 2023³⁹.

Na LDA, funções discriminantes são definidas a partir da combinação linear das variáveis originais, separando as classes de forma ideal, com máxima variação entre as classes e mínima variação dentro das classes^{39,47}.

Cada função discriminante y tem a forma geral da Equação 3:

$$y = a_1x_1 + a_2x_2 + \dots + a_px_p \quad \text{Equação 3}$$

onde a_i representam os coeficientes da função discriminante, x_i as variáveis originais e p o número de variáveis iniciais⁵⁰. As funções discriminantes são construídas uma a uma, buscando direções no espaço que maximizem o critério discriminante (λ), conforme descrito na Equação 4:

$$\lambda = \frac{SS_b}{SS_w} \quad \text{Equação 4}$$

na qual SS_b indica a soma dos quadrados das distâncias euclidianas entre as classes e SS_w a soma dos quadrados das distâncias euclidianas dentro da classe. A primeira função discriminante resulta na maior separação das classes, pois possui o máximo valor de λ . Tendo-se C classes, é possível a construção de $C-1$ funções discriminantes^{50,51}.

A LDA apresenta resultados satisfatórios quando as classes são extremamente homogêneas. Outro limitante muito importante é que o número de variáveis não pode ser maior que o número de amostras, sendo um problema quando se trabalha com dados espectrais, que geralmente apresentam um número elevado de variáveis espectrais. Uma saída para esse problema é a utilização conjunta da LDA com os algoritmos de seleção de variáveis^{35,39,50}.

2.3.6 Análise discriminante por mínimos quadrados parciais (PLS-DA)

A análise discriminante por mínimos quadrados parciais (PLS-DA – *Partial Least Squares Discriminant Analysis*) é uma variante do modelo PLS, cujo objetivo é prever a classe de uma amostra com base em coeficientes de regressão estabelecidos. A partir de uma matriz contendo as variáveis originais (X), uma matriz qualitativa (Y) é criada com base nas classes das amostras. Nessa etapa de treinamento, a matriz Y apresenta as amostras em uma classificação binária que indica quais pertencem a classe em análise (tendo o valor 1 atribuído) e quais pertencem as demais classes (recebendo o valor 0), conforme demonstrado na Figura 5^{6,40,48}.

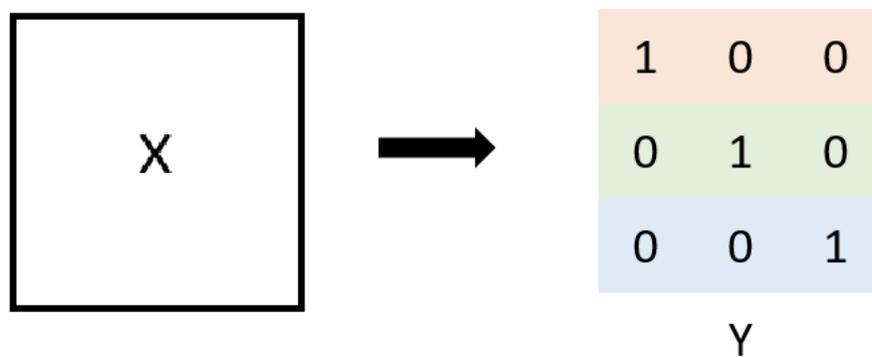


Figura 5. Esquema da etapa de treinamento do PLS-DA.

Após a etapa de treinamento, um *threshold* (limite) é calculado para cada classe em função das densidades de probabilidade, e então, no teste do modelo, as amostras que retornarem *y* acima do *threshold*, ou seja, mais

próximo de 1, serão consideradas pertencentes à classe, em caso contrário não^{6,35,48}.

2.3.7 Técnicas de seleção de variáveis

Dados multidimensionais, em especial os resultantes de métodos espectroscópicos, podem conter elevado número de variáveis, conseqüentemente com informações redundantes e até mesmo ruído, o que afeta negativamente o desempenho dos modelos quimiométricos⁴². Os algoritmos de seleção de variáveis acabam sendo essenciais para a redução do conjunto de variáveis, formando um subconjunto menor, com as informações mais relevantes, melhorando o desempenho dos modelos e até mesmo reduzindo o tempo computacional de modelagem³⁹.

As técnicas de seleção de variáveis podem ser divididas em modelos probabilísticos, como algoritmo genético e otimização da colônia de formigas, e determinísticos, como *stepwise* e algoritmo de projeções sucessivas.

2.3.7.1 Algoritmo Genético – GA (*Genetic Algorithm*)

O algoritmo genético simula a evolução do processo biológico proposto por Charles Darwin em 1859. Algumas variáveis são aleatoriamente selecionadas e formam indivíduos de uma população, em que cada indivíduo corresponde a uma possível solução do problema, ou seja, um conjunto de variáveis que discrimina bem os grupos amostrais. Após, os indivíduos são testados quanto a capacidade de resolução do problema (sua aptidão), onde os indivíduos mais aptos geram menor erro de predição do modelo. Os melhores indivíduos (grupos de variáveis) são selecionados para a etapa seguinte do modelo, que consiste em um cruzamento desses indivíduos, de forma análoga à reprodução, com o objetivo de avaliar se indivíduos melhores são formados. Nessa etapa também acontece a mutação, em que um indivíduo tem um ou mais genes (variáveis) aleatoriamente substituídos e então a nova população formada é testada, selecionando os melhores indivíduos, que são inseridos novamente no processo evolutivo. Essas etapas de seleção, cruzamento e avaliação acontecem até se atingir o objetivo determinado pelo analista^{39,52,53}. Na Figura 6 é apresentado o fluxograma proposto por Junges³⁹ que demonstra as etapas do GA.

Por ser um método probabilístico, é de extrema importância a boa escolha dos parâmetros de controle, como tamanho da população, taxa de cruzamento e de mutação e reinserção de novos indivíduos para um bom desempenho do algoritmo. Quando bem determinados os parâmetros de controle, o GA tende à excelentes resultados por avaliar diversas combinações de variáveis até chegar na melhor seleção, entretanto exige maior tempo e desempenho computacional^{39,52,53}.

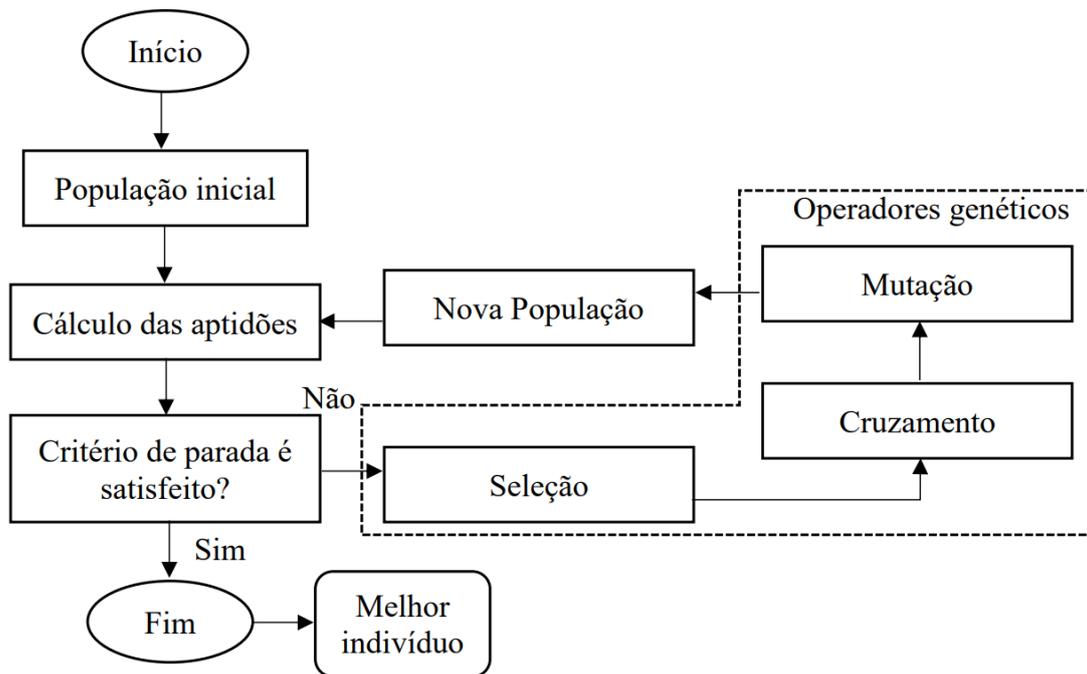


Figura 6. Fluxograma das etapas do algoritmo genético para seleção de variáveis. Fonte: Junges, C.H. 2023³⁹.

2.3.7.2 Algoritmo de Otimização da Colônia de Formigas – ACO (*Ant of Colony Optimization*)

Desenvolvido por Dorigo e colaboradores e apresentado em trabalhos no final da década de 90, o algoritmo de otimização da colônia de formigas foi inspirado no comportamento das formigas em deixar feromônios nos caminhos entre o ninho e fontes de comida^{54,55}. Inicialmente as formigas percorrem aleatoriamente diversos caminhos possíveis entre o ninho e o alimento e depositam feromônios. Caminhos mais curtos são percorridos mais rapidamente e, como os feromônios evaporam com o tempo, acabam ficando

com maior concentração dessa substância, induzindo outras formigas a fazer esse caminho menor, conseqüentemente depositando mais feromônio, até que esse se torna o único caminho para ser percorrido. Dessa forma, coletivamente as formigas otimizam o caminho entre o ninho e o alimento, como ilustrado na Figura 7^{39,54,55,56,57}.

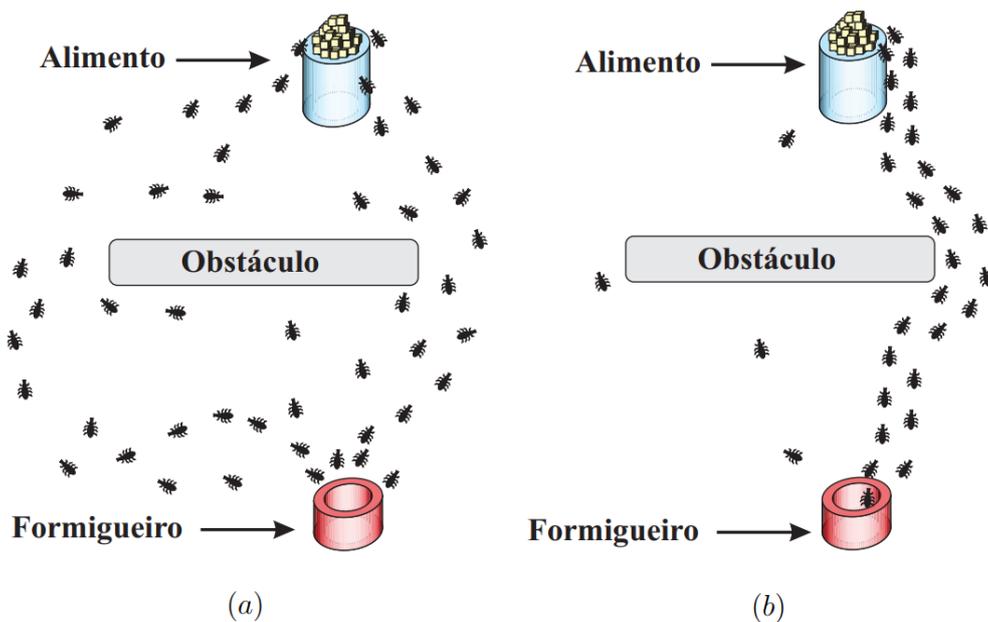


Figura 7. Exemplo do comportamento das formigas antes (a) e depois (b) de otimizarem o caminho entre o ninho e o alimento. Fonte: Sanca, H.S. 2013⁵⁶.

Na aplicação desse algoritmo para seleção de variáveis inicialmente todas as variáveis recebem a mesma quantidade de feromônio e por isso têm mesma probabilidade de ser escolhida como parte do caminho das formigas. Cada formiga seleciona um conjunto aleatório de variáveis, considerado um caminho possível a ser percorrido, e este é caracterizado em função do objetivo a ser atingido, ou seja, a melhor modelagem dos dados. A formiga que apresentar melhor resultado deposita feromônio nas variáveis utilizadas por ela, aumentando a probabilidade dessas variáveis serem escolhidas como parte do caminho em uma próxima iteração. Após determinado número de ciclos as variáveis com mais feromônios são consideradas como o caminho ótimo e são as utilizadas para modelagem dos dados^{39,53}.

2.3.7.3 Algoritmo *Stepwise* – SW

A otimização *stepwise* consiste em adicionar e remover variáveis em um modelo e avaliar o desempenho desse, selecionando as variáveis que mais importam na modelagem das classes.

Inicialmente o algoritmo avalia a variável que mais discrimina as classes e essa é adicionada à um subgrupo. Na sequência uma outra variável é adicionada a esse subgrupo inicial e é avaliado se o coeficiente de correlação aumentou ou diminuiu, sem ultrapassar o limite de correlação determinado pelo analista. Se a adição da nova variável aumentar o coeficiente de correlação, essa permanece no subgrupo, caso contrário, é eliminada. Quando uma variável permanece é testado se eliminando alguma outra variável previamente existente a correlação melhora e, em caso positivo, a variável eliminada fica fora do subgrupo ou, em caso negativo, ela retorna. Quando não há mais variáveis para testar a saída, nova variável é adicionada e a mesma sequência de avaliações é realizada até que não haja mais variáveis para serem adicionadas ao subgrupo, estando esse otimizado para o modelo^{39,52,58}.

2.3.7.4 Algoritmo de Projeções Sucessivas – SPA (*Successive Projection Algorithm*)

Com o objetivo de selecionar variáveis minimamente colineares, o SPA, proposto em 2001 por Araújo *et al.* (2001)⁵⁹, é considerado um dos métodos determinísticos mais conhecidos, devido sua simplicidade e seu menor tempo de processamento em relação a outros algoritmos^{35,59}.

Basicamente para cada variável disponível são feitas sucessivas projeções das demais variáveis e, a cada iteração, uma nova variável é adicionada ao conjunto, construindo uma cadeia de variáveis, tendo como critério de ordenamento apresentar menor colinearidade com as variáveis anteriores e não podendo selecionar um número de variáveis maior que o número de amostras de calibração. Os conjuntos formados são modelados, avaliados e o melhor selecionado. Para métodos de reconhecimento de padrões é utilizado como critério de avaliação raiz quadrada do erro médio quadrático de validação (RMSEV), enquanto para métodos de classificação se

utiliza o risco médio de classificação incorreta das amostras de validação. Como última etapa do modelo, é avaliado se alguma variável do melhor conjunto pode ser excluída sem significativa perda de capacidade de predição. Esse algoritmo resulta em uma única solução após o processo de seleção de variáveis, com mínimo de colinearidade e de informações redundantes entre as variáveis selecionadas^{39,59,60,61,62}.

2.3.8 Figuras de mérito

Uma maneira de avaliar o desempenho e a confiabilidade dos modelos quimiométricos se dá pelo cálculo das figuras de mérito, que são avaliações estatísticas dos resultados obtidos.

Os modelos de classificação geram como resultado uma matriz de confusão, onde é indicado não só quantas amostras foram classificadas em cada classe, como também à qual classe pertenciam as amostras erroneamente classificadas. Na Figura 8 têm-se como exemplo uma matriz de confusão de duas classes e as possibilidades de classificação em relação a classe A. As amostras verdadeiras são aquelas que foram atribuídas corretamente na classe que pertencem, enquanto as falsas foram atribuídas na classe errada. As amostras positivas são aquelas atribuídas a classe avaliada e as negativas são as que foram atribuídas na outra classe⁶.

		CLASSE CORRETA	
		A	B
CLASSE ATRIBUÍDA	A	VP_A	FP_A
	B	FN_A	VN_A

Figura 8. Matriz de confusão de duas classes. V=verdadeiro, F=falso, P=positivo e N=negativo

A partir da matriz de confusão pode-se calcular a acurácia do modelo, conforme a Equação 5, que demonstra o desempenho de modo geral pela avaliação de quantas amostras foram atribuídas as classes corretas em relação ao total de amostras. Também é possível avaliar alguns parâmetros por classe, como^{6,63,64}:

- Sensibilidade, que representa a capacidade do modelo em identificar e classificar corretamente todas as amostras pertencentes a classe avaliada, expressa pela Equação 6, e;
- Especificidade, que indica a capacidade do modelo em não classificar amostras erradas na classe avaliada, ou seja, evitar falsos positivos, obtida conforme a Equação 7.

$$\text{Acurácia (\%)} = \frac{\text{VN} + \text{VP}}{\text{VN} + \text{VP} + \text{FN} + \text{FP}} \times 100 \quad \text{Equação 5}$$

$$\text{Sensibilidade (\%)} = \frac{\text{VP}}{\text{VP} + \text{FN}} \times 100 \quad \text{Equação 6}$$

$$\text{Especificidade (\%)} = \frac{\text{VN}}{\text{VN} + \text{FP}} \times 100 \quad \text{Equação 7}$$

3. Objetivos

3.1 Objetivo geral

Este trabalho teve por objetivo o estudo e desenvolvimento de metodologias, empregando dados de espectrometria no infravermelho médio e ferramentas quimiométricas, visando aplicação na Documentoscopia e Datiloscopia Forenses.

3.2 Objetivos específicos

3.2.1 Desenvolver e validar modelos de análise multivariada para identificação de tintas de canetas esferográficas em diferentes papéis:

Este objetivo visa utilizar técnicas de espectroscopia no infravermelho médio com transformada de Fourier por reflexão total atenuada (ATR-FTIR) e quimiometria para diferenciar traços de canetas esferográficas em diversos tipos de papel. Métodos como PCA, LDA e PLS-DA foram aplicados para validar a eficácia dos modelos na separação de amostras conforme a tinta utilizada, destacando a influência do tipo de papel nos espectros obtidos.

3.2.2 Avaliar a preservação e degradação de impressões digitais ao longo do tempo:

Este objetivo foca em analisar como as impressões digitais latentes se preservam ou degradam ao longo do tempo sob diferentes condições de armazenamento. Espectros ATR-FTIR de impressões digitais foram obtidos e analisados usando HCA, PCA, LDA e PLS-DA para estudar as alterações químicas ocorridas entre 0 e 30 dias, permitindo a identificação de padrões de envelhecimento.

3.2.3 Aplicar e avaliar técnicas de seleção de variáveis em modelagem quimiométrica:

Este objetivo tem como foco testar e comparar diferentes métodos de seleção de variáveis (GA, AOC, SW e SPA) em modelos de análise discriminante linear (LDA) para otimizar a precisão e a eficiência na diferenciação de amostras forenses. A eficácia desses métodos foi avaliada com base nas figuras de mérito dos modelos gerados.

3.2.4 Implementação de ATR-FTIR como metodologia padrão em análises forenses:

O objetivo foi demonstrar o potencial da técnica ATR-FTIR associada a quimiometria como uma metodologia padrão em análises forenses, destacando suas vantagens como eficácia, rapidez, baixo custo e preservação da integridade amostral. A pesquisa visa consolidar o uso desta técnica em diversas aplicações forenses, promovendo seu reconhecimento e adoção em laboratórios de criminalística.

4. Diferenciação de canetas esferográficas em diferentes papéis

4.1 Materiais e métodos

4.1.1 Amostras

Traços de três canetas esferográficas, cor azul da marca Faber-Castell (AF), e cores azul e preta da marca BIC, respectivamente (AB) e (PB), todas com ponta 1.0 mm, foram confeccionados em quatro diferentes papéis ofício, sendo 3 nacionais, branco com gramatura 75 g/m² (75Br), branco com gramatura 63 g/m² (63Br) e reciclado com gramatura 75 g/m² (75Rec), e um papel branco com gramatura 75 g/m² fabricado na Argentina (75Arg). Para confecção das amostras, com cada caneta e em cada papel foram escritos os algarismos 1 e 0. Sobre cada algarismo foram escolhidos 3 pontos para aquisição dos espectros de ATR-FTIR.

4.1.2 Espectroscopia de infravermelho médio

Os traços foram analisados em um espectrofotômetro de infravermelho médio com transformada de Fourier e refletância total atenuada (ATR-FTIR) Bruker Lumus II, utilizando 32 varreduras, na região espectral de 4000 a 800 cm⁻¹, com resolução de 2 cm⁻¹. Para cada caneta, sobre os diferentes papéis, foram realizadas 96 leituras (em 6 pontos com 16 replicatas), totalizando 1152 espectros obtidos.

4.1.3 Pré-processamentos e análise multivariada

Para os espectros de ATR-FTIR foi definido como região de trabalho a faixa de 1800 a 800 cm⁻¹, região de impressão digital para tintas de caneta⁶⁵. Utilizando o software MatLab⁶⁶ e pacotes toolbox^{67,68,69}, na sequência, os dados foram suavizados empregando Savitzky-Golay (13 pontos e polinômio de segunda ordem), normalizados e foi aplicada a 1ª derivada (13 pontos e polinômio de segunda ordem), obtendo-se uma matriz de dados de dimensão 1052x487 (amostras x variáveis). Uma avaliação exploratória foi realizada pelo método não supervisionado PCA com os dados centrados na média. Após o

grupo amostral foi dividido em dois subgrupos com igual número de amostras, pelo emprego do algoritmo Kennard-Stone⁴⁶ e então modelados e testados pelos métodos supervisionados LDA e PLS-DA.

4.2 Resultados e discussão

O perfil espectral dos dados brutos pode ser observado na Figura 9. Para uma eficiente diferenciação das canetas e dos papéis, foi necessário pré-processar os dados.

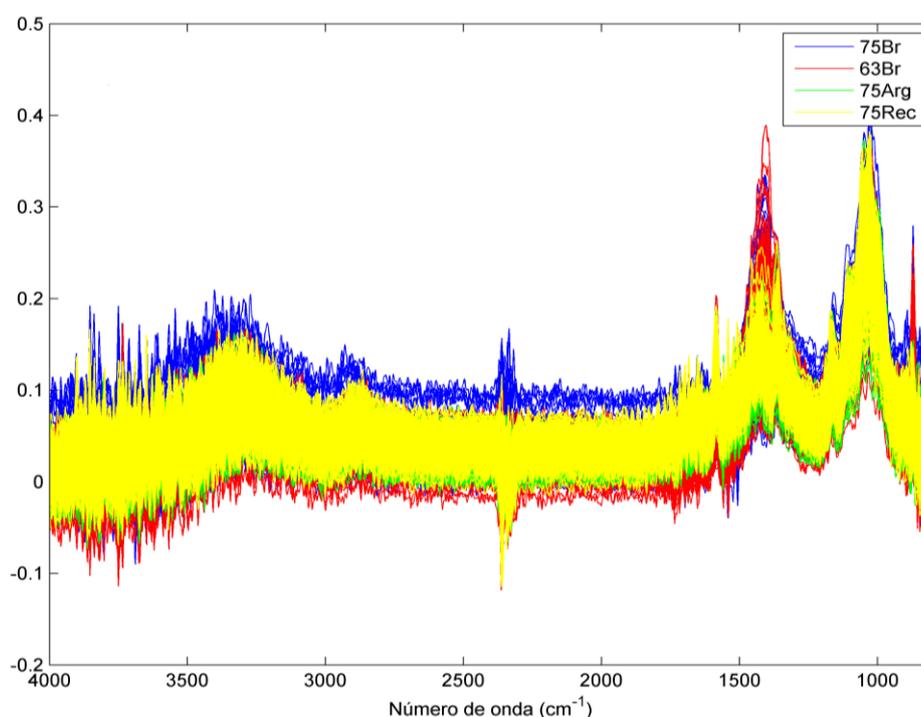


Figura 9. Espectros sem tratamento de ATR-FTIR das amostras de traço de caneta.

Os dados obtidos foram separados em 4 grupos amostrais, correspondentes aos diferentes papéis, para a aplicação dos pré-processamentos. Uma PCA dos dados recortados e suavizados apresentou tendência a separar as 3 canetas nas PCs 4 e 5, com variância explicada baixíssima. Os dados foram então normalizados entre 0 e 1 e uma nova PCA foi realizada, obtendo-se a separação das canetas nas PCs 3 e 4, porém continuando com baixa variância explicada. Para um melhor resultado aplicou-

se a 1ª derivada aos dados, sendo o espectro final de cada grupo apresentado na Figura 10.

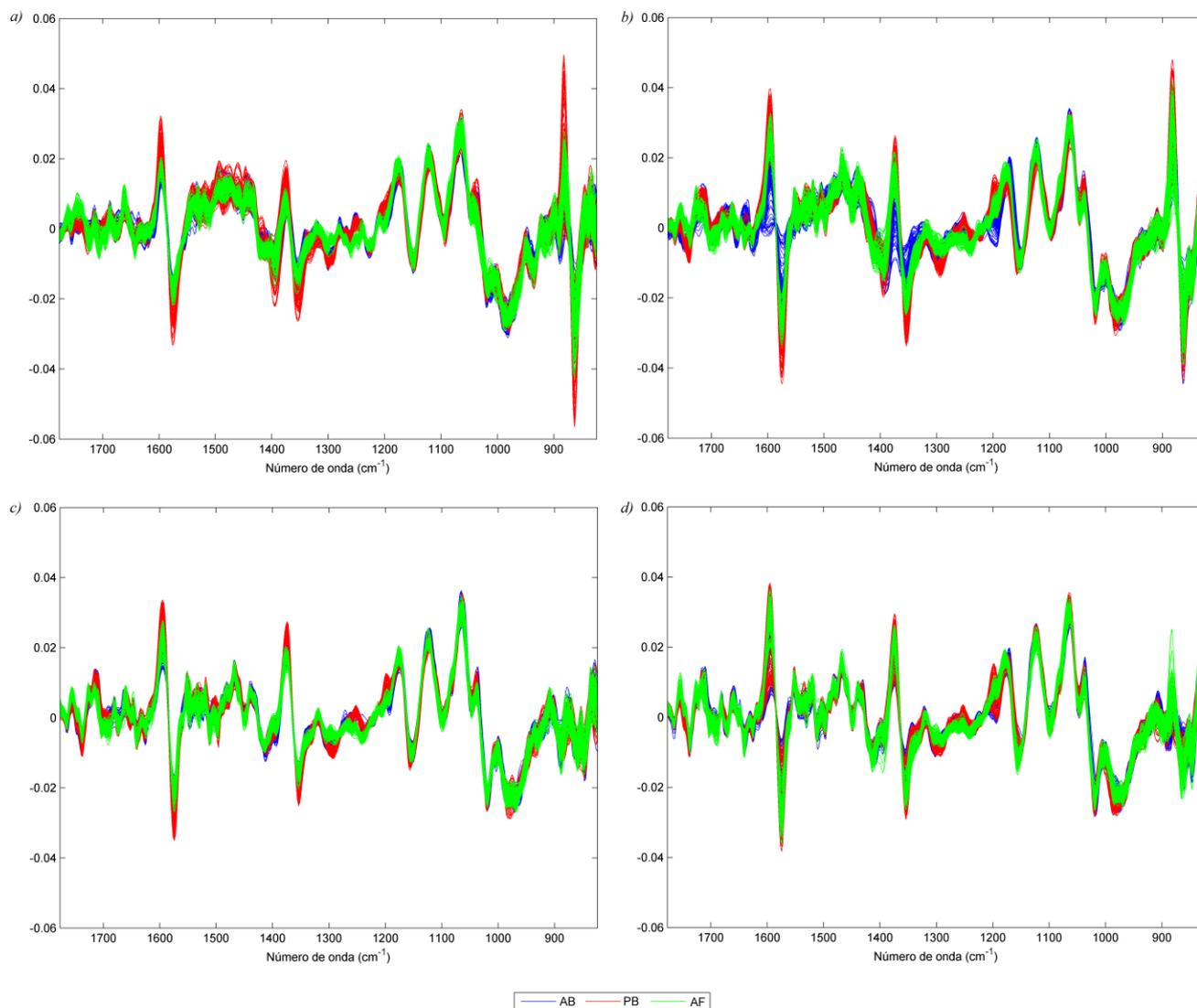


Figura 10. Espectros de ATR-FTIR após os pré-processamentos. a) 75Br b) 63Br c) 75Arg e d) 75Rec.

4.2.1 *Análise de componentes principais*

Após os pré-processamentos foi possível diferenciar as 3 classes de canetas estudadas por PCA, com os dados centrados na média, em cada papel utilizado. No resultado dos três papéis brancos, Figuras 11, 12 e 13, é possível visualizar claramente 3 grupos, cada um correspondente a uma caneta,

variando as PCs utilizadas para isso. Para o papel reciclado não foi possível obter boa separação das 3 classes de canetas, apenas uma tendência de separar as marcas, sem conseguir separar as duas cores de canetas de mesma marca (Figura 14).

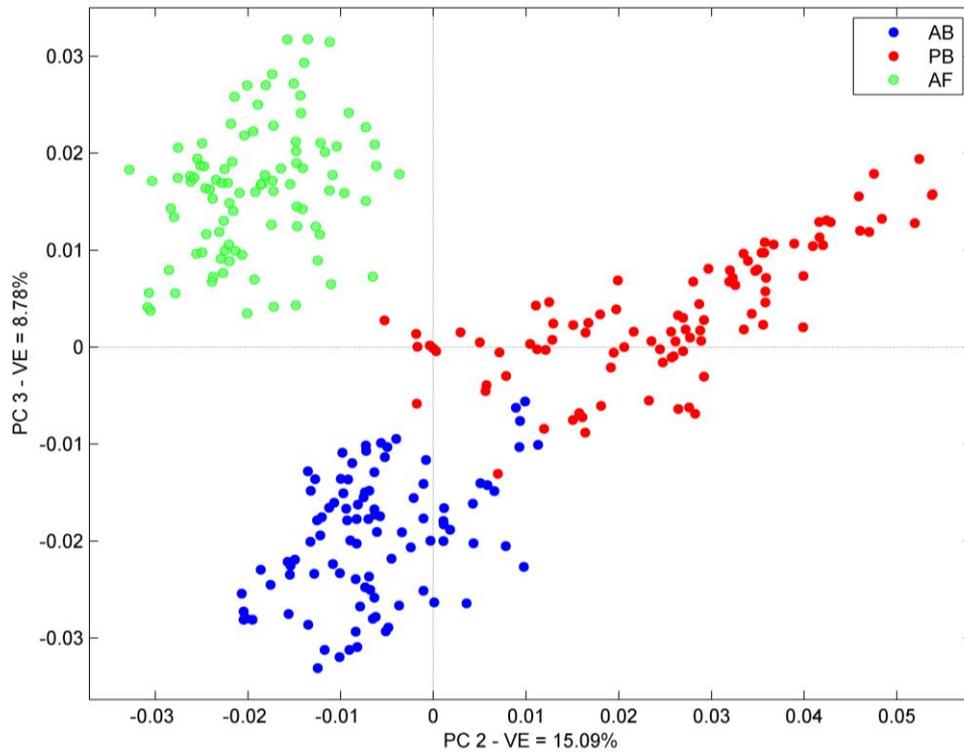


Figura 11. Gráfico dos escores PC2xPC3 da PCA com os dados obtidos sob o papel 75Br.

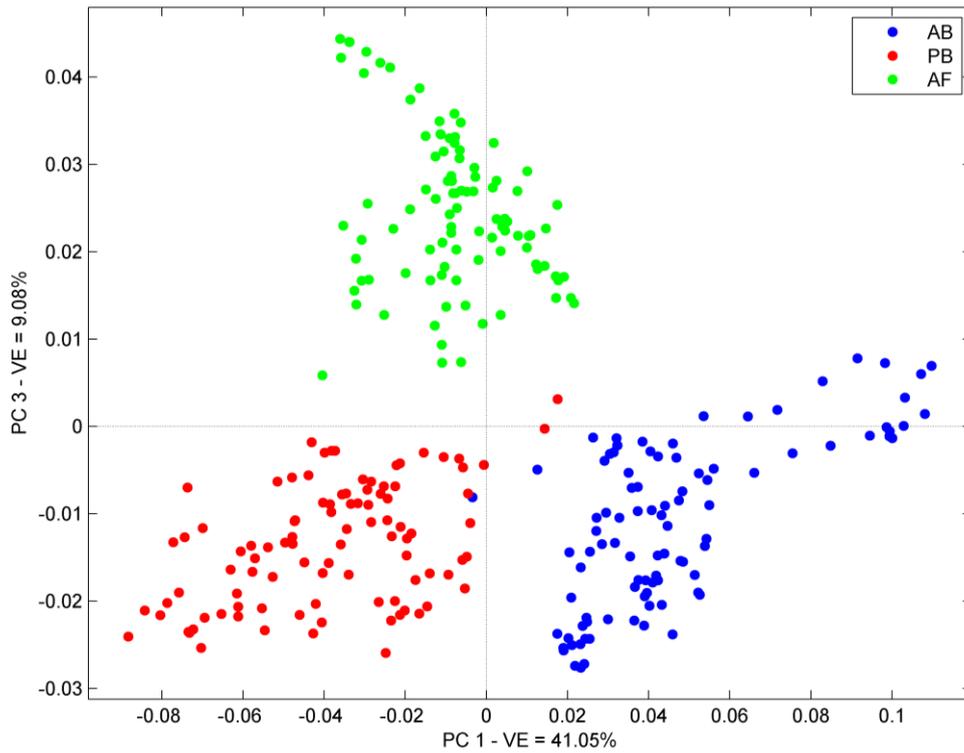


Figura 12. Gráfico dos escores PC1xPC3 da PCA com os dados obtidos sob o papel 63Br.

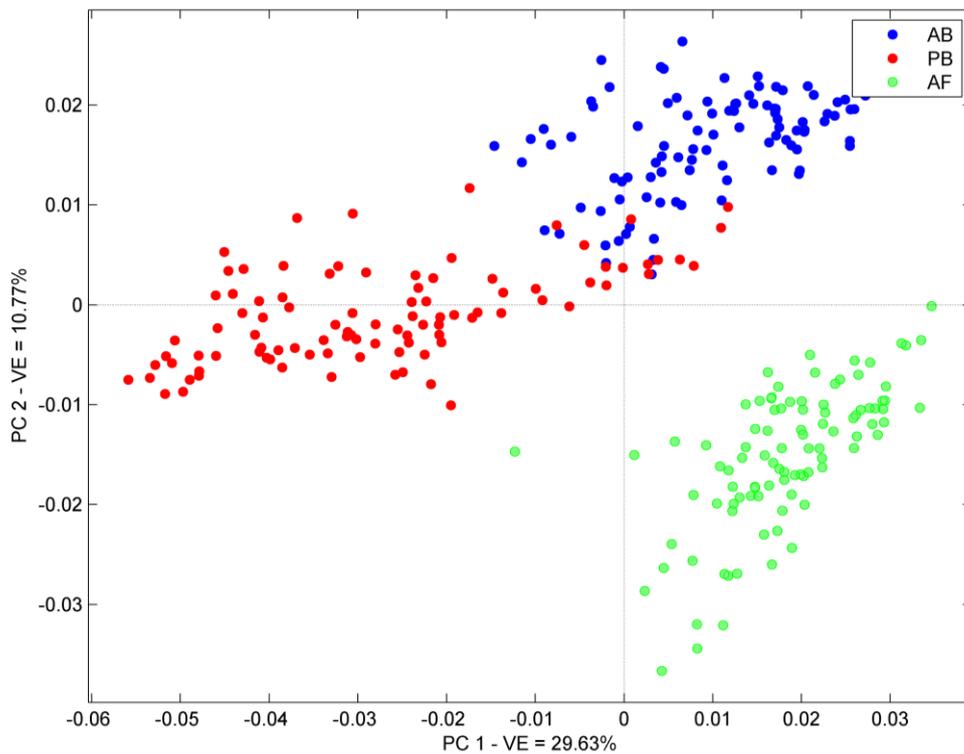


Figura 13. Gráfico dos escores PC1xPC2 da PCA com os dados obtidos sob o papel 75Arg.

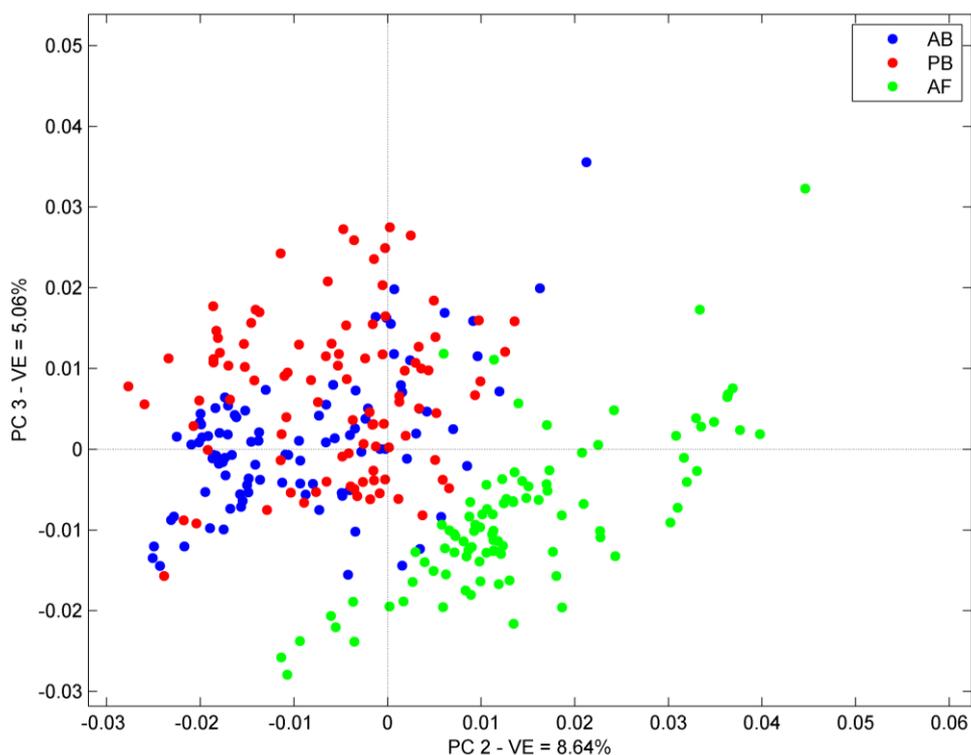


Figura 14. Gráfico dos escores PC2xPC3 da PCA com os dados obtidos sob o papel 75Rec.

Um estudo com os diferentes papéis juntos foi feito a partir dos dados pré-processados. Por limitações do pacote toolbox utilizado, não foi possível analisar as 12 classes simultaneamente (3 canetas x 4 papéis), por isso foram feitos alguns subgrupos.

Primeiro foi avaliado a separação das 3 diferentes canetas nos papéis brancos. A PC1 separou o papel de origem argentina dos brasileiros, enquanto a PC2 separou os papéis brasileiros de diferente gramatura. Esse resultado pode ser observado na Figura 15 que possui a representação dos escores de PC1xPC2. Para obter a separação em função das canetas foi necessário o emprego da PC3, a qual possui tendência em separar as canetas da mesma marca, e da PC4, onde têm-se bem distinto o grupo da marca diferente para o lado negativo (coloração verde). A Figura 16 apresenta os escores de PC3xPC4.

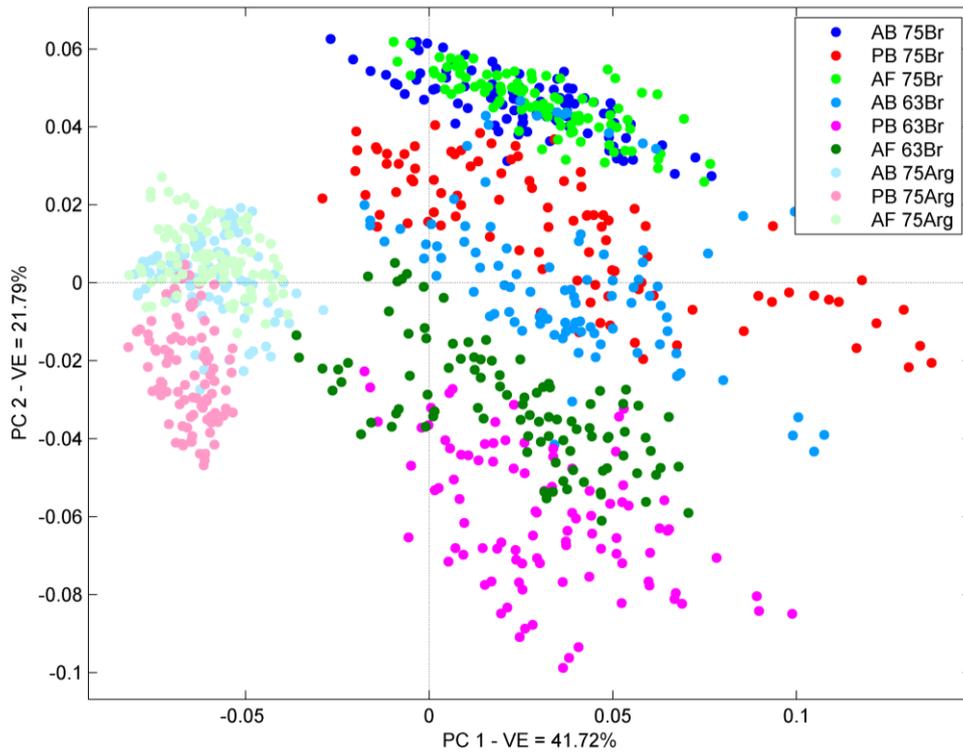


Figura 15. Gráfico dos escores PC1xPC2 da PCA com os dados obtidos sob os papéis brancos.

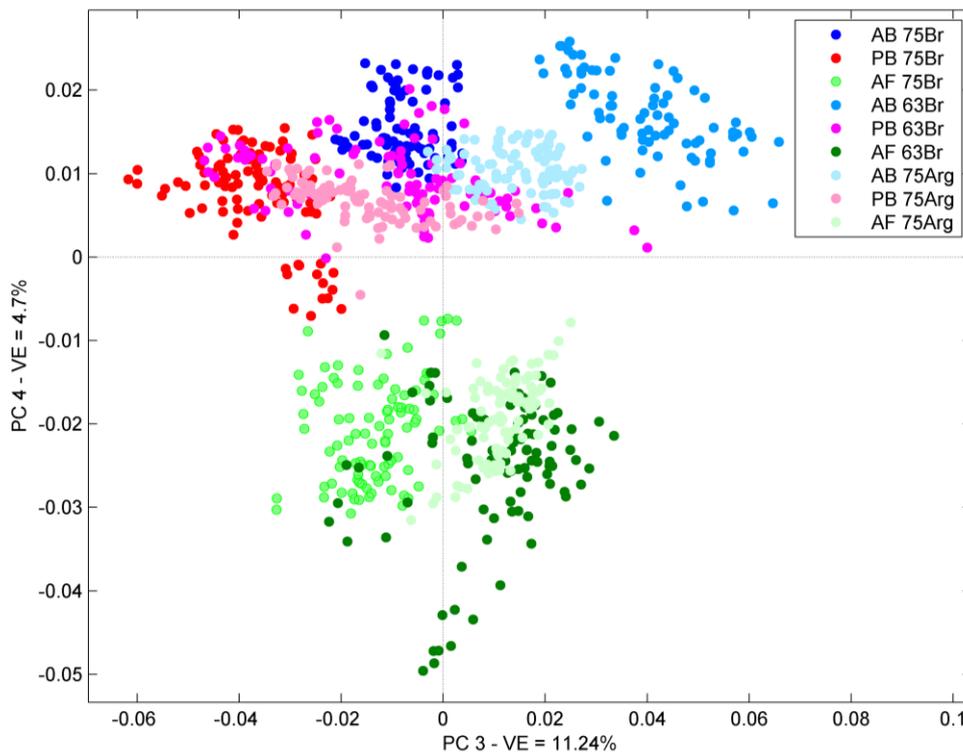


Figura 16. Gráfico dos escores PC3xPC4 da PCA com os dados obtidos sob os papéis brancos.

Esse mesmo estudo anterior foi feito com os papéis de mesma nacionalidade, ou seja, substituindo o papel de origem argentina pelo nacional reciclado. A PC1 separa bem o papel reciclado dos papéis brancos nacionais, de forma análoga ao estudo anterior, porém agora tendendo os papéis brancos para o lado negativo dessa PC. Já a PC3 tem boa capacidade de separar as diferentes gramaturas e por isso, para diferenciação dos papéis, é apresentada a Figura 17 com os escores de PC1xPC3.

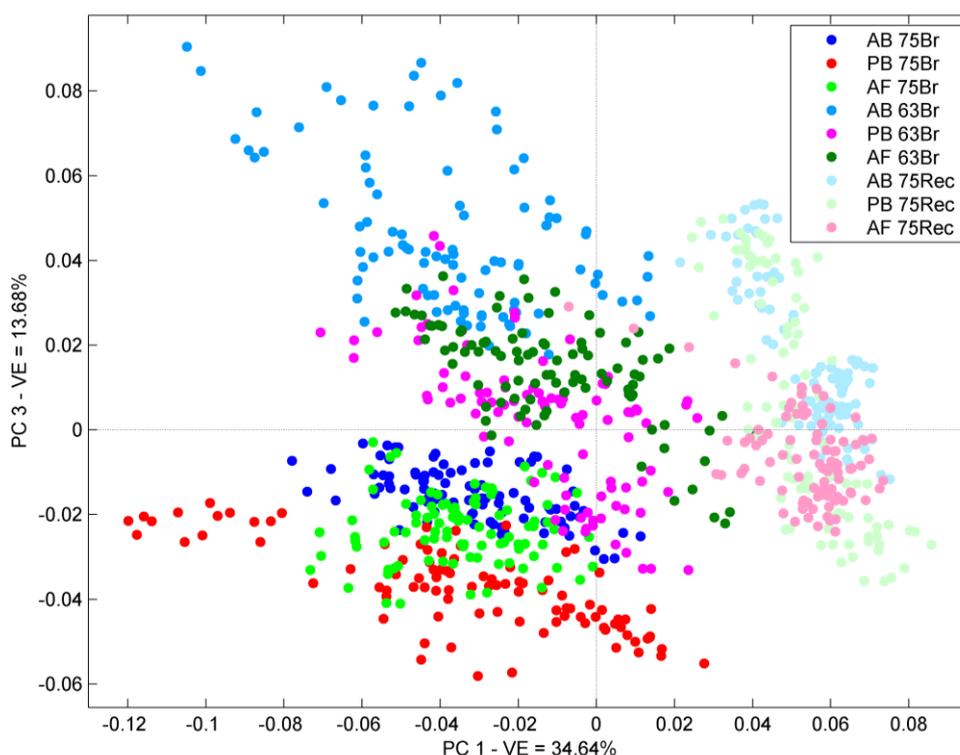


Figura 17. Gráfico dos escores PC1xPC3 da PCA com os dados obtidos sob os papéis nacionais.

Uma separação das canetas de mesma marca não foi possível em nenhuma PC, entretanto a PC4 destaca claramente a caneta de marca diferente. Na Figura 18 são apresentados os escores da PC4 e optou-se por mostrar junto os escores da PC2, que não havia ainda aparecido para essa combinação de dados.

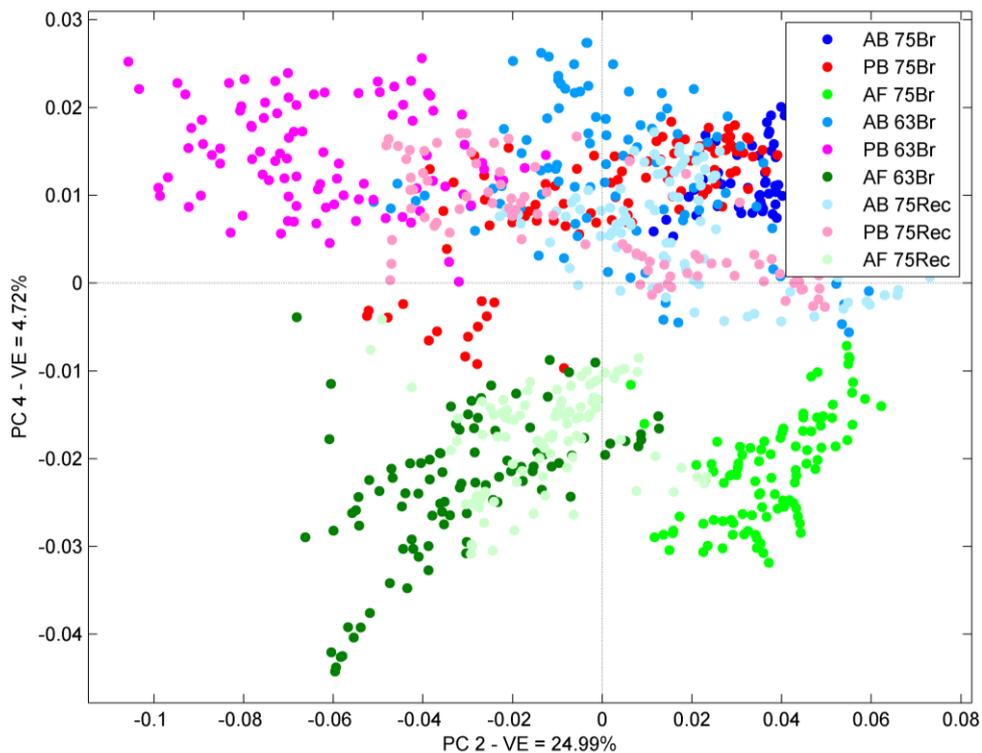


Figura 18. Gráfico dos escores PC2xPC4 da PCA com os dados obtidos sob os papéis nacionais.

Por último agrupou-se os papéis de mesma gramatura, ou seja, 75Br, 75Arg e 75Rec. Neste caso, conforme apresentado na Figura 19, a PC1 distingue com sucesso o papel branco de nacionalidade brasileira dos demais, enquanto a PC4 tem sucesso na distinção da caneta de marca distinta. Essa foi a melhor separação encontrada com o uso de PCA nesse grupo de amostras. Na figura 20 podemos ver que a PC2 apresenta uma tendência em separar as amostras da caneta de coloração preta das demais, mas sem sucesso para os traços sobre o papel reciclado, enquanto a PC3 não apresenta contribuição clara para esse estudo.

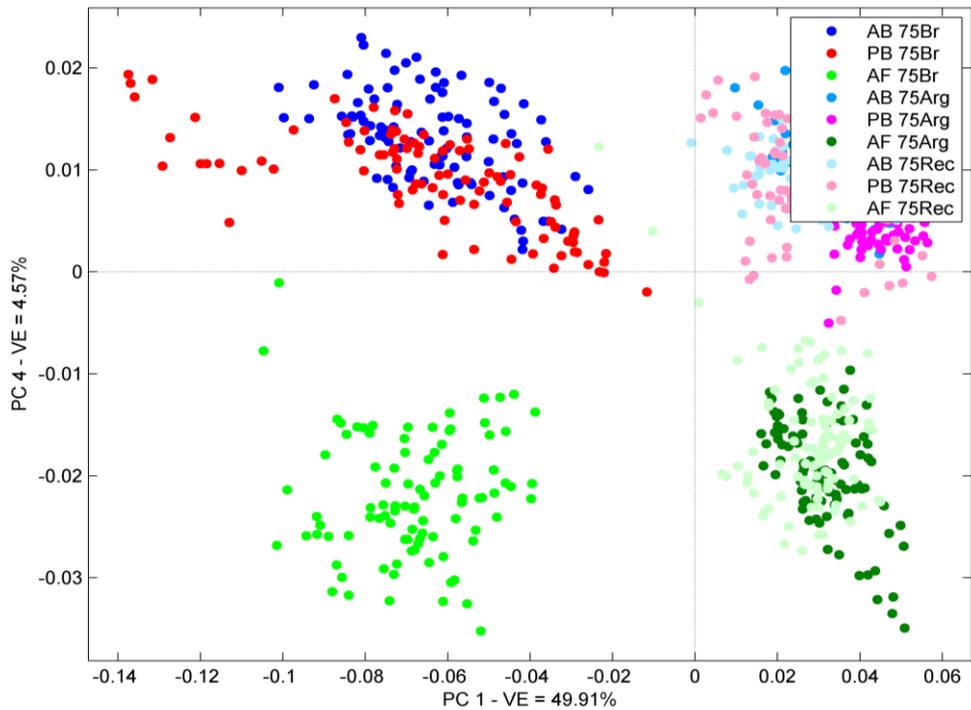


Figura 19. Gráfico dos escores PC1xPC4 da PCA com os dados obtidos sob os papéis de mesma gramatura.

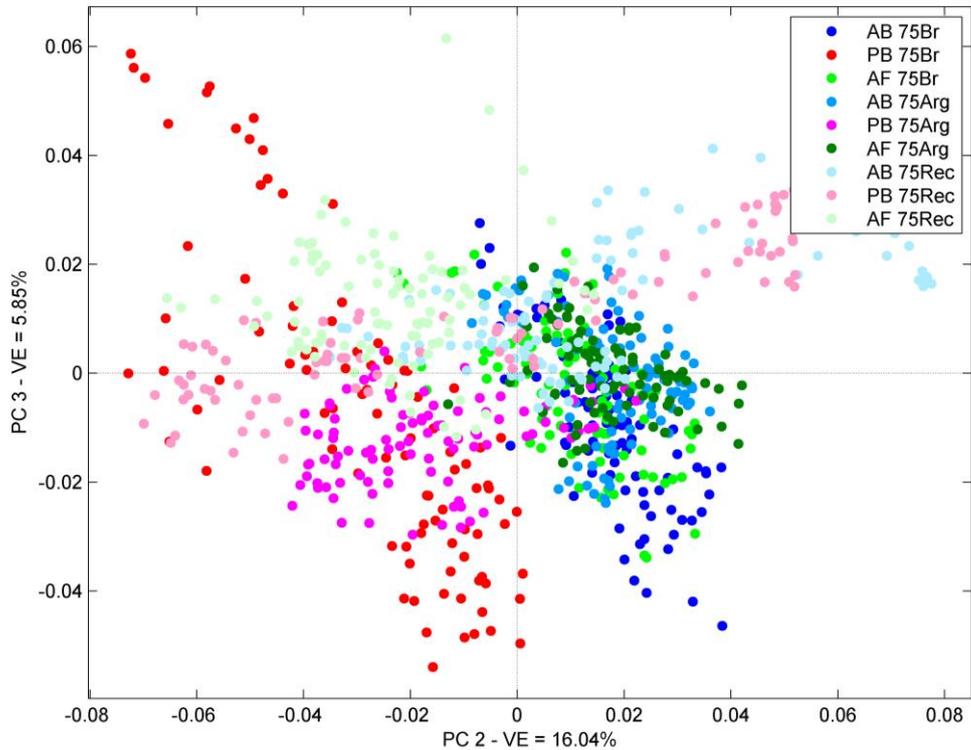


Figura 20. Gráfico dos escores PC2xPC3 da PCA com os dados obtidos sob os papéis de mesma gramatura.

4.2.2 Análise discriminante linear

Os dados espectrais pré-processados (após recorte, suavização, normalização e 1ª derivada) foram avaliados separadamente como dois grupos: considerando apenas 3 classes, correspondentes às canetas, e considerando 4 classes, correspondentes aos papéis. Para modelagem utilizou-se metade das amostras, sendo o restante empregado no teste dos modelos. Para a construção dos modelos de LDA, comparou-se a utilização dos algoritmos de seleção de variáveis SPA e GA. As configurações de cada modelo foram as seguintes: para SPA, tamanho de cadeia de variáveis de 1 a 13; e para GA, população de 50 indivíduos, número de gerações de 100, taxa de mutação de 5%, número máximo de variáveis selecionadas de 20 e taxa de cruzamento de 60, sendo escolhido o melhor desempenho após 3 modelagens.

Primeiramente todas as amostras foram consideradas como tendo única diferença entre as classes as canetas utilizadas e estudou-se o desempenho de dois diferentes algoritmos de seleção de variáveis diferentes, SPA e GA. Nos dois modelos gerados a LDA apresentou bons valores nas figuras de mérito, com ótimos resultados para a classe da caneta de marca distinta, resultando em valores acima de 95%. Os modelos apresentaram uma leve confusão para as classes referentes às canetas da mesma marca, comportamento semelhante à PCA, resultando em valores acima de 80%, conforme é apresentado na Tabela 2.

Tabela 2. Figuras de mérito LDA grupo canetas.

	Ac ¹		AB	PB	AF
SPA-LDA Treinamento	89,76	Sen ²	91,15	82,29	95,83
		Esp ³	91,41	94,79	98,44
SPA-LDA Teste	88,19	Sen ²	85,94	81,77	96,88
		Esp ³	90,62	92,19	99,48
GA-LDA Treinamento	95,14	Sen ²	97,40	90,52	97,40
		Esp ³	95,31	97,40	100
GA-LDA Teste	94,62	Sen ²	96,88	88,02	98,96
		Esp ³	94,01	97,92	100

¹Acurácia (%); ²Sensibilidade (%); ³Especificidade (%)

Um modelo de LDA com 3 classes resulta em duas funções discriminantes de Fischer. Na Figura 21 pode-se ver pelos escores da SPA-LDA que a primeira função discriminante separa para seu lado negativo a classe AF, correspondente a marca diferente, e em seu lado positivo ficam as classes AB e PB, que pertencem a mesma marca. Ao avaliar a Figura 22b vemos que, para a primeira função discriminante, uma variável selecionada que tem alta contribuição para o lado positivo está em $\sim 1250\text{ cm}^{-1}$, correspondente à algum composto presente na tinta das canetas da marca BIC, que a diferencia da Faber-Castell. Segundo Sharma e Kumar (2017)¹⁷, essa absorção em $\sim 1250\text{ cm}^{-1}$ corresponde à vibração de éter aromático e está associada ao solvente 2-fenoxietanol, utilizado em tintas. Pela Figura 22a essa variável não se mostra tão intensa, mas quimicamente muito importante para a distinção das classes. A segunda função discriminante é muito útil para a separação das classes de mesma marca, mostrando tendência em agrupar em seu lado positivo a classe AB e, no lado negativo, a classe PB. Associando esse agrupamento com a Figura 22 vemos que para o lado positivo da FD2 quem contribui é um sinal bastante intenso em $\sim 1575\text{ cm}^{-1}$, provavelmente corresponde à vibração do grupo amino proveniente do corante cristal violeta¹⁷, enquanto em $\sim 850\text{ cm}^{-1}$ temos um sinal, de intensidade também significativa, que contribui para o lado negativo da FD2 e pode ser relacionado com algum composto presente apenas ou majoritariamente na classe PB.

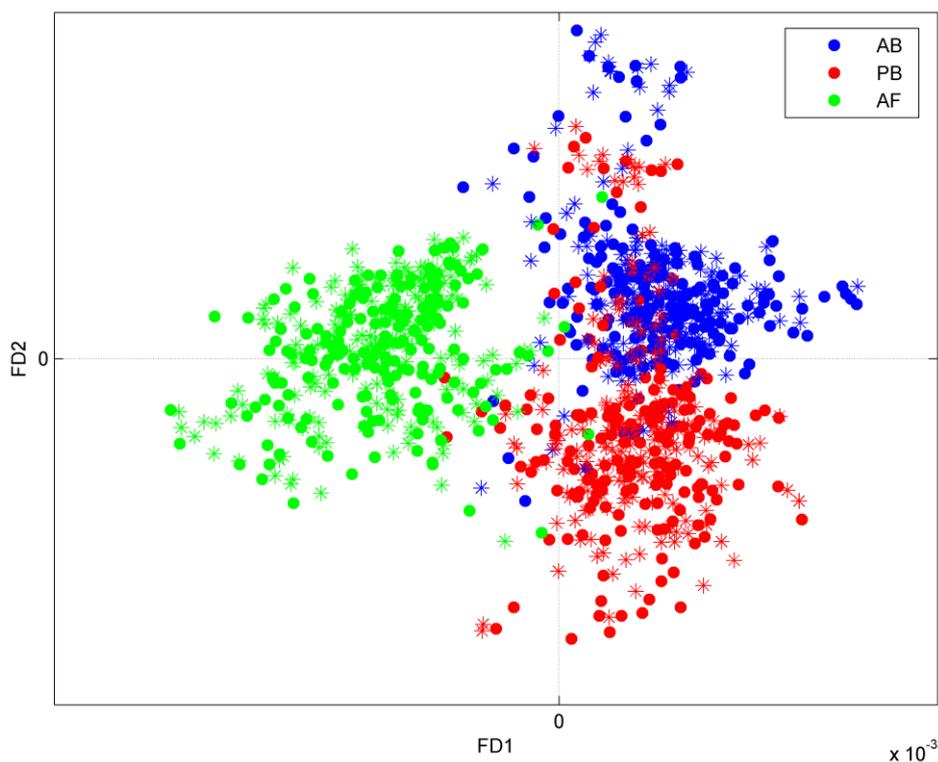


Figura 21. Gráfico dos escores $FD1 \times FD2$ da SPA-LDA com os dados classificados pelas tintas de caneta. ● corresponde as amostras de treinamento e * as amostras de teste.

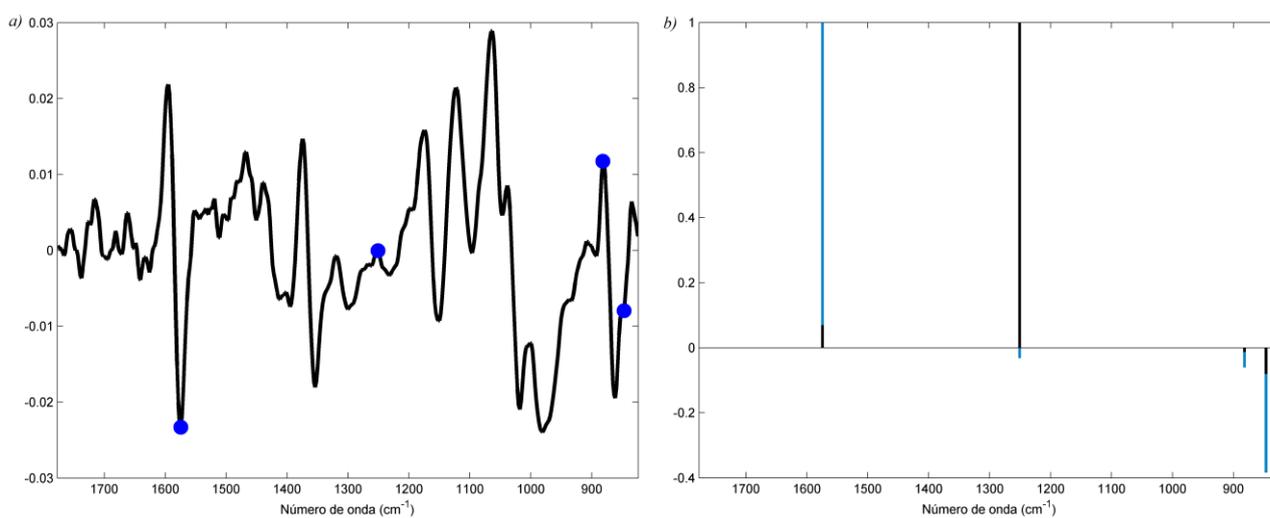


Figura 22. Dados classificados pelas tintas de caneta. a) Seleção das variáveis mais relevantes do modelo SPA-LDA; b) Gráfico de pesos das funções discriminantes 1 (preto) e 2 (azul) da SPA-LDA.

O modelo GA-LDA apresentou figuras de mérito sutilmente melhores que o SPA-LDA e resultados de escores e pesos muito semelhantes. Na Figura 23 estão os escores da GA-LDA onde podemos ver a primeira função discriminando a marca diferente e a segunda função discriminando as cores de mesma marca, assim como no modelo SPA-LDA.

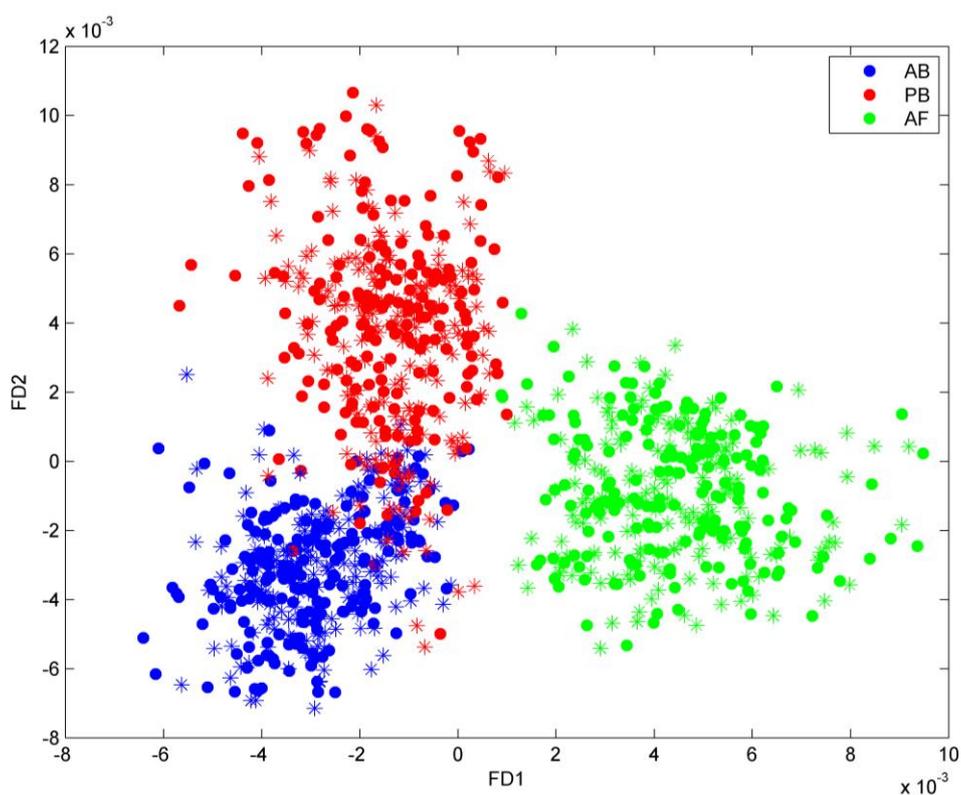


Figura 23. Gráfico dos escores FD1xFD2 do modelo GA-LDA com os dados classificados pelas tintas de caneta. ● corresponde as amostras de treinamento e × as amostras de teste.

As variáveis selecionadas pelo modelo GA-LDA, Figura 24a, mostram similaridade com o SPA-LDA na região de $\sim 1575 \text{ cm}^{-1}$ e $\sim 1250 \text{ cm}^{-1}$, e essas variáveis são responsáveis pela discriminação das mesmas classes nos dois modelos. Os pesos das funções discriminantes, Figura 24b, mostram que a banda em $\sim 1250 \text{ cm}^{-1}$ é responsável por segregar a marca BIC na FD1, assim como a banda em $\sim 1575 \text{ cm}^{-1}$ é responsável pela classe AB, mesmo comportamento já descrito para o modelo SPA-LDA.

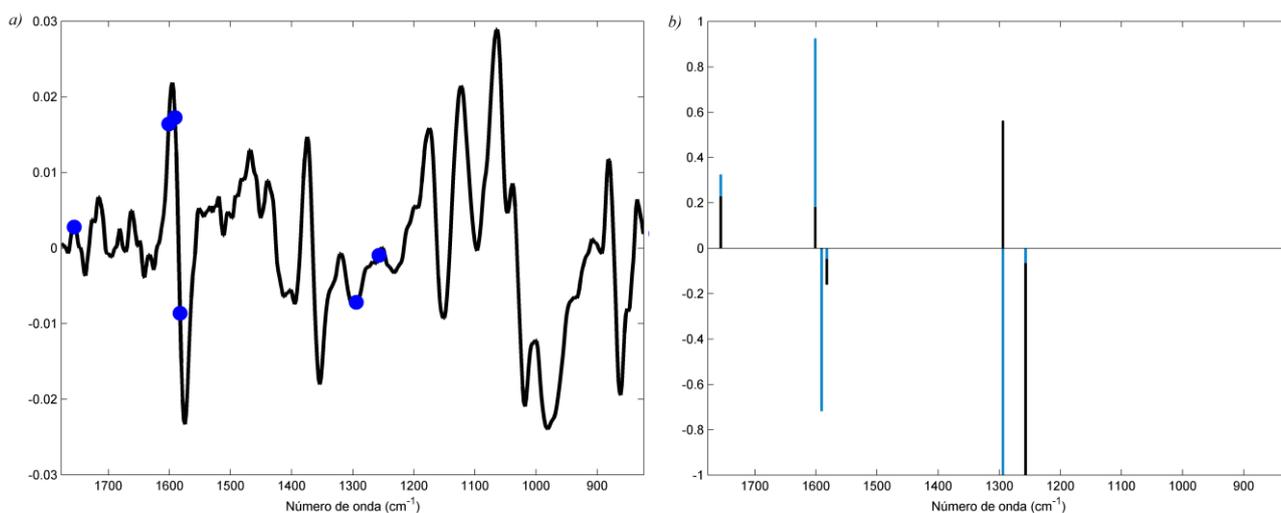


Figura 24. Dados classificados pelas tintas de caneta. a) Seleção das variáveis mais relevantes do modelo GA-LDA; b) Gráfico de pesos das funções discriminantes 1 (preto) e 2 (azul) da GA-LDA.

Este estudo demonstra a capacidade dos algoritmos SPA e GA em selecionar o mínimo necessário de variáveis para caracterizar as classes, assim como da LDA em modelar as amostras de cada caneta como um único grupo, independente dos diferentes papéis utilizados. Possivelmente o uso do papel reciclado em algumas amostras foi o responsável pela confusão entre as classes de mesma marca e a não obtenção de resultados mais próximos de 100%.

Em um segundo estudo foi considerado apenas os papéis como a diferença entre as classes e também foi avaliado o desempenho dos algoritmos de seleção de variáveis SPA e GA. Nos dois modelos, a LDA apresentou uma acurácia acima de 90%, com ótimos resultados para as classes 75Br e 63Br, ambos papéis nacionais brancos com gramaturas diferentes, resultando em figuras de mérito com valores acima de 93%. Os modelos apresentaram uma pequena dificuldade para modelar as classes referentes ao papel de origem argentina e ao papel nacional reciclado, resultando em figuras de mérito a partir de 80%. Os valores exatos das figuras de mérito são apresentados na tabela 3.

Tabela 3. Figuras de mérito LDA grupo papéis.

	Ac ¹		75Br	63Br	75Arg	75Rec
SPA-LDA Treinamento	91,67	Sen ²	100	93,75	86,11	86,81
		Esp ³	100	99,77	95,37	93,75
SPA-LDA Teste	92,36	Sen ²	100	95,14	88,89	85,42
		Esp ³	100	100	94,44	95,38
GA-LDA Treinamento	96,53	Sen ²	100	98,61	93,06	94,44
		Esp ³	100	100	97,68	97,68
GA-LDA Teste	96,18	Sen ²	100	97,91	97,22	89,58
		Esp ³	100	100	96,30	98,61

¹Acurácia (%); ²Sensibilidade (%); ³Especificidade (%).

Um modelo de LDA com 4 classes resulta em 3 funções discriminantes. Pelo gráfico de escores da SPA-LDA para esse grupo, Figura 25, pode-se ver que a classe 75Br está separada das demais e localizada no lado positivo da primeira função discriminante. Pelos pesos das funções discriminantes, Figura 26b, em $\sim 845\text{ cm}^{-1}$ está a variável selecionada responsável por essa discriminação, que pode ser correspondente ao carbonato de cálcio utilizado no processo de alisamento de alguns papéis^{70,71}. A classe 63Br também está bem separada das demais, porém localizada na região central da FD1; as demais classes não são separadas pela FD1 e estão posicionadas no lado negativo dessa função. A separação das classes 75Arg e 75Rec foi obtida na terceira função discriminante. A FD3 tem quatro variáveis selecionadas que contribuem para o lado negativo, porém, a variável selecionada em $\sim 1700\text{ cm}^{-1}$ tem significativa contribuição negativa para FD3 e FD1, provavelmente sendo responsável pela classe do papel reciclado, que ficou posicionada no quadrante negativo para as duas classes. Ganzerla *et al.* (2009)⁷² indicam que essa banda pode ser proveniente da degradação da celulose (oxicelulose). Em $\sim 1600\text{ cm}^{-1}$ tem uma banda bem intensa, vista na Figura 26a, que, pela Figura 26b, tem contribuição, apesar de pequena, para o lado negativo da FD1 e positivo da FD3, estando possivelmente relacionada com a classe de nacionalidade argentina e correspondendo a presença de breu colofônia⁷⁰. A FD2 não colabora de forma significativa para a separação espacial das classes.

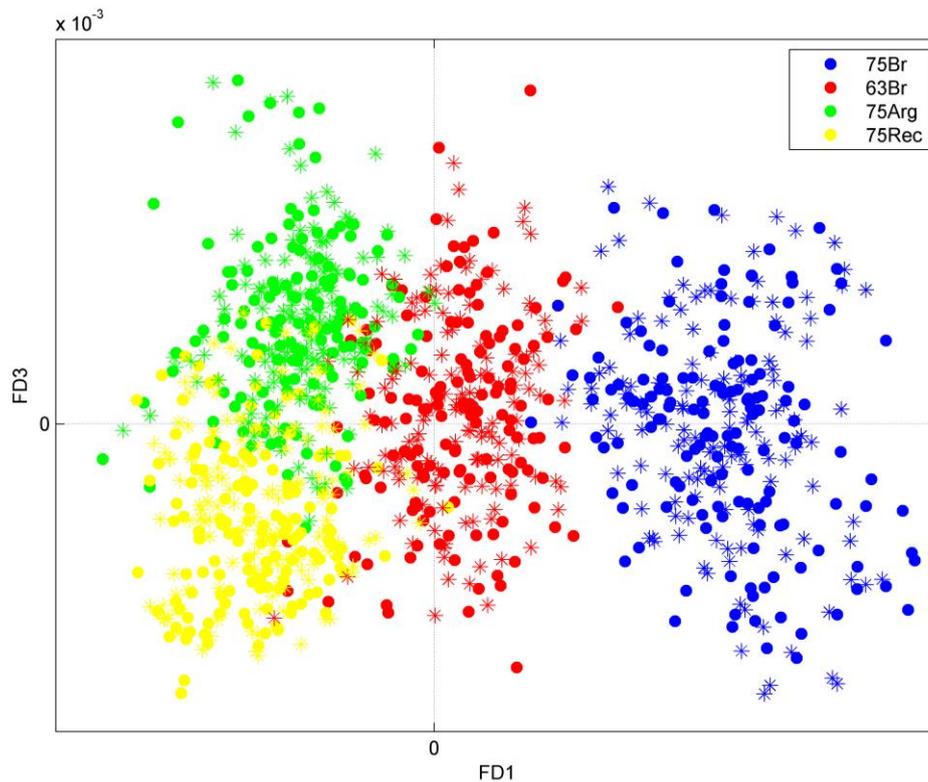


Figura 25. Gráfico dos escores FD1xFD3 do modelo SPA-LDA com os dados classificados pelos papéis. ● corresponde as amostras de treinamento e × as amostras de teste.

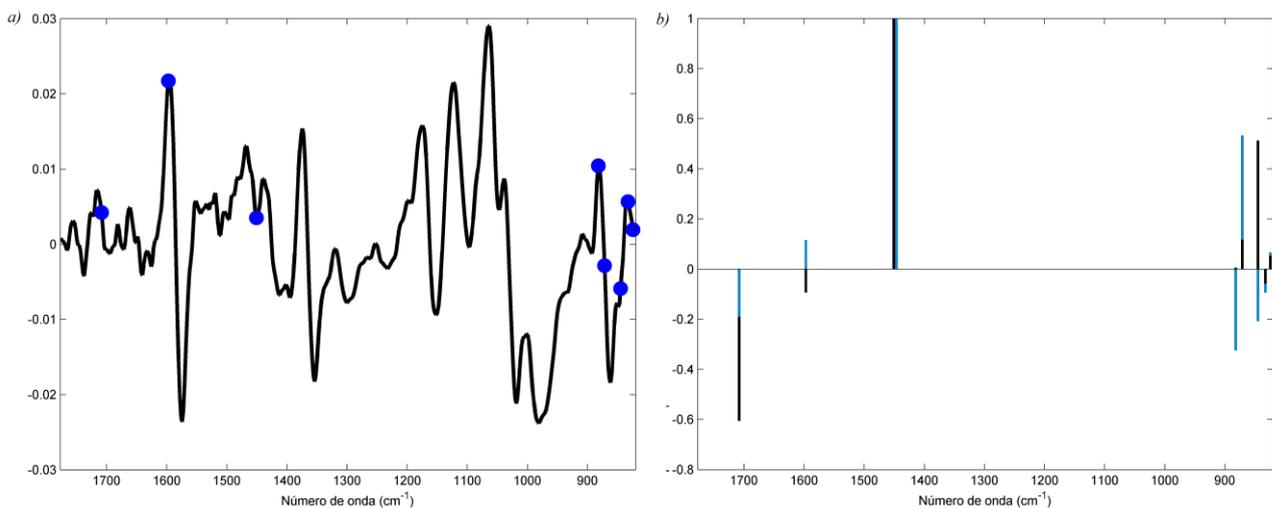


Figura 26. Dados classificados pelos papéis. a) Seleção das variáveis mais relevantes do modelo SPA-LDA; b) Gráfico de pesos das funções discriminantes 1 (preto) e 3 (azul) da SPA-LDA.

De forma semelhante e como pode ser visto no gráfico de escores, na Figura 27, o modelo GA-LDA também teve a classe 75Br discriminada para o lado positivo da FD1, enquanto a FD3 foi responsável por segregar as classes 75Arg e 75Rec. Entretanto, para esse modelo, a FD2 foi muito importante para a separação da classe 63Br, como pode ser visto na Figura 28. As variáveis selecionadas pelo modelo GA-LDA são apresentadas na Figura 29a, sendo algumas semelhantes ao modelo SPA-LDA, como por exemplo em $\sim 845\text{ cm}^{-1}$ no lado positivo da FD1, que segrega a classe 75Br. Na Figura 29b é possível ver que a variável mais importante para o lado negativo da FD2 está em $\sim 1420\text{ cm}^{-1}$ sendo correspondente ao estiramento assimétrico de carbonato, correspondendo ao CaCO_3 utilizado no alisamento de papéis^{71,73}.

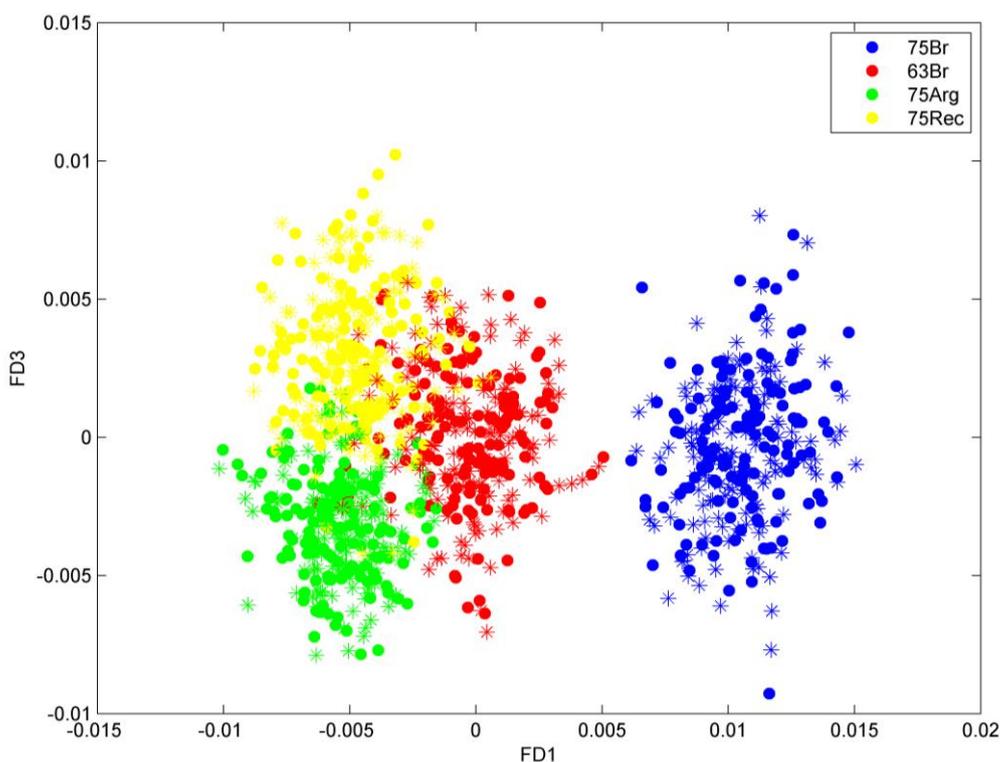


Figura 27. Gráfico dos escores FD1xFD3 do modelo GA-LDA com os dados classificados pelos papéis. ● corresponde as amostras de treinamento e ✕ as amostras de teste.

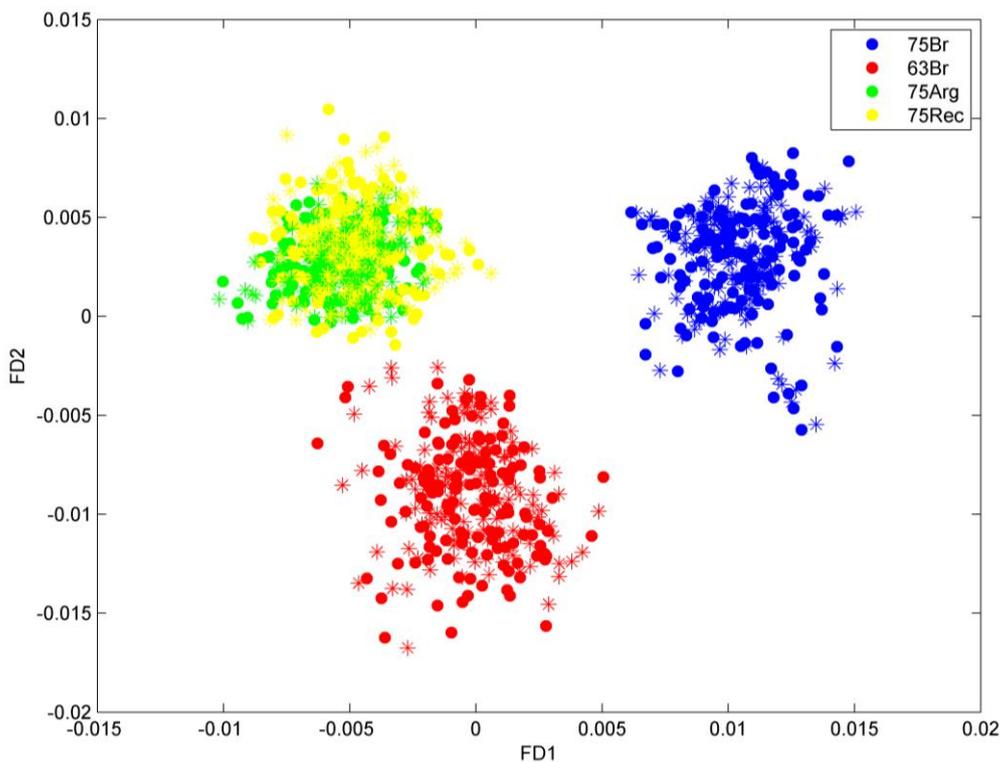


Figura 28. Gráfico dos escores FD1xFD2 do modelo GA-LDA com os dados classificados pelos papéis. ● corresponde as amostras de treinamento e × as amostras de teste.

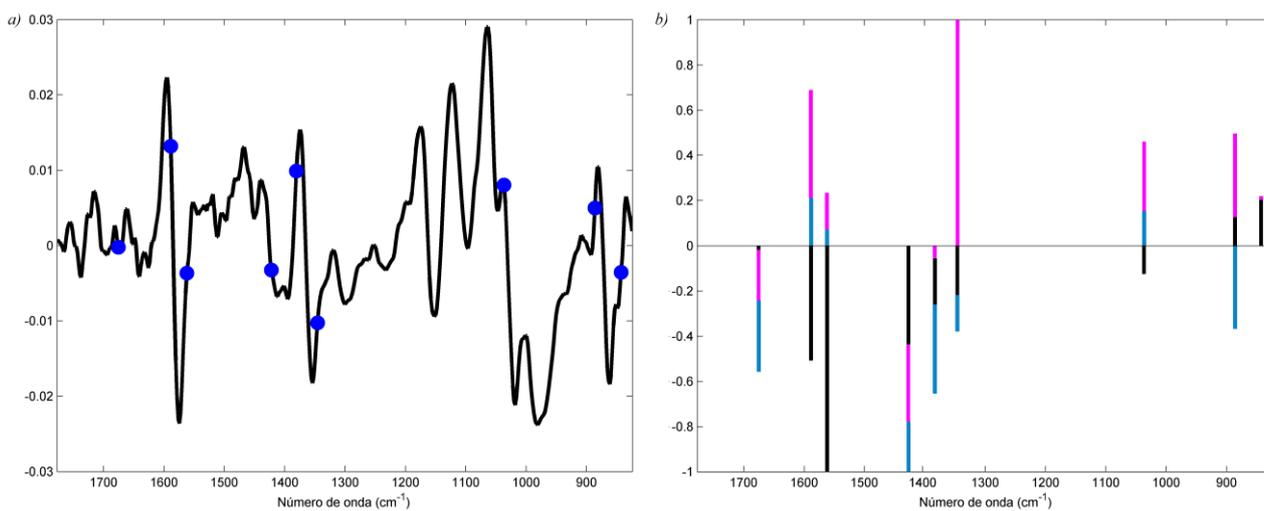


Figura 29. Dados classificados pelos papéis. a) Seleção das variáveis mais relevantes do modelo GA-LDA; b) Gráfico de pesos das funções discriminantes 1 (preto), 2 (azul) e 3 (rosa) da GA-LDA.

4.2.3 Análise discriminante por mínimos quadrados parciais

Os mesmos grupos, de canetas e de papéis, avaliados pela LDA foram modelados e testados por PLS-DA. Este modelo apresentou melhores resultados, com figuras de mérito acima de 96%, conforme dados da tabela 4.

De forma análoga ao LDA, os resultados para o grupo dos papéis foram superiores, havendo apenas 1 amostra que não foi associada a nenhuma classe na predição do modelo. Para o grupo das canetas, como esperado, o modelo apresenta melhores resultados para a classe AF, correspondente a marca distinta. A classificação de cada amostra para o grupo das canetas é mostrada na Figura 30 e, para o grupo dos papéis, na Figura 31.

O limite para cada classe aparece na forma de uma linha tracejada vermelha. Podemos ver que, para as classes dos papéis, as amostras são classificadas mais distantes desse limite, quando comparado ao resultado das classes de canetas, demonstrando que os espectros carregam maior informação química correspondente aos papéis, o que explica os melhores resultados de figuras de mérito para esse grupo.

Tabela 4. Figuras de mérito PLS-DA

	Ac ¹		AB	PB	AF	
Canetas Treinamento	98,61	Sen ²	100	96,88	98,96	
		Esp ³	99,74	100	100	
Canetas Teste	99,48	Sen ²	100	98,44	100	
		Esp ³	99,22	100	100	
	Ac ¹		75Br	63Br	75Arg	75Rec
Papéis Treinamento	100	Sen ²	100	100	100	100
		Esp ³	100	100	100	100
Papéis Teste	99,83	Sen ²	100	100	100	99,31
		Esp ³	100	100	100	100

¹Acurácia (%); ²Sensibilidade (%); ³Especificidade.

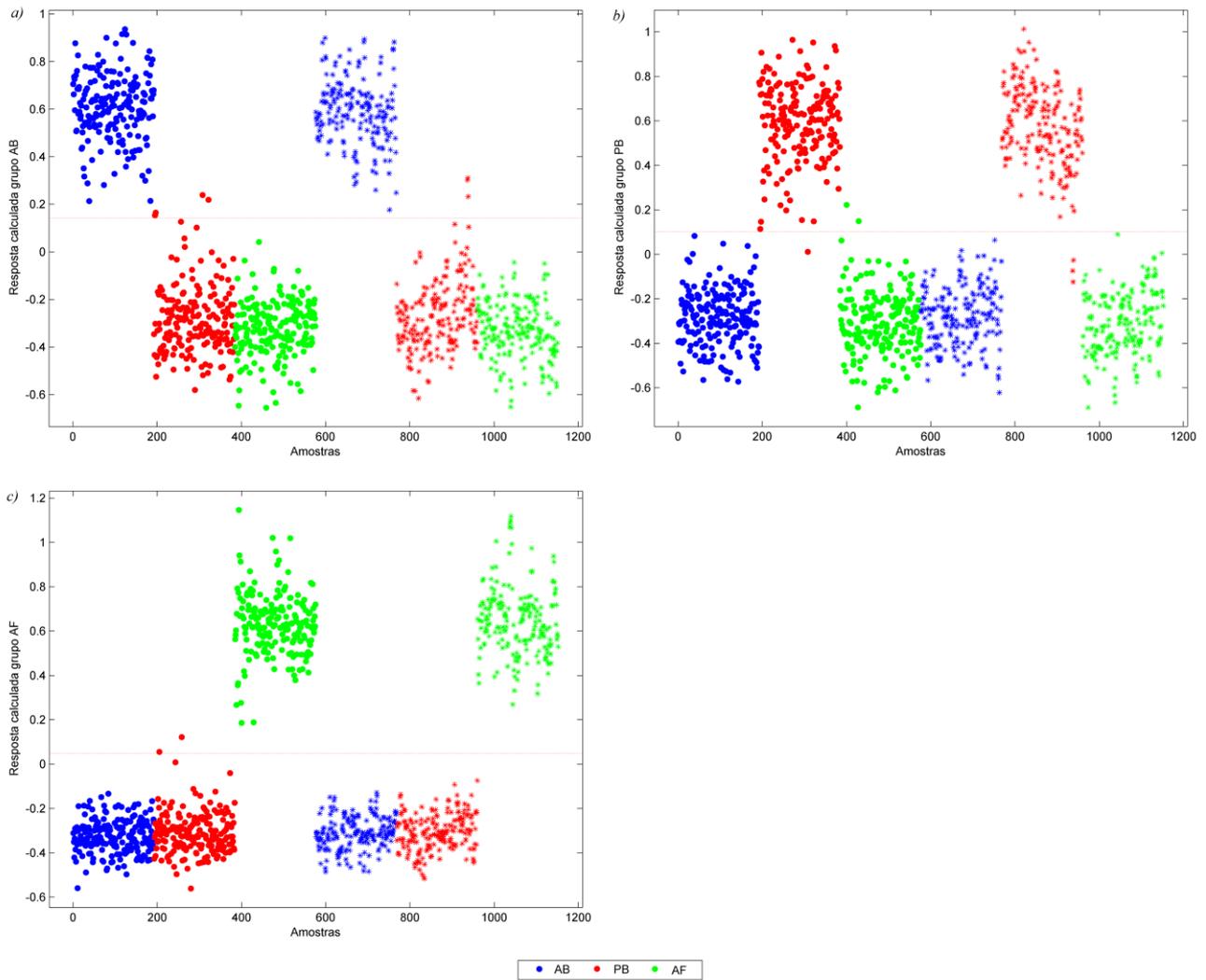


Figura 30. Classificação das amostras por PLS-DA considerando as canetas como classes. ● corresponde as amostras de treinamento e × as amostras de teste.

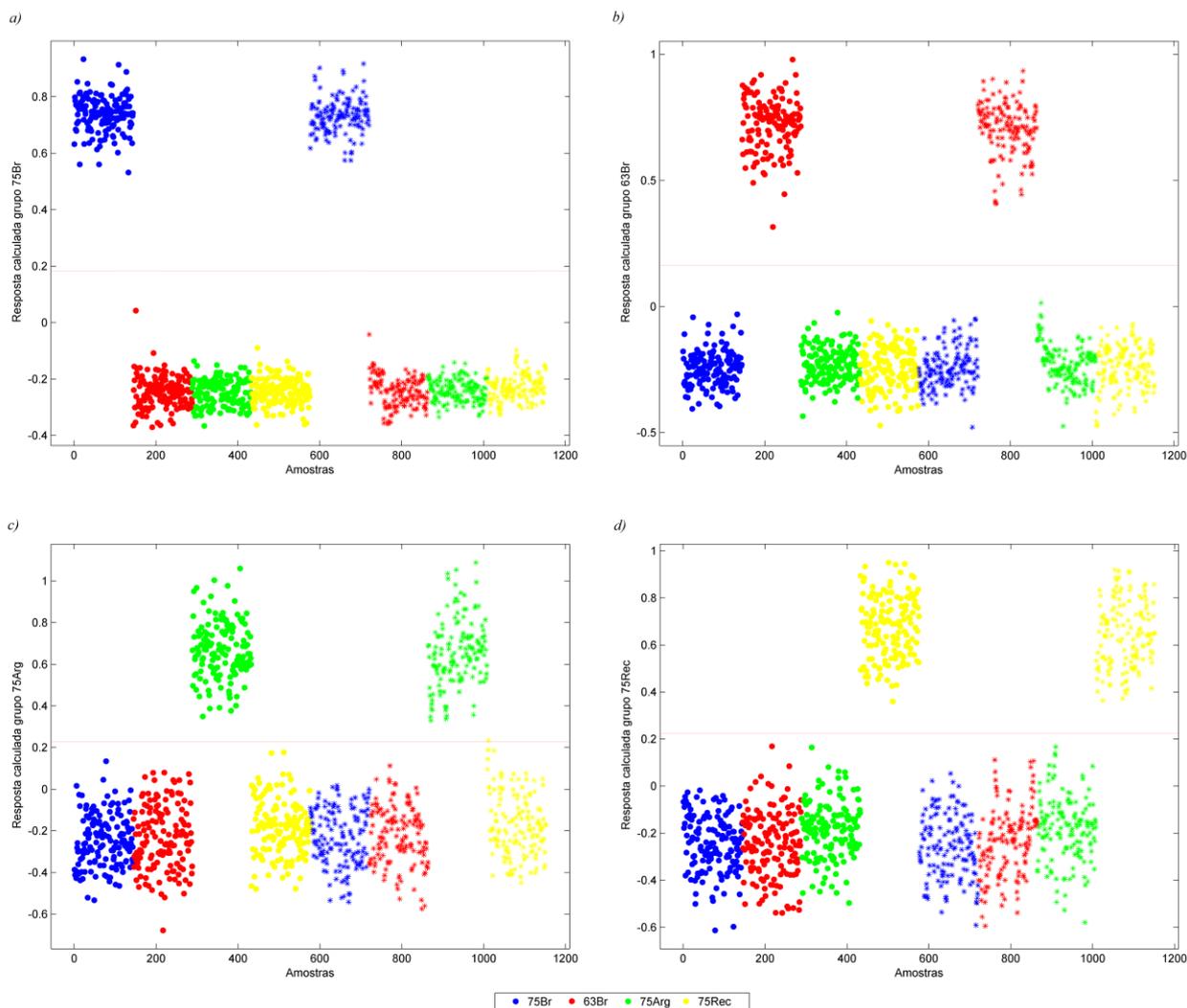


Figura 31. Classificação das amostras por PLS-DA considerando os papéis como classes. ● corresponde as amostras de treinamento e × as amostras de teste.

Com este estudo vê-se que é possível classificar diferentes tintas de caneta, assim como agrupar as amostras pelos papéis, mesmo considerando que os espectros de FTIR foram obtidos pela leitura dos traços de caneta. Isso vai ao encontro do relatado por Sharma *et al.* sobre a grande dificuldade em comparar amostras controle de tinta de carimbo (pura) com amostras dessa tinta impressas no papel, analisadas por ATR-FTIR, pela elevada contribuição desse substrato²⁴. Wang *et al.* (2012)⁷⁴ ao analisar tintas de carimbo também observaram interferências significativas do papel e, inclusive, optaram por procedimento de subtração dos espectros do papel em branco.

5. Envelhecimento de impressões digitais latentes

5.1 Materiais e métodos

5.1.1 Amostras

Impressões digitais de 19 pessoas do sexo masculino foram coletadas em duplicata, constituindo dois grupos amostrais, sendo um deles armazenado ao abrigo da luz e o outro exposto à luz. Os voluntários foram orientados a não utilizar cosméticos nas 24 horas anteriores a coleta. No momento da coleta todos lavaram as mãos com o mesmo sabonete neutro, secaram e aguardaram 10 minutos sem encostar em nada. Após, friccionaram o dedo polegar sobre a testa durante 3 segundos e então a impressão digital foi depositada em lâmina de vidro medindo 1 cm².

5.1.2 Espectroscopia de infravermelho médio

As amostras foram analisadas em um espectrofotômetro de infravermelho com transformada de Fourier (FTIR) Bruker Vertex 70 utilizando método de refletância com ângulo de 40°, com 64 segundos e 64 varreduras, na região espectral de 4000 a 400 cm⁻¹, com resolução de 2 cm⁻¹. As medições foram realizadas no dia da coleta (D0), 7 dias após a coleta (D7) e 30 dias após a coleta (D30), sendo em cada tempo escolhido três diferentes pontos em cada amostra para obtenção da triplicata. As amostras armazenadas com exposição à luz são identificadas com a letra C (de claro) antes da sigla correspondente ao tempo. De mesmo modo, as amostras armazenadas ao abrigo da luz são identificadas com a letra E (de escuro).

5.1.3 Pré-processamentos e análise multivariada

Os espectros de FTIR foram recortados na faixa de 1900 a 700 cm⁻¹. As amostras foram submetidas à uma HCA, utilizando o software ChemoStat⁷⁵ com o objetivo de selecionar os pares mais semelhantes. Após a seleção das amostras foi feita a média dos espectros das duplicatas.

Utilizando o software MatLab⁶⁶ e pacotes tolbox^{67,68,69} os dados foram suavizados por Savitzky-Golay (13 pontos e polinômio de segunda ordem),

normalizados e na sequência foi aplicada a 1ª derivada (13 pontos e polinômio de segunda ordem). Os dados pré-processados foram analisados pelo método não supervisionado PCA. O grupo amostral foi dividido em dois subgrupos utilizando o algoritmo Kennard-Stone⁴⁶ e então modelos de classificação foram construídos empregando os métodos supervisionados LDA e PLS-DA.

5.2 Resultados e discussão

O perfil espectral bruto das amostras obtidas pode ser observado na Figura 32.

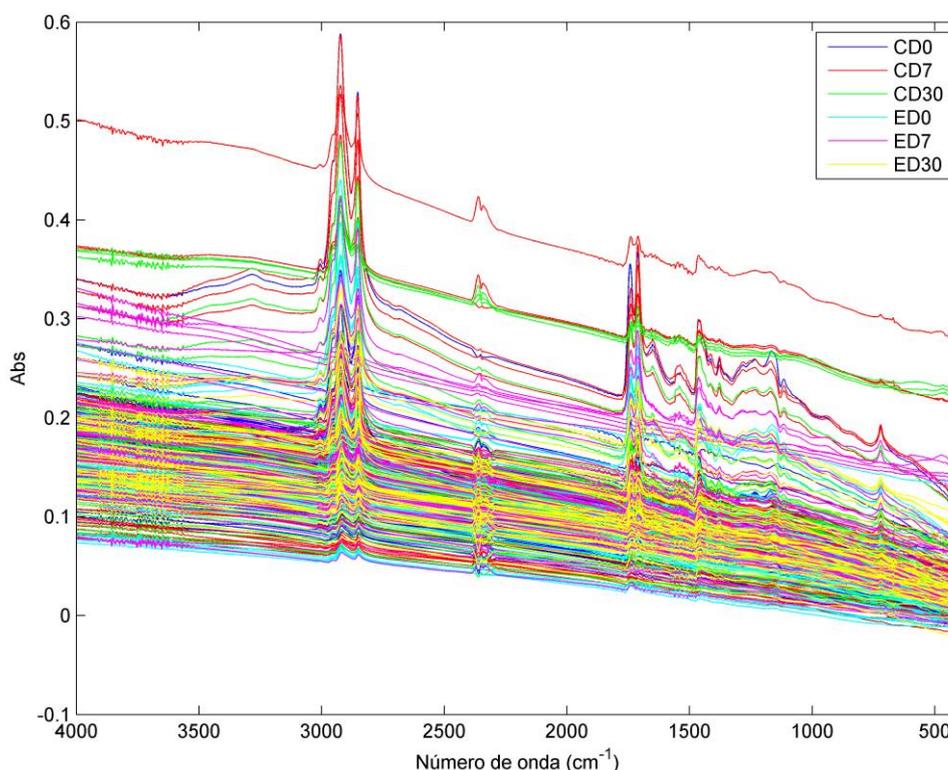


Figura 32. Espectros FTIR das amostras sem pré-processamento. C=Claro, amostras expostas à luz; E=Escuro, amostras armazenadas ao abrigo da luz; D0=Dia da coleta; D7=7 dias após coleta; D30=30 dias após coleta.

Os espectros foram recortados entre 1900 e 700 cm^{-1} por ser essa a faixa de impressão digital espectral, isto é, a região mais informativa e característica de cada estrutura molecular²⁷. As amostras foram previamente separadas considerando condições de armazenamento (claro/escuro) e tempo de análise. Cada um desses conjuntos amostrais foi submetido a uma análise por HCA

com o objetivo de identificar os pares de replicatas mais semelhantes. A título ilustrativo, o dendrograma do grupo de amostras mantidas no claro (C) e no dia (D) zero (CD0) é apresentado na Figura 33 com a indicação dos 8 pares de amostras que ficaram lado-a-lado, sendo essas selecionadas como as replicatas mais similares e que seguiram como parte do grupo amostral. O grupo CD0 foi, portanto, o que retornou menor número de pares de amostras, isto é, aquele que apresentou amostras mais dissemelhantes ou com maior variabilidade na sua composição espectroscópica. Nos demais grupos foram selecionados 10 (CD7 e ED0), 11 (ED7 e ED30) e 14 (CD30) pares de amostras. De cada um desses pares foi obtido o espectro médio, totalizando 64 amostras.

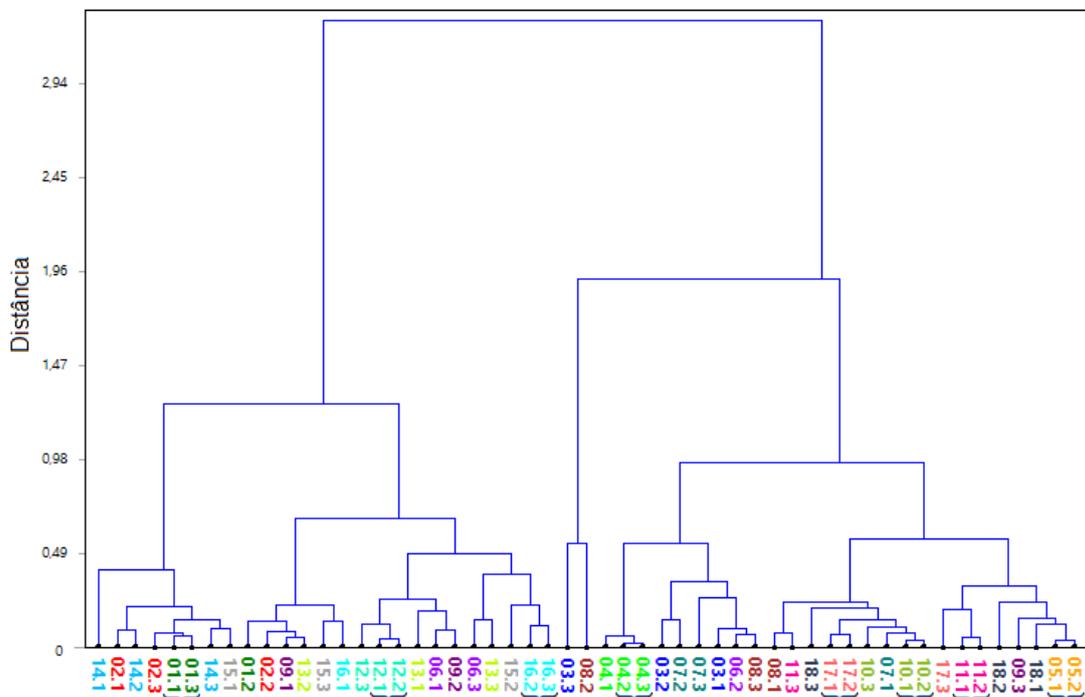


Figura 33. Dendrograma das amostras pertencentes ao grupo CD0.

Como pré-processamento os espectros ainda foram suavizados, normalizados e foi aplicada a 1ª derivada. O espectro resultante dos pré-processamentos é mostrado na Figura 34.

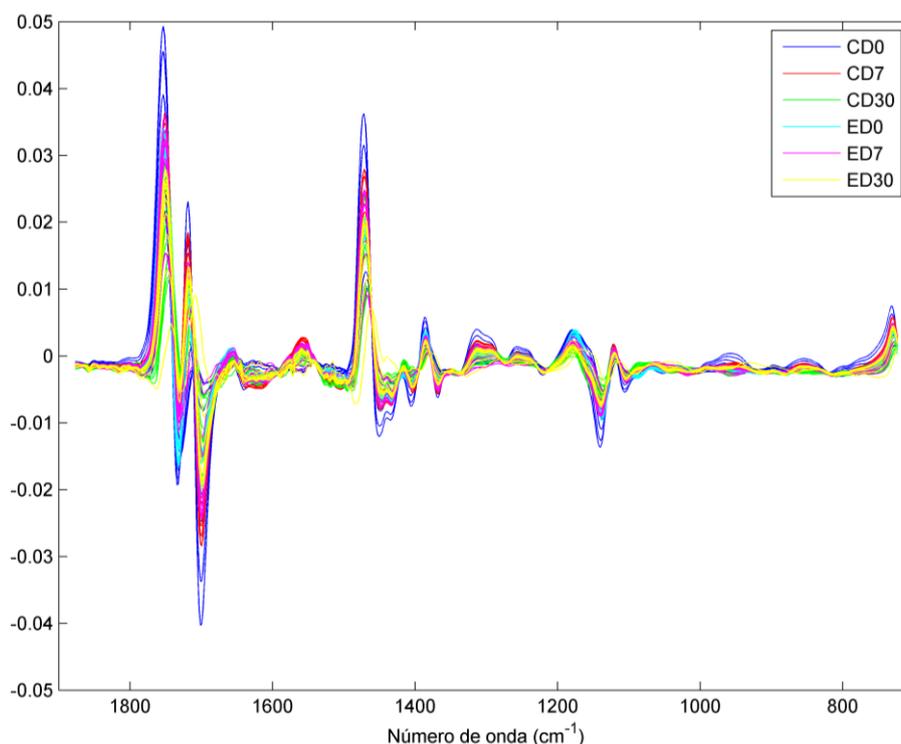


Figura 34. Espectros FTIR dos dados após recorte, suavização, normalização e derivada.

5.2.1 Análise de componentes principais

Os espectros pré-processados foram avaliados por PCA, com os dados centrados na média. Na Figura 35 são apresentados os escores da primeira componente principal, que não apresenta clara contribuição para a segregação das classes, e da segunda componente principal que demonstra uma tendência em separar os dados pelo tempo de armazenamento. Para o lado negativo da PC2 estão agrupados os dados referentes ao tempo zero, enquanto da região central para o lado positivo ficaram os demais dados, sendo possível ver uma tendência de separação entre os três tempos. As classes ED0 e ED7 estão mais próximas entre si que as classes CD0 e CD7, indicando maior similaridade entre as amostras de tempos iniciais armazenadas no escuro, ou seja, essa condição de armazenamento preservou as amostras. De forma semelhante, as classes CD7 e CD30 estão mais próximas que os mesmos tempos armazenados no escuro, resultados que indicam uma potencial fotodegradação com consequente perda informações químicas nas amostras expostas à luz. Resultados semelhantes foram encontrados por Souza *et al.*

(2022)³ que diferenciou impressões digitais latentes de homens e mulheres, empregando espectroscopia Ramam. As classes CD30 e ED30 não são classificadas separadamente pela PCA, possivelmente por terem atingido similar teor de degradação, demonstrando que, para maiores tempos de armazenamento, o abrigo ou a exposição à luz podem ser indiferentes na conservação das amostras. Girod et al. avaliaram o perfil de envelhecimento de resíduos de impressões digitais e demonstraram que aquelas armazenadas no escuro encontraram um padrão de envelhecimento até 36 dias, enquanto as amostras expostas à luz demonstram atingir um patamar de envelhecimento na metade desse tempo²⁸.

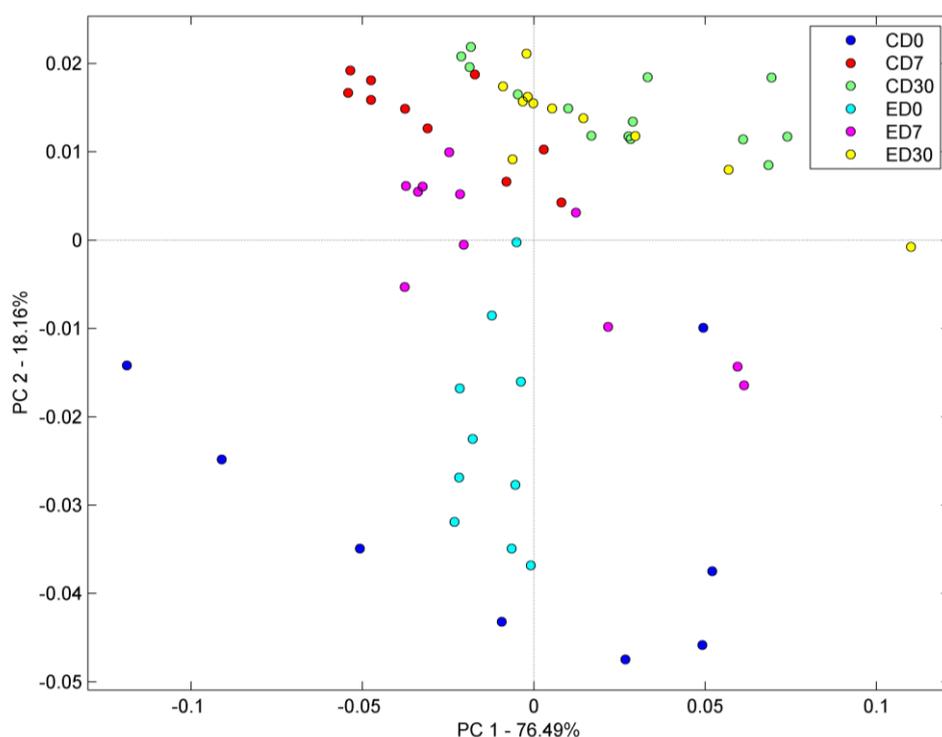


Figura 35. Gráfico dos escores PC1xPC2 da PCA.

Pela Figura 36 podemos ver que as duas primeiras bandas mais intensas, na faixa de 1780 a 1670 cm^{-1} , são as principais responsáveis por segregar as amostras do tempo zero em relação às demais. Nessa região têm-se absorção de energia para estiramento e deformação axial da ligação carbonila de éster, componente presente em impressões digitais latentes e proveniente de secreções sebáceas^{27,76,77}. Diversos autores demonstram que esses

componentes decaem ao longo do tempo por um período de aproximadamente 30 dias, sendo essa diminuição de concentração que possibilita diferenciar as classes ao longo do tempo^{3,28,78,79}.

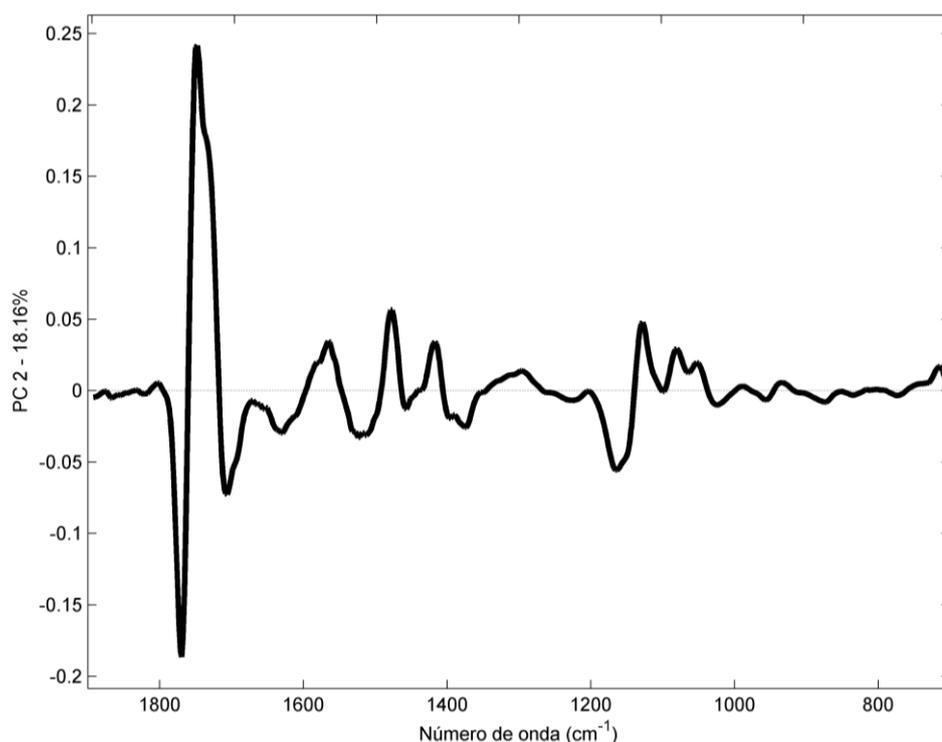


Figura 36. Gráfico de pesos da segunda componente principal da PCA.

5.2.2 Análise discriminante linear

Para a construção dos modelos de LDA, comparou-se a utilização dos algoritmos de seleção de variáveis GA, AOC, SW e SPA. As configurações de cada modelo foram as seguintes: para GA, população de 50 indivíduos, número de gerações de 100, taxa de mutação de 5%, número máximo de variáveis selecionadas de 20 e taxa de cruzamento de 60%; para AOC, o número de formigas foi 50, número de colônias de 100, taxa de formigas cegas de 35%, número máximo de variáveis selecionadas de 20 e taxa de evaporação de feromônio de 65%; para SW correlação máxima de 0,9, após teste de 0,2 a 0,9, com variação de 0,5; e para SPA, tamanho de cadeia de variáveis de 1 a 13. Para os modelos probabilísticos (GA e AOC) foi escolhido o melhor desempenho após 3 modelagens.

A LDA resultou em excelentes figuras de mérito com todos os algoritmos testados, conforme apresentado na Tabela 5. Para os modos estocásticos, GA e AOC, todas as métricas alcançaram 100%, porém muitas variáveis selecionadas estão na região abaixo de 1100 cm⁻¹ e são pouco informativas²⁷. Os modelos SW-LDA e SPA-LDA exigem menor tempo computacional, mas apresentaram resultados inferiores.

Tabela 5. Figuras de mérito dos modelos LDA

	Ac ¹		CD0	CD7	CD30	ED0	ED7	ED30
GA-LDA Treinamento	100	Sen ²	100	100	100	100	100	100
		Esp ³	100	100	100	100	100	100
GA-LDA Teste	100	Sen ²	100	100	100	100	100	100
		Esp ³	100	100	100	100	100	100
AOC-LDA Treinamento	100	Sen ²	100	100	100	100	100	100
		Esp ³	100	100	100	100	100	100
AOC-LDA Teste	100	Sen ²	100	100	100	100	100	100
		Esp ³	100	100	100	100	100	100
SW-LDA Treinamento	97,50	Sen ²	100	100	100	83,33	100	100
		Esp ³	100	100	100	100	96,97	100
SW-LDA Teste	100	Sen ²	100	100	100	100	100	100
		Esp ³	100	100	100	100	100	100
SPA-LDA Treinamento	97,50	Sen ²	100	83,33	100	100	100	100
		Esp ³	100	100	96,77	100	100	100
SPA-LDA Teste	95,83	Sen ²	100	100	100	100	75	100
		Esp ³	100	100	100	95	100	100

¹Acurácia (%); ²Sensibilidade (%); ³Especificidade (%).

O modelo SW-LDA teve uma pequena confusão para modelar as classes ED0 e ED7 e após atingiu resultados de 100% nas métricas de teste. Já o modelo SPA-LDA apresentou alguma confusão ao modelar as classes CD7 e CD30 e ao testar as classes ED0 e ED7. Esses resultados vão ao encontro daqueles obtidos por PCA, em que as amostras do tempo 7 armazenadas ao

abrigo da luz apresentaram maior similaridade com o tempo 0, enquanto as amostras do tempo 7 expostas à luz apresentaram maior similaridade com o tempo 30.

Apesar do inferior desempenho em relação às figuras de mérito, o modelo SPA-LDA selecionou variáveis quimicamente mais importantes que os demais, como pode ser visto na Figura 37, e por isso foi escolhido para discussão. As variáveis selecionadas pelos demais algoritmos estão representadas nas Figuras 43, 44 e 45 do anexo.

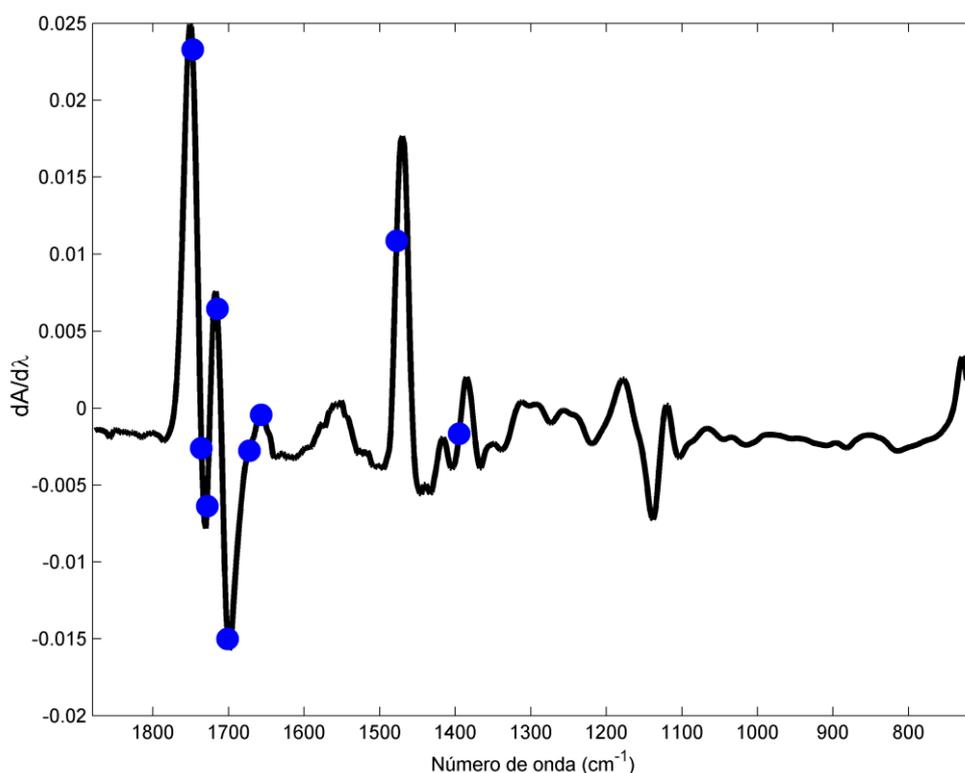


Figura 37. Variáveis selecionadas pelo modelo SPA-LDA.

Para melhor visualização das bandas que contém as variáveis selecionadas foi gerada a Figura 38, que apresenta o espectro de uma amostra sem pré-processamentos.

Como já discutido na PCA (em 5.2.1), as duas primeiras bandas correspondem ao estiramento e deformação axial da ligação carbonila de éster, proveniente de secreções sebáceas^{27,76,77}. Além dessas, as bandas em 1465 cm^{-1} e 1377 cm^{-1} correspondem CH_3 e CH_2 de cadeias alifáticas, também relacionadas às secreções sebáceas^{27,79}. Andersson et al. (2017)⁸⁰ indicam que

as bandas em 1740 cm^{-1} e 1377 cm^{-1} correspondem ao esqualeno, um composto que, segundo Dorakumbura *et al.* (2020)⁸¹, pode ser usado como biomarcador de digitais. A banda em 1653 cm^{-1} é característica de alongamento da ligação C=O de amida secundária e foi a única selecionada pelo modelo que corresponde as secreções écrinas de glândulas sudoríparas²⁸.

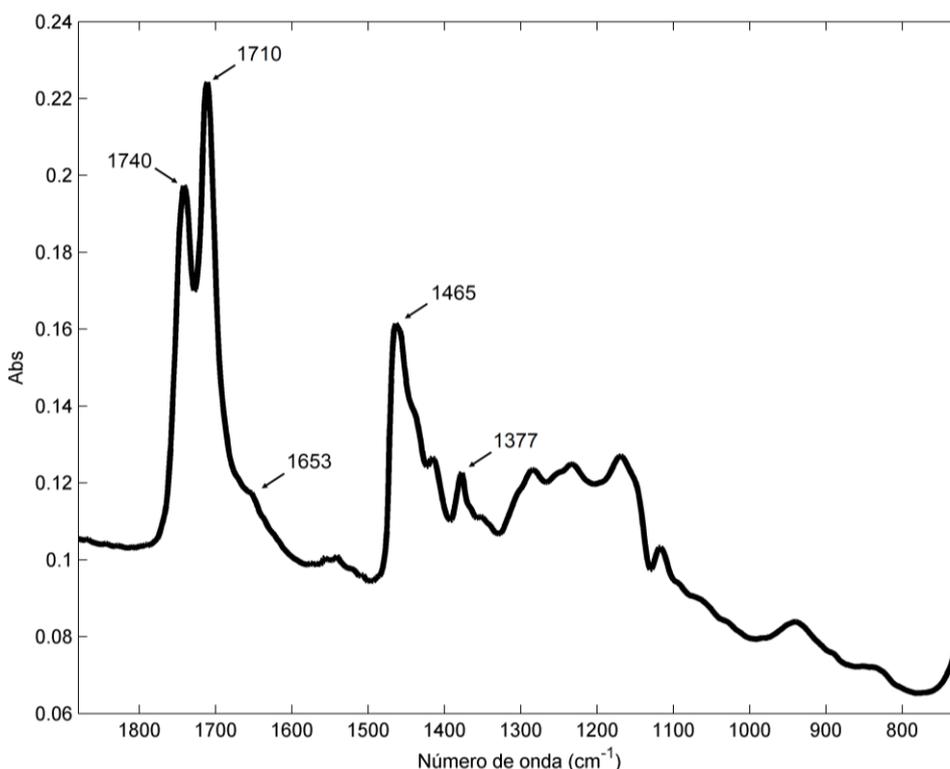


Figura 38. Espectro de uma amostra do tempo 0 sem tratamento, recortado na região de interesse e com indicativo das bandas com informações químicas mais relevantes para o estudo.

Na Figura 39 é apresentado o gráfico de escores das 2 primeiras funções discriminantes. A FD1 classificou para o extremo do seu lado positivo a classe CD0 e, para o lado negativo, as duas classes de 30 dias. A FD2 demonstra separar essas duas classes de 30 dias, agrupando ao lado positivo a classe ED30 e, para o lado negativo, a classe CD30, algo que não foi possível observar pela PCA.

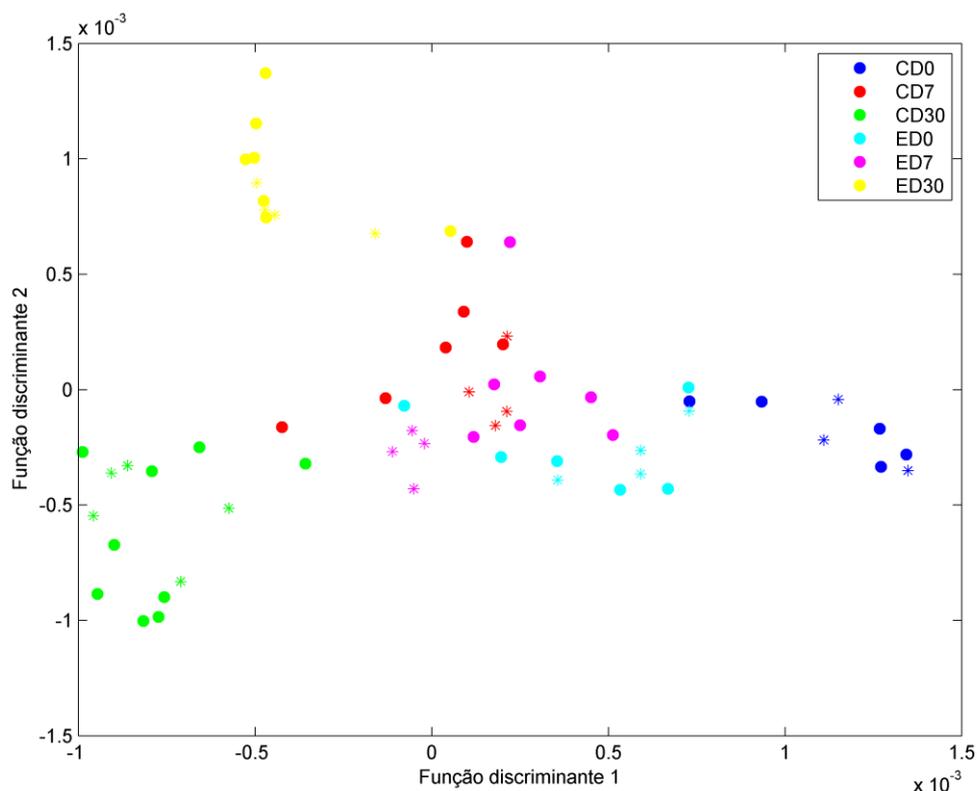


Figura 39. Gráfico dos escores FD1XFD2 da SPA-LDA. ● corresponde as amostras de treinamento e * as amostras de teste.

A função discriminante 3 é responsável pela separação das classes ED0, para o lado positivo e, ED7, para o lado negativo, o que pode ser visto pelo gráfico de escores FD3XFD4 apresentado na Figura 40. A FD4 indica separar as classes de tempo intermediário, segregando para seu lado positivo a classe CD7 e, para o negativo, a classe ED7.

As duas primeiras bandas, com máximos em 1740 cm^{-1} e 1710 cm^{-1} , carregam as 6 primeiras variáveis selecionadas e, como pode ser observado pelos gráficos de pesos, Figura 41, teve significativas contribuições nas funções discriminantes 2 e 4, que são funções que discriminam a condição de armazenamento para os tempos 30 e 7 dias, respectivamente. Nos escores da FD2, temos o escuro para o lado positivo e o claro para o negativo, enquanto na FD4 vê-se comportamento contrário, o que justifica o sentido oposto nos pesos dessa banda para essas duas funções.

Ao analisar apenas as classes do tempo 7, além da FD4, é possível observar nos escores das funções discriminantes 2 e 3 uma tendência em

separar essas classes. Nessas três funções descritas, a classe CD7 tende para o lado positivo e, a classe ED7, para o lado negativo. A banda em 1653 cm^{-1} , nos gráficos de pesos dessas 3 funções, está tendo contribuição para o lado negativo e pode ser associada à classe ED7. Por ser essa uma banda relacionada às secreções écrinas e essas tenderem a degradar rapidamente quando expostas à luz²⁷, esse resultado indica que apenas as amostras armazenadas ao abrigo da luz preservaram essas secreções nesse período inicial. Além disso, essa banda não aparece nos pesos da FD1, pois essa função não separa essas duas classes.

A banda em 1377 cm^{-1} aparenta segregar a classe CD30 em relação à classe ED7. Por ser relacionada à um possível biomarcador aparenta estar segregando a classe mais degradada de uma classe bastante conservada.

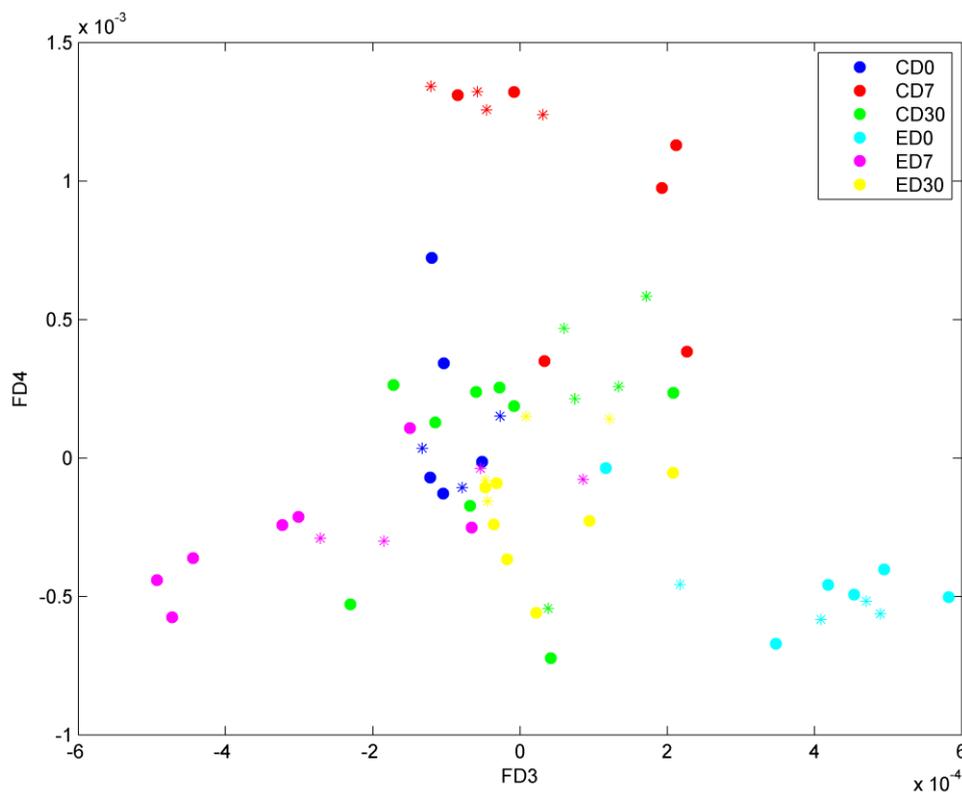


Figura 40. Gráfico dos escores FD3XFD4 da SPA-LDA. ● corresponde as amostras de treinamento e * as amostras de teste.

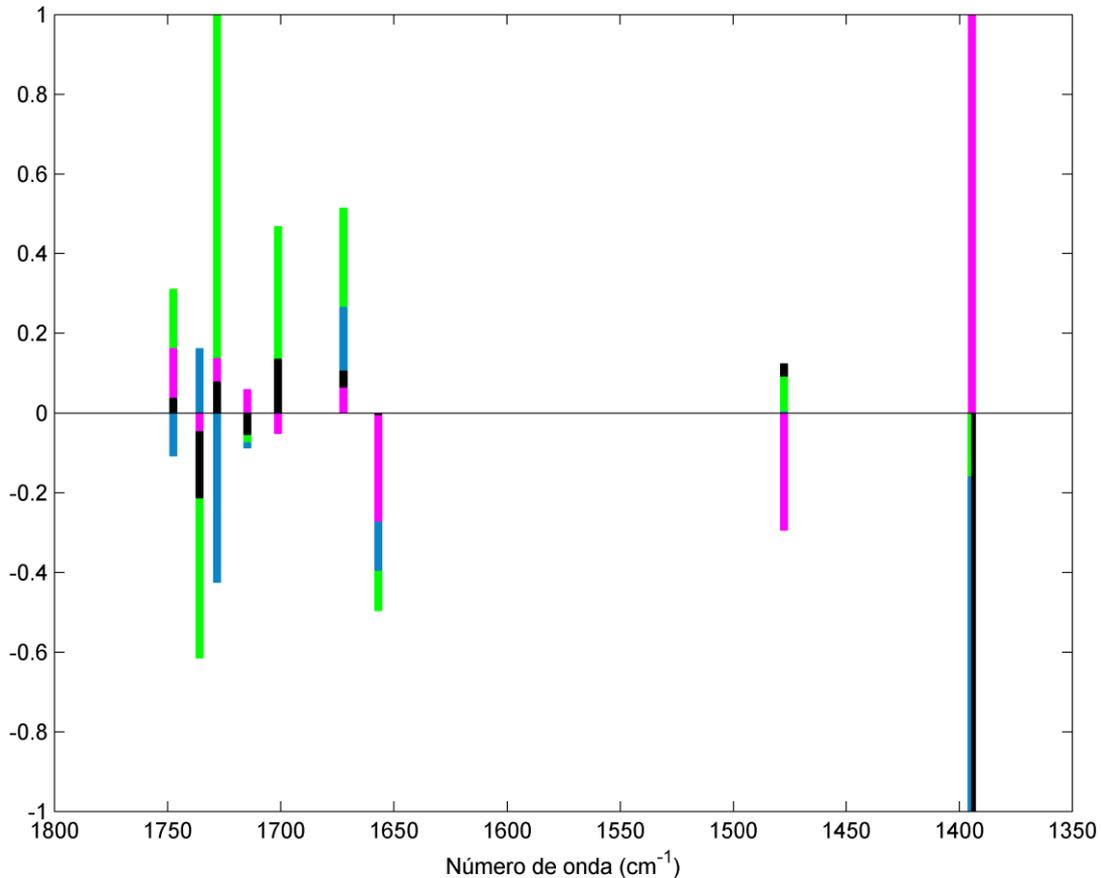


Figura 41. Gráfico de pesos das funções discriminantes 1 (preto), 2 (azul), 3 (rosa) e 4 (verde) da SPA-LDA.

Cabe ainda ressaltar que a quinta função discriminante não foi apresentada por não ter contribuição clara para o estudo.

5.2.3 Análise discriminante por mínimos quadrados parciais

Visando uma comparação dos resultados dos modelos de análise discriminante, foram desenvolvidos modelos PLS-DA. No treinamento do modelo obteve-se resultados excelentes, sendo perfeitamente possível modelar as 6 classes estudadas. Entretanto, essa perfeita separação não se manteve no teste do modelo, que classificou erroneamente 4 amostras, duas da classe CD30 e duas da classe ED7, o que resultou em desempenho inferior ao modelo LDA. Conforme já visto na PCA e corroborado na SPA-LDA, a classe ED7 tem perfil muito similar à classe ED0, o que pode justificar a dificuldade no

teste do modelo PLS-DA em classificar as amostras desses grupos. Na Tabela 6 é apresentado o desempenho do modelo pelas figuras de mérito.

Tabela 6. Figuras de mérito PLS-DA

		Ac ¹	CD0	CD7	CD30	ED0	ED7	ED30
Treinamento		Sen ²	100	100	100	100	100	100
	100	Esp ³	100	100	100	100	100	100
Teste		Sen ²	100	100	60	100	50	100
	83,33	Esp ³	100	100	100	80,95	100	100

¹Acurácia (%); ²Sensibilidade (%); ³Especificidade (%).

Na Figura 42 são apresentados os escores do PLS-DA para as 6 classes de impressões digitais e o limite de resposta de cada classe é representado na forma de uma linha tracejada vermelha. Podemos observar elevada dificuldade do modelo em relação a classe ED0 não só pelas amostras erroneamente classificadas, mas pela proximidade das classes com a linha de limite.

O modelo PLS-DA é amplamente conhecido e empregado em estudos forenses, como demonstrados nos recentes artigos de revisão de Mariotti *et al.* (2023)⁸² e Zhang *et al.* (2024)⁸³, normalmente utilizado como comparação na avaliação de outros modelos. Souza *et al.* (2022)³ empregaram a combinação de espectroscopia Raman e PLS-DA com o objetivo de discriminar sexo utilizando amostras de impressões digitais latentes e descreve como esse sendo um problema difícil, obtendo figuras de mérito que variaram de 63 a 91%.

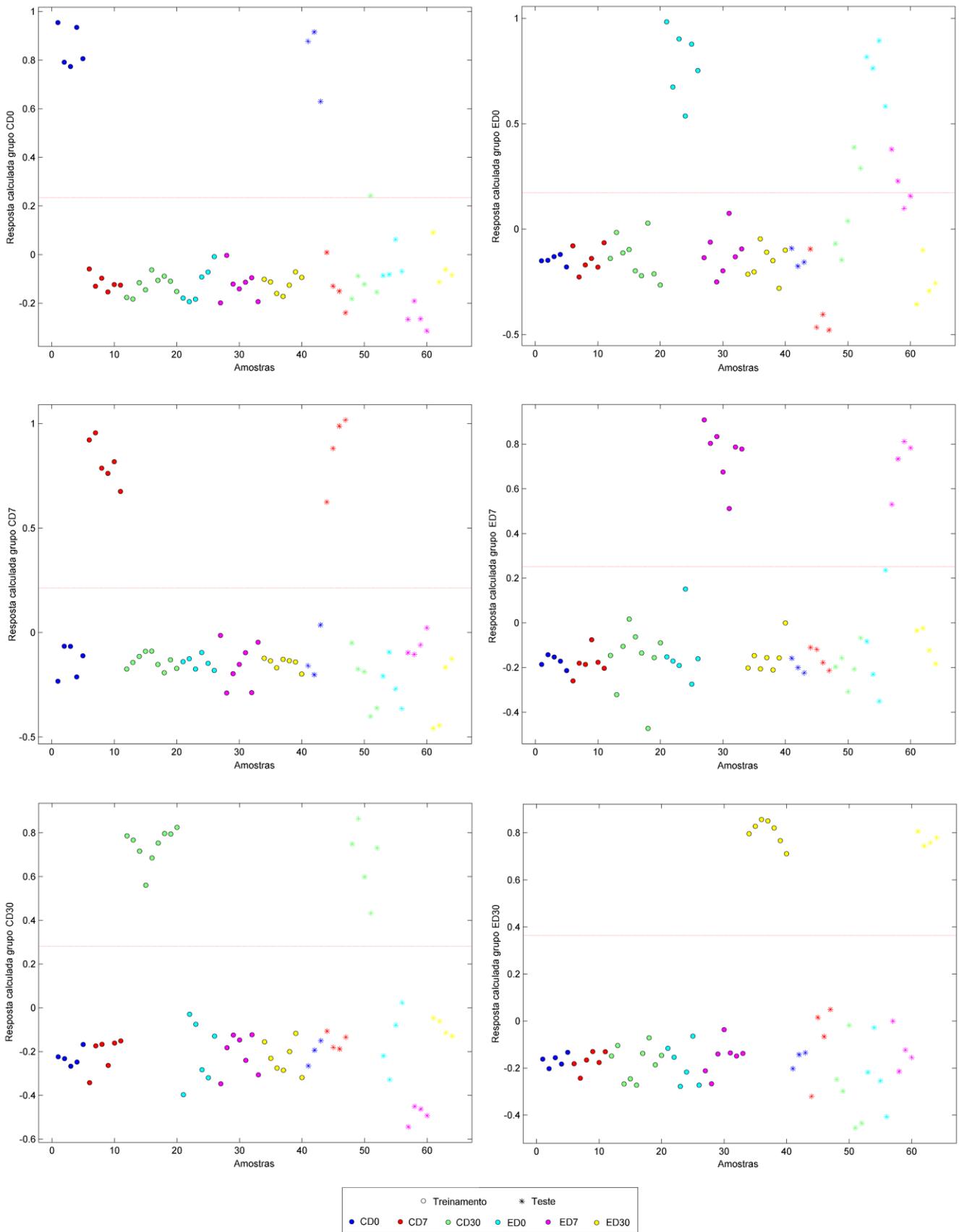


Figura 42. Respostas do modelo PLS-DA por classe. ● corresponde as amostras de treinamento e * as amostras de teste.

Conclusões

A aplicação de espectroscopia de infravermelho médio juntamente com técnicas quimiométricas demonstrou excelentes resultados para os estudos propostos, aliada à praticidade de obtenção dos dados e pelo emprego de uma técnica não destrutiva. Os pré-processamentos foram de suma importância para evidenciar as informações químicas mais importantes na discriminação das diferentes classes de amostras.

No estudo de documentos foi possível classificar as diferentes tintas de caneta empregadas, bem como os diferentes papéis utilizados. No caso das tintas de caneta, ficou evidente a dificuldade em discriminar quimicamente traços de tinta de canetas de mesmo fabricante, devido a base da tinta ser a mesma. Foi demonstrado a elevada contribuição que o papel forneceu aos dados, havendo melhores resultados de discriminação entre os papéis em relação às tintas, mesmo considerando que os espectros foram obtidos sobre os traços de caneta.

Como esperado, o estudo de impressões digitais latentes demonstrou maior dificuldade em obter uma clara separação de classes, principalmente no modelo não supervisionado de PCA, por se tratar de amostras mais complexas e similares. Ainda foi possível identificar o impacto do tipo de armazenamento na conservação das amostras.

Os diferentes algoritmos de seleção de variáveis empregados demonstraram excelentes resultados nos modelos de LDA. Entende-se que os modelos determinísticos podem ser considerados melhores devido à rapidez no emprego deles em relação aos modelos probabilísticos, que exigem maior tempo e melhores condições computacionais. Ambos os modelos estudados, LDA e PLS-DA, demonstraram ótimos desempenhos, sendo considerado melhor o modelo LDA por ser mais simples de aplicação e interpretação, principalmente pela seleção das variáveis mais importantes.

Espera-se que este trabalho contribua com o contínuo aprimoramento das ciências forenses, mostrando possibilidades de caminhos investigativos e a importância do estudo e aplicação de técnicas simples e rápidas.

6. Referências bibliográficas

1. Barros, F.; Kuhnen, B.; Serra, M.C.; Fernandes, C.M.S. Ciências forenses: princípios éticos e vieses. *Rev. Bioét.* **2021**, 29, 1, 55-65. <https://doi.org/10.1590/1983-80422021291446>
2. Gorziza, R.P.; Carvalho, C.M.B.; Ortiz, R.S.; Limberger, R.P. O estudo de tintas de canetas esferográficas: uma revisão para a abordagem pericial em documentoscopia forense. *Revista Brasileira de Ciências Policiais*, **2022**, 13, 9, 207-234. <https://doi.org/10.31412/rbcp.v13i9.867>
3. Souza, M.A.; Santos, A.S.; Silva, S.W.; Braga, J.W.B.; Sousa, M.H. Raman spectroscopy of fingerprints and chemometric analysis for forensic sex determination in humans. *Forensic Chem.* **2022**, 27, 100395. <https://doi.org/10.1016/j.forc.2021.100395>
4. Mariotti, K.C. Exames de impressões digitais: métodos espectrométricos e espectroscópicos aplicados a análise auímica de impressões digitais. *Braz. J. Forensic Sci.*, **2020**, 9, 210-228. [http://dx.doi.org/10.17063/bjfs9\(2\)y2020210](http://dx.doi.org/10.17063/bjfs9(2)y2020210)
5. Deviterne-Lapeyre, M.; Ibrahim, S. Interpol questioned documents review 2019–2022. *Forensic Sci. Int. Synergy*, **2023**, 6, 100300. <https://doi.org/10.1016/j.fsisyn.2022.100300>
6. Silva, C.S. Espectroscopia no infravermelho para aplicações forenses: documentoscopia e identificação de sêmen em tecidos. Tese (Doutorado em Química). **2017**, Universidade Federal de Pernambuco, Recife. <https://repositorio.ufpe.br/handle/123456789/29654>
7. Oliveira, M. Ciências forenses: o uso do método e do conhecimento científicos em investigações judiciais. Agência de Comunicação, UFSC. 2022. <<https://noticias.ufsc.br/2022/09/ciencias-forenses-o-uso-do-metodo-e-do-conhecimento-cientificos-em-investigacoes-judiciais/>> (acesso em 10/12/2023)
8. Rodrigues, C.H.P.; Amaral, M.E.A.; Mariotto, L.S.; Castro, J.S.; Mascarelli, M.E.; Velho, J.A.; Bruni, A.T. Forensic science or forensic sciences? a conceptual analysis. *Res., Soc. Dev.* **2022**, 11, 12, e177111234215. <http://dx.doi.org/10.33448/rsd-v11i12.34215>

17. Sharma, V.; Kumar, R. Fourier transform infrared spectroscopy and high performance thin layer chromatography for characterization and multivariate discrimination of blue ballpoint pen ink for forensic applications. *Vibr. Spectrosc.* **2017**, *92*, 96-104. <https://doi.org/10.1016/j.vibspec.2017.05.006>
18. Lian, Z.; Yang, R.; Zhao, L.; Shi, G.; Liang, L.; Qin, D.; Zou, J.; Yin, B. Analysis of volatile components in inkjet printouts by GC–MS: A classification method. *Forensic Sci. Int.* **2021**, *318*, 110562. <https://doi.org/10.1016/j.forsciint.2020.110562>
19. Sharif, M.; Jalees, M.I.; Tirmazi, S.A.A.Z.; Athar, M.M.; Durrani, A.I.; Batool, M. Discrimination of pakistani fountain pen inks by gas chromatography-mass spectrometry (GC-MS). *Int. J. Anal. Chem.*, **2022**, *2022*, 7186625, 10. <https://doi.org/10.1155/2022/7186625>
20. Copper, C.; Erbach, T.; McDonough, M.; Drury, N.; Moini, M. Analysis of ballpoint pen inks directly from paper using capillary electrophoresis. *Forensic Chem.* **2019**, *13*, 100145. <https://doi.org/10.1016/j.forc.2019.100145>
21. Djozan, D.; Baheri, T.; Karimian, G.; Shahidi, M. Forensic discrimination of blue ballpoint pen inks based on thin layer chromatography and image analysis. *Forensic Sci. Int.* **2008**, *179*, 2, 199-205. <https://doi.org/10.1016/j.forsciint.2008.05.013>
22. Kher, A.; Mulholland, M.; Green, E.; Reedy, B. Forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy with principal components analysis and linear discriminant analysis. *Vibr. Spectrosc.* **2006**, *40*, 2, 270-277.
23. Gautam, R.; Chauhan, R.; Kumar, R.; Sharma, V. PLS-DA and infrared spectroscopy based rapid and non-destructive discrimination of black ball and gel pen inks for forensic application. *Forensic Sci. Int. Reports.* **2021**, *3*, 100162. <https://doi.org/10.1016/j.fsir.2020.100162>
24. Sharma, S.; Garg, D.; Chophi, R.; Singh, R. On the spectroscopic investigation of stamp inks using ATR-FTIR and chemometrics: Application in forensic document examination. *Forensic Chem.* **2021**, *26*, 100377. <https://doi.org/10.1016/j.forc.2021.100377>
25. Hou, H.; Wu, Q.; Li, Z.; Wang, D.; Debrah, A.A.; Zou, J.; Du, Z. A non-invasive and rapid method for discrimination of inkjet printouts by ion mobility

spectrometry combined with chemometrics. *Mirochem. J.* **2023**, 194, 109250. <https://doi.org/10.1016/j.microc.2023.109250>

26. Tian, R.; Yu, Q.; Mei, L.; Zhu, F.; Qin, Q.; Ma, R.; Wang, Y.; Li, C.; Zhu, M. Single-molecule multi-target fluorescent imaging for latent fingerprints based on aggregation-induced emission. *Sens Actuators B Chem*, **2023**, 396, 134634. <https://doi.org/10.1016/j.snb.2023.134634>

27. Girod, A.; Xiao, L.; Reedy, B.; Roux, C.; Weyermann, C. Fingerprint initial composition and aging using Fourier transform infrared microscopy (μ -FTIR). *Forensic Sci. Int.* **2015**, 254, 185-196. <https://doi.org/10.1016/j.forsciint.2015.07.022>

28. Girod, A.; Spyratou, A.; Holmes, D.; Weyermann, C. Aging of target lipid parameters in fingerprint residue using GC/MS: Effects of influence factors and perspectives for dating purposes. *Sci. Justice* **2016**, 56, 3, 165-180.

29. Antoine KM, Mortazavi S, Miller AD, Miller LM, Ph D. Chemical differences are observed in children's versus adults' latent fingerprints as a function of time. *J. Forensic Sci.*, **2010**, 55, 513-518 <https://doi.org/10.1111/j.1556-4029.2009.01262.x>

30. Ruschel, C.F.C. Aplicação de ferramentas quimiométricas e técnicas espectroscópicas na análise de combustíveis. Tese (Doutorado em Química) **2017** Instituto de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre. <http://hdl.handle.net/10183/163889>

31. Holler, F.J.; Skoog, D.A.; Crouch, S.R. Princípios de Análise Instrumental. 6ª ed., Porto Alegre, Bookman, 2009, 1056 p.

32. John, D.K. Potencialidades da espectrometria no infravermelho e quimiometria para caracterização de perfis de medicamentos apreendidos. Dissertação (Mestrado em Química) **2020** Instituto de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre. <http://hdl.handle.net/10183/218037>

33. Vogel, A.I. Análise química quantitativa. 6ª ed., Rio de Janeiro. LTC, 2022, 488p.

34. Lilo, T.; Morais, C.L.M.; Shenton, C.; Ray, A.; Gurusinghe, N. Revising Fourier-transform infrared (FT-IR) and Raman spectroscopy towards brain cancer detection. *Photodiagnosis Photodyn Ther*, **2022**, 38, 102785. <https://doi.org/10.1016/j.pdpdt.2022.102785>

35. Schneider, M.P. Origem geográfica e perfil varietal de vinhos tintos produzidos no Rio Grande do Sul via técnicas espectroscópicas assistidas por ferramentas quimiométricas. Dissertação (Mestrado em Química) **2021** Instituto de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre. <http://hdl.handle.net/10183/220382>

36. Veras, G.; Alves, V.D.; Ramos, H.A.; Gomes, M; Figueiredo, L.; Matias, E.V.S. Perfil ciutométrico da quimiometria no brasil. *Quim. Nova* **2022**, 45, No. 10, 1315-1321. <http://dx.doi.org/10.21577/0100-4042.20170930>

37. Ataide, V.N.; Pradela Filho, L.A.; Guinati, B.G.S.; Moreira, N.S.; Gonçalves, J.; Ribeiro, C.M.G.; Grasseschi, D.; Coltro, W.K.T.; Salles, M.O.; Paixão, T.R.L.C. Combining chemometrics and paper-based analytical devices for sensing: An overview *TrAC Trends Anal. Chem.* **2023**, 164, 117091. <https://doi.org/10.1016/j.trac.2023.117091>

38. Silva, E.G.C. Avaliação do envelhecimento do papel para determinar falsificações de documentos utilizando espectroscopia vibracional e técnicas quimiométricas. Dissertação (Mestrado em Química) **2020**. Departamento de Química Fundamental, Universidade Federal de Pernambuco, Recife. <https://repositorio.ufpe.br/handle/123456789/41711>

39. Junges, C.H. Novas metodologias analíticas na autenticação de sucos de uva integral orgânicos empregando dados de técnicas de espectrometria molecular e ferramentas quimiométricas. Tese (Doutorado em Química) **2023** Instituto de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre. <http://hdl.handle.net/10183/259616>

40. Ferreira, M.M.C. Quimiometria: conceitos, métodos e aplicações. Campinas, Unicamp, 2015, 1ª ed., 493 p.

41. Souza, A.M.; Poppi, R.J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. *Quim. Nova* **2012**, 35, No. 1, 223-229. <https://doi.org/10.1590/S0100-40422012000100039>

42. Rech, A.M. Caracterização de bebidas à base de soja empregando espectroscopia no infravermelho médio com transformada de fourier por reflexão total atenuada e quimiometria. Dissertação (Mestrado em Química)

2018 Instituto de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre. <http://hdl.handle.net/10183/180663>

43. Helfer, G.A. Chemostat: desenvolvimento de software para análise exploratória de dados multivariados. Dissertação (Mestrado em Sistemas e Processos Industriais) **2014** Universidade de Santa Cruz do Sul – UNISC. <http://hdl.handle.net/11624/685>

44. Marcelo, M.C.A. Identificação da procedência geográfica da erva mate (*Ilex paraguariensis*) mediante análise elementar, reflexão no infravermelho próximo e quimiometria. Dissertação (Mestrado em Química) **2013** Instituto de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre. <http://hdl.handle.net/10183/98121>

45. Lattin, J.; Carroll, J.D.; Green, P.E., *Análise de Dados Multivariados*. Cengage Learning: São Paulo, 2011.

46. Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, 11 (1), 137–148. <https://doi.org/10.1080/00401706.1969.10490666>

47. Daszykowski, M.; Walczak, B.; Massart, D.L. Representative subset selection. *Anal. Chim. Acta* **2002**, 468,1, 91-103. [https://doi.org/10.1016/S0003-2670\(02\)00651-7](https://doi.org/10.1016/S0003-2670(02)00651-7)

48. Marini, F. Classification methods in chemometrics. *Curr. Anal. Chem.* **2010**, 6. 72-79.

49. Berrueta, L.A.; Alonso-Salces, R.M.; Héberger, K. Supervised pattern recognition in food analysis. *J. Chromatogr. A* **2007**, 1158, 1, 196-214. <https://doi.org/10.1016/j.chroma.2007.05.024>

50. Smith, R. Chemometrics. In *Forensic Chemistry*; John Wiley & Sons, Ltd: Chichester, UK, 2015; pp 469–503. <https://doi.org/10.1002/9781118897768.ch12>

51. Garcia, M.J.L. Introduction. In: *Characterization and Authentication of Olive and Other Vegetable Oils*. Springer Theses. Springer, Berlin, Heidelberg, 2012; pp 1-43. https://doi.org/10.1007/978-3-642-31418-6_1

52. Vinciguerra, L.L. Novas metodologias para determinação de origem de vinhos gaúchos empregando espectrometria por fluorescência associada a quimiometria. Tese (Doutorado em Ciências Farmacêuticas) **2021** Faculdade

de Farmácia, Universidade Federal do Rio Grande do Sul, Porto Alegre.
<http://hdl.handle.net/10183/226154>

53. Gomes, A.A. Algoritmo das projeções sucessivas aplicado à seleção de variáveis em regressão PLS. Dissertação (Mestrado em Química) **2012** Universidade Federal da Paraíba, João Pessoa.
<https://repositorio.ufpb.br/jspui/handle/tede/7056>

54. Dorigo, M.; Maniezzo, V.; Coloni, A. The ant system: optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man Cybern. Part B, Cybern.*, **1996**, 26, 29-41.<http://dx.doi.org/10.1109/3477.484436>

55. Dorigo, M.; Gambardella, L.M. Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE Trans. Evol. Comput.*, **1997**, 1, 53-66. <http://dx.doi.org/10.1109/4235.585892>

56. Sanca, H. S. Reconfiguração ótima de sistemas de distribuição de energia elétrica aplicando o algoritmo MAX-MIN Ant System. Dissertação (Mestrado em Engenharia Elétrica) **2013** Universidade Federal da Bahia, Escola Politécnica, Salvador.
https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=89178

57. Ranzan, C. Desenvolvimento de modelos quimiométricos utilizando o algoritmo de otimização de colônia de formigas. Tese (Doutorado em Engenharia Química) **2014** Universidade Federal do Rio Grande do Sul, Porto Alegre.<https://lume.ufrgs.br/bitstream/handle/10183/101212/000930612.pdf?sequence=1&isAllowed=y>

58. Alves, M.F.; Lotufo, A.D.P.; Lopes, M.L.M. Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas. *Proc. Ser. Braz. Soc. Appl. Comput. Math*, **2013**, 1, 010144.
<https://doi.org/10.5540/03.2013.001.01.0144>

59. Araújo, M.C.U.; Saldanha, T.C.B.; Galvão, R.K.H.; Yoneyama, T.; Chame, H.C.; Visani, V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemom. Intell. Lab. Syst.* **2001**, 57, 2, 65-73. [https://doi.org/10.1016/S0169-7439\(01\)00119-8](https://doi.org/10.1016/S0169-7439(01)00119-8)

60. Vallese, F.D.; Paoloni, S.G.; Springer, V.; Fernandes, D.D.S.; Diniz, P.H.G.D.; Pistonesi, M.F. Exploiting the successive projections algorithm to

improve the quantification of chemical constituents and discrimination of botanical origin of Argentinean bee-pollen. *J. Food Compos. Anal.* **2024**, 126, 105925. <https://doi.org/10.1016/j.jfca.2023.105925>

61. Milanez, K.D.T.M. Seleção de variáveis robustas para transferência de modelos de classificação empregando o Algoritmo das Projeções Sucessivas. Tese (Doutorado em Química) **2017** Universidade Federal da Paraíba, João Pessoa. <https://repositorio.ufpb.br/jspui/handle/123456789/11879>

62. Pontes, M.J.C.; Galvão, R.K.H.; Araújo, M.C.U.; Moreira, P.N.T.; Neto, O.D.P.; José, G.E.; Saldanha, T.C.B. The successive projections algorithm for spectral variable selection in classification problems. *Chemom. Intell. Lab. Syst.* **2005**, 78, 1-2, 11-18. <https://doi.org/10.1016/j.chemolab.2004.12.001>

63. Grasel, F.S.; Ferrão, M.F. Rapid discrimination of natural polyphenols (vegetable tannins) from different plants by molecular spectroscopy and PLS-DA. *Anal. Methods*, **2018**, 10, 968-974. <https://doi.org/10.1039/C8AY00075A>

64. de Andrade, B.M.; Margalho, L.P.; Batista, D.B.; Lucena, I.O.; Kamimura, B.A.; Balthazar, C.F.; Brexó, R.P.; Pia, A.K.R.; Costa, R.A.S.; Cruz, A.G.; Granato, D.; Sant'Ana, A.S.; Luna, A.S.; de Gois, J.S. Chemometric classification of Brazilian artisanal cheeses from different regions according to major and trace elements by ICP-OES. *J. Food Compos. Anal.* **2022**, 109, 104519. <https://doi.org/10.1016/j.jfca.2022.104519>

65. Silva, C.S.; Borba, F.S.L.; Pimentel, M.F.; Pontes, M.J.C.; Honorato, R.S.; Pasquini, C. Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis. *Mirochem. J.* **2013**, 109, 122-127. <https://doi.org/10.1016/j.microc.2012.03.025>

66. MATLAB. Version 8.0.0.783 R2012b. The MathWorks, Inc., Natick, Massachusetts, EUA. August, 2012. Disponível em: <http://www.mathworks.com>

67. Ballabio, D. A MATLAB Toolbox for Principal Component Analysis and Unsupervised Exploration of Data Structure. *Chemom. Intell. Lab. Syst.* **2015**, 149, 1–9. <https://doi.org/10.1016/j.chemolab.2015.10.003>

68. Ballabio, D.; Consonni, V. Classification Tools in Chemistry. Part 1: Linear Models. PLS-DA. *Anal. Methods* 2013, 5 (16), 3790. <https://doi.org/10.1039/c3ay40582f>.

69. Homemade routine available under request to araujo.gomes@ufrgs.br

70. Manente, S.; Micheluz, A.; Ganzerla, R.; Ravagnan, G.; Gambaro, A. Chemical and biological characterization of paper: A case study using a proposed methodological approach. *Int. Biodeterior. Biodegrad.* **2012**, 74, 99-108. <https://doi.org/10.1016/j.ibiod.2012.03.008>
71. Farid, S.; Kasem, M.A.; Zedan, A.F.; Mohamed, G.G.; El-Hussein, A. Exploring ATR Fourier transform IR spectroscopy with chemometric analysis and laser scanning microscopy in the investigation of forensic documents fraud. *Opt. Laser Technol.* **2021**, 135, 106704. <https://doi.org/10.1016/j.optlastec.2020.106704>
72. Ganzerla, R.; Gambaro, A.; Cappelletto, E.; Fantin, M.; Montalbani, S.; Orlandi, M. Characterization of selected paper documents from the archives of Palazzo Ducale (Venice), Italy using various analytical techniques. *Mirochem. J.* **2009**, 91, 1, 70-77. <https://doi.org/10.1016/j.microc.2008.08.003>
73. Silva, E.G.C.; Silva, C.S.; Pimentel, M.F. Infrared spectroscopy for document dating - Advantages, challenges and limitations. *Talanta Open.* **2022**, 5, 100102. <https://doi.org/10.1016/j.talo.2022.100102>
74. Wang, X.F.; Yu, J.; Zhang, A.L.; Zhou, D.-W.; Xie, M.-X. Nondestructive identification for red ink entries of seals by Raman and Fourier transform infrared spectrometry. *Spectrochim. Acta. Part. A, Mol. Biomol. Spectrosc.* **2012**, 97, 986-994. <https://doi.org/10.1016/j.saa.2012.07.127>
75. Helfer G.A.; Bock, F.; Marder, L.; Furtado, J. C.; Costa, A. B.; Ferrão, M. F.; Chemostat, um software gratuito para análise exploratória de dados multivariados. *Quim. Nova.* **2015**. 38, 4, 575. <http://dx.doi.org/10.5935/0100-4042.20150063>
76. Crane N.J.; Bartick E.G.; Perlman R.S.; Huffman S. Infrared spectroscopic imaging for noninvasive detection of latent fingerprints. *J Forensic Sci.* **2007** 52, 1, 48–53. doi: 10.1111/j.1556-4029.2006.00330.x
77. Souza, M.A. A química instrumental e a identificação humana: aplicações e potenciais. *Revista Brasileira de Ciências Policiais*, Brasília, **2022** v. 13, n. 8, 71-93. <https://doi.org.br/10.31412/rbcp.v13i8>
78. Chen, H.; Shi, M.; Ma, R.; Zhang, M. Advances in fingermark age determination techniques. *Anlst.* **2021**, 146, 1, 33-47. doi: 10.1039/D0AN01423K

79. González, M.; Mariotti, K. C.; Gomes, A. A.; Ferrão, M. F.; Limberger, R. P. Fingerprint Analysis by Fourier Transform Infrared Microscopy Using Chemometric Tools. *Braz. J. Anal. Chem.*, **2021**, 8, 32, 141–154. doi: 10.30744/brjac.2179-3425.AR-25-2021

80. Andersson, P.O.; Lejon, C.; Mikaelsson, T.; Landström, L. Towards Fingerprint Dating: A Raman Spectroscopy Proof-of-Concept Study. *Chem. Open* **2017**, 6, 706-709. <https://doi.org/10.1002/open.201700129>

81. Dorakumbura, B.N.; Buseti, F.; Lewis, S.W. Analysis of squalene and its transformation by-products in latent fingerprints by ultrahigh-performance liquid chromatography-high resolution accurate mass Orbitrap™ mass spectrometry. *Forensic Chem.* **2020**, 17, 100193

82. Mariotti, K.C.; Ortiz, R.S.; Ferrão, M.F. Hyperspectral imaging in forensic science: An overview of major application areas. *Sci. Justice*, **2023**, 63, 3, 387-395. <https://doi.org/10.1016/j.scijus.2023.04.003>

83. Zhang, Z.; Wang, J.; Lu, X.; Wang, H. Recent progress in the application of chemical attribution in forensic profiles. *Forensic Chem.*, **2024**, 37-100556. <https://doi.org/10.1016/j.forc.2024.100556>

7. Anexos

7.1 Variáveis selecionadas nos modelos GA-LDA, AOC-LDA e SW-LDA

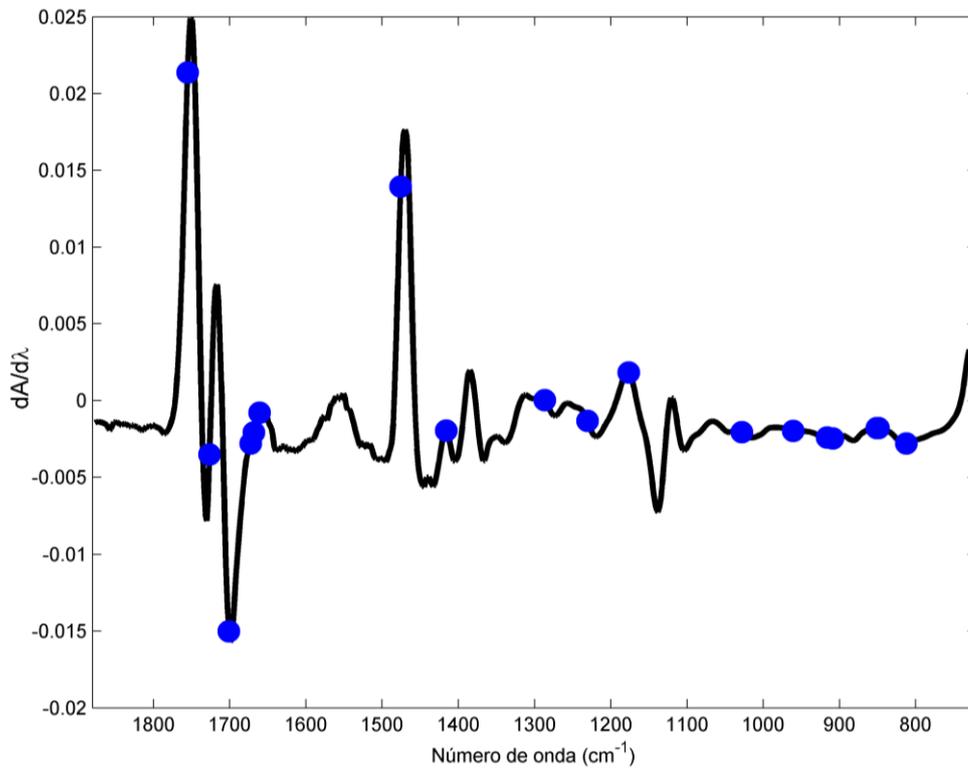


Figura 43. Variáveis selecionadas modelo GA-LDA.

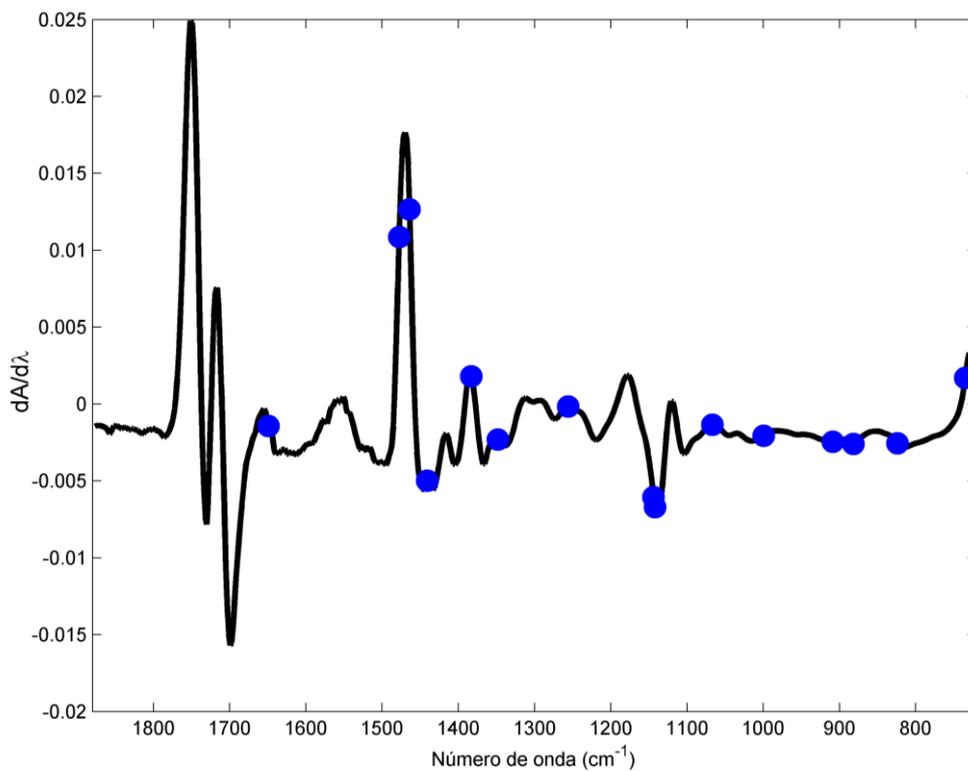


Figura 44. Variáveis selecionadas modelo AOC-LDA.

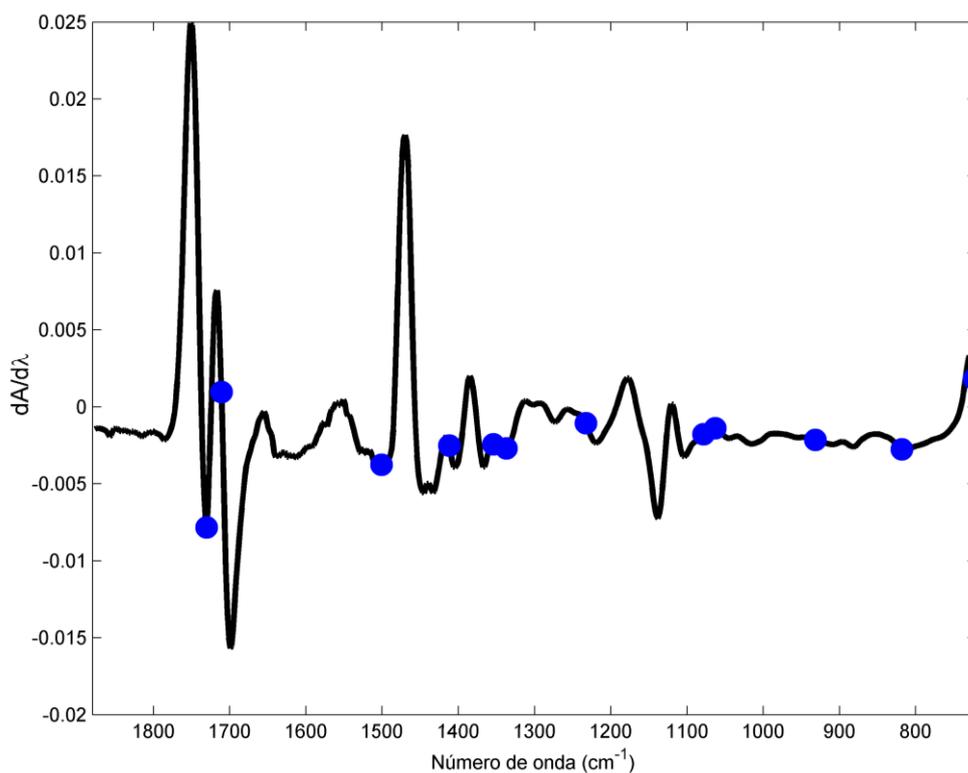


Figura 45. Variáveis selecionadas modelo SW-LDA.