

**CLASSIFICAÇÃO MULTIVARIADA DE BIODIESEL B100 E B5
USANDO IMAGENS DIGITAIS**

GEAN BEZERRA DA COSTA

UNIVERSIDADE ESTADUAL DA PARAÍBA

CAMPINA GRANDE-PB

FEVEREIRO DE 2015

**CLASSIFICAÇÃO MULTIVARIADA DE BIODIESEL B100 E B5
USANDO IMAGENS DIGITAIS**

GEAN BEZERRA DA COSTA

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Agrárias da Universidade Estadual da Paraíba/Embrapa Algodão, como parte das exigências para obtenção do título de Mestre em Ciências Agrárias / Área de Concentração: Energias Renováveis e Biocombustíveis.

Orientador: Prof. Dr. José Germano Vêras Neto

Coorientador: Prof. Dr. Paulo Henrique Gonçalves Dias Diniz

**CAMPINA GRANDE- PB
FEVEREIRO DE 2015**

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

C837c Costa, Gean Bezerra da.
Classificação multivariada de Biodiesel B100 e B5 usando
imagens digitais [manuscrito] / Gean Bezerra da Costa. - 2015.
80 p. : il. color.

Digitado.

Dissertação (Programa de Pós-Graduação em Ciências
Agrárias) - Universidade Estadual da Paraíba, Pró-Reitoria de Pós-
Graduação e Pesquisa, 2015.

"Orientação: Prof. Dr. José Germano Vêras Neto, Pró-
Reitoria de Pós-Graduação e Pesquisa".

"Co-Orientação: Prof. Dr. Paulo Henrique Gonçalves Dias
Diniz, Departamento de Química".

1. Biocombustíveis. 2. Webcam. 3. Quimiometria. I. Título.
21. ed. CDD 662.8

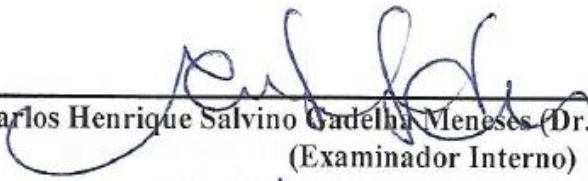
CLASSIFICAÇÃO MULTIVARIADA DE BIODIESEL B100 E B5 USANDO IMAGENS DIGITAIS

GEAN BEZERRA DA COSTA


Dissertação apresentada ao Programa de Pós-Graduação em Ciências Agrárias da Universidade Estadual da Paraíba/Embrapa Algodão, como parte das exigências para obtenção do título de Mestre em Ciências Agrárias / Área de Concentração: Energias Renováveis e Biocombustíveis.

Aprovada em 13 de Fevereiro de 2015

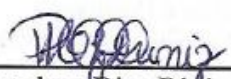
Banca Examinadora




Prof. Carlos Henrique Salvino Gadelha Menezes (Dr., Química Analítica) - UEPB
(Examinador Interno)



Wellington da Silva Lyra (Dr., Química Analítica) – UFPB
(Examinador Externo)



Prof. Paulo Henrique Gonçalves Dias Diniz (Dr., Química Analítica) – UEPB
(Coorientador)



Prof. José Germano Vêras Neto (Dr., Química Analítica) - UEPB
(Orientador)

Àqueles que me ensinaram o que é o amor, a vida e a felicidade; Àqueles que me deram segurança e conforto; Àqueles que me deram estímulo, carinho e orientação; Àqueles que passaram a noite em claro por minha causa e tiveram paciência comigo; Àqueles que riram e choraram ao meu lado; Àqueles que são meus amigos, meu ombro, meu colo; Àqueles que vibraram comigo e também estenderam a mão para mim; Àqueles para quem sou o mais bonito, mais inteligente e campeão; Aos meus fãs incondicionais; Àqueles que dariam a própria vida por mim; Àqueles a quem devo a própria vida; Àqueles a quem amo com toda altura e profundidade do meu coração; Aos meus eternos heróis; Aos meus pais”.

Dedico

*" Diga-me eu esquecerei,
ensina-me e eu poderei lembrar,
envolva-me e eu aprenderei "*

Benjamin Franklin

AGRADECIMENTO

- Agradecer é antes de tudo reconhecer que nada fazemos sozinhos. É reconhecer aqueles que nos apoiaram e depuseram confiança em nossa capacidade.
- Inicialmente quero agradecer a Deus, pois nos momentos difíceis foi Nele que busquei forças para continuar.
- Agradeço a minha família que sempre unida criou um ambiente agradável e propício ao desenvolvimento humano e social, que me fez o que sou hoje. Em especial aos meus pais e minha irmã Marilene. A minha esposa Isabel e ao meu filho Guilherme, que me deram muita inspiração.
- Ao Programa de Pós-Graduação em Ciências Agrárias UEPB/ EMBRAPA pela oportunidade de formação acadêmica e crescimento profissional.
- Meus sinceros agradecimentos ao meu orientador Professor Drº. José Germano Vêras Neto toda a paciência, amizade e ensinamentos ao longo destes anos.
- Ao Prof. Dr. Paulo Diniz, que esteve à frente da minha pesquisa com muita competência e por todos os conhecimentos que a mim foram transmitidos. Pela coorientação, paciência, e especialmente pela confiança em mim depositada ao longo deste trabalho.

- Aos professores do Programa de Pós-Graduação em Ciências Agrárias UEPB/ EMBRAPA, que através de seus ensinamentos ‘plantaram’ em mim boas sementes que estimulam minhas atividades profissionais.
- A equipe do Laboratório de Química Analítica e Quimiometria (LQAQ) pela infraestrutura e apoio que possibilitaram a realização desta pesquisa. Valber Elias (PC1), a Wellen Duarte (PC2) são as que contém maiores informações, Thomás mais conhecido com Tuba (PC3), as funções discriminantes Martina Maia (DF1) e (DF2) Jéssica Priscila, ao novos lobinhos Gizelly, Rossana e Juliana, meus agradecimentos a todos.
- Ao Algoritmo das Projeções Sucessivas (SPA) David Fernandes, pelo apoio e grande contribuição em todos os trabalhos que realizamos juntos desde a iniciação científica, mestrado (dissertação) e em um possível doutorado. Grato
- Aos colegas de mestrado Ingredy, Adriana, Ana Lígia, Ciro, Geisenilma, Isaias, Jean Pierre, Luanna, Suziane, Thiago e Wellison, pelos inesquecíveis dois anos de convivência participando do meu dia-a-dia. Cada um com uma parcela de contribuição. Por todas as alegrias, grupos de estudo antes das provas e trabalhos e, por todo carinho. Saibam que direta ou indiretamente sempre me ajudaram e, por isso, vocês também fazem parte dessa conquista.
- A CAPES pelo fornecimento da bolsa.

“Não há figura de mérito que consiga expressar a gratidão que tenho por todos vocês”.

SUMÁRIO

	Pg.
LISTA DE TABELAS	viii
LISTAS DE FIGURAS	ix
LISTA DE ABREVEATURAS	xi
RESUMO	xii
ABSTRACT	xiii
1. INTRODUÇÃO	14
2. REVISÃO DA LITERATURA	18
2.1. Biodiesel	18
2.2. Misturas biodiesel/diesel	18
2.3. Imagens digitais	19
2.3.1. Pixel	20
2.3.2. Histograma	20
2.4. Sistemas de cores	21
2.4.1. Sistemas de cores RGB	22
2.4.2. Sistemas de cores HSI	24
2.4.3. Escala de Cinza	26
2.5. Quimiometria	27
2.5.1. Técnicas de reconhecimento de padrões	27
2.5.1.1. Técnicas de reconhecimento de padrões não supervisionadas	27
2.5.1.1.1. Análise de Componentes Principais (PCA)	28
2.5.1.2. Técnicas de reconhecimento de padrões supervisionadas	29
2.5.1.2.1. Modelagem Independente e Flexível por Analogia de Classe (SIMCA)	30

2.5.1.2.2. Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA)	30
2.5.1.2.3. Análise Discriminante Linear (LDA)	31
2.5.2. Seleção de variáveis	32
2.5.2.1. Algoritmo das Projeções Sucessivas (SPA)	33
2.6. Aplicações de ferramentas quimiométricas e imagens digitais em problemas de classificação	34
3. OBJETIVOS	36
3.1. Objetivo geral	36
3.2. Objetivo específicos	36
4. MATERIAL E MÉTODOS	37
4.1. Localização do experimento	37
4.2. Amostras de biodiesel	37
4.3. Amostras de misturas biodiesel/diesel (B5)	40
4.4. Aparato	41
4.5. Aquisição de imagens	42
4.6. Histograma de cores e análise dos dados	43
5. RESULTADOS E DISCUSSÃO	45
5.1. Classificação de biodiesel B100 usando imagens digitais e quimiometria.	45
5.1.1. Análise exploratória - PCA	46
5.1.2. Classificação	48
5.1.2.1. Classificação SIMCA	50
5.1.2.2. Classificação PLS-DA	51
5.1.2.3. Classificação SPA-LDA	53
5.2. Classificação de misturas biodiesel/diesel (B5) por imagem digital e quimiometria	56
5.2.1. Análise exploratória - PCA	57
5.2.2. Classificação	58
5.2.2.1. Classificação SIMCA	60
5.2.2.2. Classificação PLS-DA	61
5.2.2.3. Classificação SPA-LDA	62
6. CONCLUSÃO	66
REFERÊNCIAS	68

LISTA DE TABELAS

	Pg.
Tabela 1 – Composição de ácidos graxos (%) de alguns óleos vegetais comuns usados como fonte de biodiesel.	48
Tabela 2 - Funções Discriminantes obtidas pelo SPA-LDA na classificação das 120 amostras de misturas biodiesel/diesel B5: (●) algodão (●) girassol (●) milho, (●) soja.	49
Tabela 3 - Matriz de confusão da classificação SIMCA para as amostras de biodiesel B100 usando o histograma de Escala de Cinza + RGB.	50
Tabela 4 - Matriz de confusão da classificação PLS-DA para as amostras de biodiesel B100 usando o histograma de Intensidade	52
Tabela 5 - Matriz de confusão da classificação SPA-LDA para as amostras de biodiesel B100 usando o histograma RGB.	55
Tabela 6 - Taxa de classificação correta obtida com os modelos SIMCA, PLS-DA e SPA-LDA para a classificação das amostras de misturas biodiesel/diesel B5.	59
Tabela 7 - Matriz de confusão da classificação SIMCA para as amostras de misturas B5 usando o histograma de Escala de Cinza + RGB + HSI.	60
Tabela 8 – Matriz de confusão da classificação PLS-DA para as amostras de misturas B5 usando o histograma de RGB.	61
Tabela 9 – Matriz de confusão da classificação SPA-LDA para as amostras de misturas B5 usando o histograma de I (Intensidade).	64

LISTA DE FIGURAS

	Pg.
Figura 1. Histogramas referentes à distribuição de frequência de todos os pixels nos níveis de Cinza, Vermelho, Verde e Azul da imagem apresentada.	21
Figura 2. Cores primárias aditivas e as cores resultantes de suas combinações R: Vermelho; G: Verde; B: Azul.	23
Figura 3. Cubo de cores RGB	24
Figura 4. Sistemas HSI de cores.	25
Figura 5. Escala dos níveis de Cinza	26
Figura 6. Decomposição da matriz X em Componentes Principais	28
Figura 7. Posicionamento dos parâmetros da PCA entre duas variáveis originais	29
Figura 8. Representação de um modelo preditivo ideal de PLS-DA para 4 classes distintas	31
Figura 9. Representação esquemática da Análise Discriminante Linear	32
Figura 10. Sistema de refluxo e processo de formação do metóxido de potássio	38
Figura 11. Adição do óleo ao balão de duas bocas acoplado ao sistema de refluxo, para reação de transesterificação	38
Figura 12. Separação da Glicerina após 24 horas de repouso da mistura	39
Figura 13. (a) Processo de purificação do biodiesel; (b) biodiesel purificado	39
Figura 14. Processo de tratamento térmico do biodiesel	40
Figura 15. Esquema ilustrativo para mistura do biodiesel ao diesel.	41
Figura 16. Sistema utilizado na aquisição das imagens das amostras de biodiesel B100 e B5. (a) notebook; (b) placa de Petri; (c) caixa; (d) Lâmpada fluorescente e (e) webcam	42
Figura 17. Representação de amostras na matriz de dados	43

Figura 18. Histogramas médios nos canais de Escala de Cinza, Vermelho, Verde, Azul, Matiz, Saturação e de Intensidade para cada classe de amostra biodiesel: algodão (linha azul), milho (linha vermelha), girassol (linha verde) e soja (linha amarela).	46
Figura 19. PCA gráficos dos escores PC1 x PC2, obtidos para as componentes de cores (a) RGB + HSI + Escala de Cinza (b) RGB e (c) Cinza + HSI, para as quatro classes das amostras de biodiesel. (●) algodão (●) girassol, (●) milho, (●) soja	47
Figura 20. Gráfico de escores de PC1 × PC2 de SIMCA em relação à classe de soja. Amostras de treinamento (círculo) e teste (triângulo) para os biodieseis B100 de algodão (azul), milho (vermelho), girassol (verde) e soja (amarelo).	51
Figura 21. Gráfico de escores de PLS-DA utilizando o canal Intensidade. Amostras de treinamento (círculo) e teste (triângulo) para os biodieseis B100 de algodão (azul), milho (vermelho), girassol (verde) e soja (amarelo).	53
Figura 22. (a) As 20 variáveis selecionadas pelo SPA para o sistema RGB e (b) distribuição de frequências das cores predominantes para a classificação de amostras de biodiesel B100.	54
Figura 23. Funções Discriminantes obtidas pelo SPA-LDA na classificação das 120 amostras de biodiesel B100: (●) algodão (●) girassol (●) milho, (●) soja.	56
Figure 24. Histogramas médios nos canais de Escala de Cinza, Vermelho, Verde, Azul, Matiz, Saturação e Intensidade para cada classe de amostras de misturas biodiesel/diesel B5 após a remoção dos zeros: algodão (linha azul), milho (linha vermelha), girassol (linha verde) e soja (linha amarela).	57
Figura 25. Gráficos dos escores PC1 x PC2 de PCA obtidos para (a) RGB + HSI + Escala de Cinza, (b) RGB, (c) HSI e (d) Intensidade para as quatro classes das amostras de misturas biodiesel/diesel B5: (●) algodão (●) girassol (●) milho, (●) soja.	58
Figura 26. Gráfico de escores de PLS-DA utilizando o sistema de cores RGB. Amostras de treinamento (círculo) e teste (triângulo) para as amostras de misturas B5 de algodão (azul), milho (vermelho), girassol (verde) e soja (amarelo).	62
Figura 27. Histograma médio do canal de Intensidade com as doze variáveis selecionadas (●) pelo SPA.	63
Figura 28. Funções Discriminantes obtidas pelo SPA-LDA na classificação das 120 amostras de misturas biodiesel/diesel B5: (●) algodão (●) girassol (●) milho, (●) soja.	65

LISTA DE ABREVIATURAS

ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
CMY	Ciano, Magenta, Amarelo
FD	Função discriminante
GA	Algoritmo Genético
HSI	Matiz, Saturação, Intensidade
KS	Kennard-Stone
LDA	Análise discriminante linear
NIR	Espectroscopia no infravermelho próximo
PCA	Análise de componentes principais
PCs	Componentes principais
PLS-DA	Análise discriminante pelos mínimos quadrados parciais
RGB	Vermelho, Verde, Azul
SIMCA	Modelagem Independente e Flexível por Analogia de Classe
SPA	Algoritmo das Projeções Sucessivas
TCC	Taxa de classificação correta
UV-VIS	Ultravioleta e visível

RESUMO

COSTA, Gean Bezerra. Universidade Estadual da Paraíba / Embrapa Algodão, fevereiro, 2015. **CLASSIFICAÇÃO MULTIVARIADA DE BIODIESEL B100 E B5 USANDO IMAGENS DIGITAIS**. José Germano Vêras Neto

Este trabalho propõe uma metodologia simples, rápida, de baixo custo e não destrutiva baseada na utilização de imagens digitais e técnicas de reconhecimento de padrões para a classificação de biodiesel (B100) e de misturas biodiesel/diesel (B5) com relação ao tipo de óleo de origem (algodão, girassol, milho e soja). Para isto, diferentes histogramas de cor (extraídos das imagens digitais) nos canais RGB, HSI, Escala de Cinza e suas combinações foram usados como informação analítica, a qual foi estatisticamente avaliada usando Modelagem Independente e Flexível por Analogia de Classe (SIMCA), Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e seleção de variáveis usando o Algoritmo das Projeções Sucessivas associado à Análise Discriminante Linear (SPA-LDA). Apesar dos bons desempenhos dos modelos de classificação SIMCA e PLS-DA, SPA-LSA forneceu os melhores resultados em termos de precisão, sensibilidade e especificidade para ambos conjuntos de treinamento e teste. As variáveis selecionadas pelo Algoritmo das Projeções Sucessivas continham claramente a informação necessária para a classificação de biodiesel (B100) e de misturas biodiesel/diesel (B5) de acordo como o óleo de origem. Isto é importante porque o produto final pode ter diferentes propriedades dependendo da variedade da matéria-prima utilizada, o que influencia diretamente na sua qualidade e, conseqüentemente, nos preços repassados ao consumidor. Além disso, vantagens intrínsecas são obtidas porque a análise é rápida, não requer uso de reagentes e evita caracterização química das amostras, isto é, a geração de resíduos é notadamente reduzida, contribuindo para um objetivo primário da Química Verde.

Palavras-chave: Biocombustíveis; Webcam; Quimiometria

ABSTRACT

COSTA, Gean Bezerra. Universidade Estadual da Paraíba / Embrapa Cotton, August, 2014. **MULTIVARIATE CLASSIFICATION OF BIODIESEL B100 AND B5 USING DIGITAL IMAGES.** José Germano Vêras Neto

This work proposes a simple, rapid, inexpensive, and non-destructive methodology based on digital images and pattern recognition techniques, for classification of biodiesel (B100) and biodiesel/diesel blends (B5) according to oil type (cottonseed, sunflower, corn, or soybean). For this, differing color histograms (extracted from digital images) in RGB, HSI, Grayscale channels, and their combinations were used as analytical information; which was then statistically evaluated using Soft Independent Modeling by Class Analogy (SIMCA), Partial Least Squares Discriminant Analysis (PLS-DA), and variable selection using the Successive Projections Algorithm associated with Linear Discriminant Analysis (SPA-LDA). Despite good performances by the SIMCA and PLS-DA classification models, SPA-LDA provided better results in terms of accuracy, sensitivity, and specificity for both the training and test sets. The variables selected Successive Projections Algorithm clearly contained the information necessary for classification of biodiesel (B100) and biodiesel/diesel blends (B5) according to oil type. This is important since a product may exhibit different properties, depending on the feedstock used. Such variations directly influence the quality, and consequently the price. Moreover, intrinsic advantages such as quick analysis, requiring no reagents, and a noteworthy reduction (the avoidance of chemical characterization) of waste generation, all contribute towards the primary objective of Green Chemistry.

Keywords: Biofuels; webcam; Chemometrics.

1. INTRODUÇÃO

O interesse por combustíveis renováveis tem se tornado cada vez maior nas últimas décadas, em contraposição aos derivados de petróleo (VÉRAS *et al.*, 2010), visto que o uso intensivo destes ao longo do tempo está causando danos ambientais em função do seu processo de extração e utilização (KNOTHE, 1999; BALABIN *et al.*, 2011; BALABIN e SAFIEVA, 2011). Assim, a produção de biocombustíveis é uma alternativa para diversificação da matriz energética e contribui para o meio ambiente por ser renovável, isento de metais pesados e com potencial para que possam ser substitutos, parciais ou totais, dos combustíveis fósseis (VÉRAS *et al.*, 2010). Dentre os biocombustíveis, tais como bioetanol, biogás e bio-óleo, destaca-se o biodiesel.

Considerado como uma alternativa ao diesel de petróleo em diversos países, o biodiesel vem sendo misturado com diesel em proporções diferentes, de acordo com a política energética, agrária e ambiental dos países que o introduziram na matriz energética nacional. São os casos da Argentina, com 10% de biodiesel no diesel, e da Alemanha, que permite ao consumidor final a proporção da mistura que desejar. No Brasil, a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) estabeleceu normas para adição de 5% de biodiesel no diesel de petróleo, mistura denominada B5, e para certificação da qualidade do biodiesel (BAKEAS *et al.*, 2011).

A partir das políticas nacionais e da comunidade europeia, observou-se uma crescente demanda da utilização do biodiesel, o que conduz à procura de diversas matérias-primas, incluindo os óleos vegetais e gorduras de origem animal (ANDERSEN, WEINBACH, 2010; FORE *et al.*, 2011). Soja (*Glycine max*), girassol (*Helianthus annuus*) e algodão (*Gossypium hirsutum latifolium* Hutch LR) são considerados as principais fontes de biomassa para produção do biodiesel no Brasil, além de outras oleaginosas em potencial, tais como mamona (*Ricinus communis* L.), pinhão-manso (*Jatropha curcas* L.), amendoim (*Arachis hypogaea*), canola (*Brassica napus* L.), dendê (*Elaeis guineensis* Jacq.), coco (*Cocos nucifera*), colza (*Brassica napus*), castanha-do-pará (*Bertholletia excelsa*), babaçu (*Attalea speciosa*), tucumã (*Astrocaryum aculeatum* Meyer), inajá

(*Attalea maripa*), macaúba (*Acrocomia* spp.), licuri (*Syagrus coronata*) e buriti (*Mauritia flexuosa*) (SUARES *et al.*, 2009; BERGMANN *et al.*, 2013; COSTA *et al.*, 2013).

Uma consequência do aumento da produção de biodiesel e da diversificação de suas matérias primas é a relevância que o controle de qualidade assume. Recentemente, alguns trabalhos foram publicados mostrando a importância do desenvolvimento de métodos para determinar o tipo de óleo da produção de biodiesel (VÉRAS *et al.*, 2010; BALABIN *et al.*, 2011; VÉRAS *et al.*, 2012; SILVA, *et al.*, 2012; INSAUSTI, *et al.*, 2012; VÉRAS *et al.* 2013).

A este respeito, Veras *et al.* desenvolveram metodologias analíticas utilizando as espectroscopias no infravermelho próximo (NIR, do inglês: *Near Infrared Spectroscopy*) (VÉRAS *et al.*, 2010) e no ultravioleta-visível (UV-vis) (VÉRAS *et al.*, 2012) para classificar amostras de biodiesel a partir do seu óleo de origem utilizando Modelagem Independente e Flexível por Analogia de Classe (SIMCA, do inglês: *Soft Independent Modeling by Class Analogy*) como técnica de reconhecimento de padrões (classificação multivariada).

Balabin *et al.* (2011) classificaram dez grupos de amostras de biodiesel a partir do óleo vegetal de origem utilizando espectroscopia NIR e técnicas de classificação multivariadas como Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA, do inglês *Partial Least Square Discriminant Analysis*) e K-ésimo Vizinho mais Próximo (KNN, do inglês *K-nearest Neighbors*).

Insausti *et al.* (2012) descreveram um estudo comparativo entre as espectroscopias NIR, UV-vis e fluorescência sincrônica (FS) para classificar amostras de biodiesel de girassol, milho e canola utilizando a Análise Discriminante Linear (LDA, do inglês *Linear Discriminant Analysis*) associada ao Algoritmo das Projeções Sucessivas (SPA, do inglês *Successive Projections Algorithm*) como técnica de seleção de variáveis.

Silva *et al.* (2012) realizaram o rastreamento das misturas biodiesel/diesel com respeito ao óleo de partida utilizando espectroscopia NIR associada com Modelagem Independente e Flexível por Analogia de Classe (SIMCA) e Análise de Componentes Principais associada à Análise Discriminante Linear (PCA-LDA, do inglês *Principal Component Analysis-Linear Discriminant Analysis*) e modelos utilizando seleção de variáveis pelo Algoritmo Genético (GA, do inglês *Genetic algorithm*) e Algoritmo das Projeções Sucessivas (SPA) acoplados ao LDA.

Véras *et al.* (2013) determinaram o tipo de óleo de origem de misturas ternárias utilizando espectroscopia NIR associada a ferramentas quimiométricas de classificação multivariada, tais como SIMCA, PLS-DA e SPA-LDA.

É preciso enfatizar que as normas nacionais e internacionais apresentam metodologias analíticas para avaliação da qualidade do biodiesel e da mistura biodiesel/diesel que utilizam

Cromatografia Gasosa (LEE *et al.*, 1998), Cromatografia Líquida de Alta Eficiência (EL-HAMDY e EL-FIZGA, 1995) e Ressonância Magnética Nuclear (SACCHI *et al.*, 1996) como métodos analíticos de referência estabelecidos pela ASTM (*American Society of Testing and Materials*), através da norma ASTM D6751, e pela União Europeia, através da norma EN 14214 do Comitê Europeu de Normalização (*Comité Européen de Normalisation - CEN*) (KNOTHE, 2005). No entanto, estas técnicas são destrutivas e possuem custo elevado de instrumentos e de manutenção, bem como requerem etapas de pré-tratamento de amostras (KNOTHE, 2006).

Desta forma, faz-se necessário o desenvolvimento de uma metodologia alternativa que seja simples, rápida, barata e não destrutiva para a classificação de biodiesel a partir do óleo de origem. Esta metodologia pode ser realizada baseando-se na análise de imagens digitais associadas a técnicas de reconhecimento de padrões e seleção de variáveis.

A imagem digital como ferramenta analítica tem se tornando cada dia mais importante devido a sua capacidade de realizar análises rápidas, confiáveis e de baixo custo. Elas podem ser produzidas por diversos equipamentos, incluindo câmeras digitais (BRIONES e AGUILERA, 2005), scanners (VAN DALEN, 2004), aparelhos celulares (KONG e TAN, 2012), webcam (BRIONES e AGUILERA, 2005), dentre outros. Além disso, outro fator positivo do uso de imagem digital é que ele substitui, de modo geral, o sistema visual humano (DU e SUN, 2004), eliminando o caráter subjetivo da análise, a qual é substancialmente influenciada por condições ambientais e sujeita a inconsistências (ANTONELLI *et al.*, 2004).

Neste sentido, diversos sistemas de cor são utilizados para tratar os dados provenientes de imagens digitais e têm como finalidade facilitar a especificação das cores de forma padronizada e de aceitação geral (GONZALES e WOODS, 2000). Essencialmente, um sistema de cor é uma especificação de um sistema espacial de coordenadas tridimensionais onde cada cor é representada por um único ponto (GONZALES e WOODS, 2000). Existem diversos sistemas de cor, sendo o RGB (Vermelho, Verde e Azul, do inglês: *red-green-blue*), o HSI (Matiz, Saturação e Intensidade, do inglês: *hue-saturation-intensity*) e a Escala de Cinza os mais utilizados para fins analíticos (GARRIDO-NOVELL *et al.*, 2012; ULRICI *et al.*, 2012).

O sistema RGB é baseado no princípio óptico de que os diferentes efeitos cromáticos são obtidos por projeções de luz branca, onde todas as cores visíveis podem ser criadas utilizando as cores primárias aditivas: vermelho, verde e azul. Os atributos HSI descrevem a formação de cores de uma maneira muito mais próxima àquela pela qual o sistema visual humano as percebe, em que a saturação ou pureza expressa o intervalo de comprimentos de onda (λ) ao redor do λ médio no qual a energia é transmitida ou refletida, o Matiz de um objeto é a medida do λ médio da luz que ele reflete ou emite definindo, portanto, a cor do objeto. Já a Intensidade é a medida de energia

total envolvida em todos os comprimentos de onda, sendo responsável pela sensação de brilho dessa energia incidente sobre o olho.

Vale a pena notar que as cores puras do espectro são completamente saturadas. Matiz e saturação em conjunto são chamados de cromaticidade e estão intimamente relacionadas com a percepção de cor humana. Portanto, uma cor pode ser caracterizada pela sua cromaticidade e luminosidade. O único atributo da luz acromática é a intensidade, cuja medida escalar produz a Escala de Cinza. A Escala de Cinza tem 256 níveis, onde o valor 0 (zero) representa o preto e o valor 255 representa branco (SANTOS e PEREIRA-FILHO, 2014; DOMINGUEZ *et al.*, 2014).

Para fins analíticos, imagens digitais foram decompostas em histogramas de cor, que descrevem a distribuição estatística dos pixels como uma função do componente de cor gravada, e foram utilizados com sucesso como dados de entrada para classificação de chás (DINIZ *et al.*, 2013), méis (DOMÍNGUEZ *et al.*, 2014), bactérias (ALMEIDA *et al.*, 2014) e cafés (SOUTO *et al.*, 2015). Neste sentido, histogramas de cor de RGB, HSI, Escala de Cinza e suas combinações foram utilizados como informação analítica e avaliados estatisticamente por meio de diferentes classificadores multivariados: Modelagem Independente e Flexível por Analogia de Classe (SIMCA), Análise de Componentes Principais associada à Análise Discriminante Linear (PCA-LDA), Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA) e Análise Discriminante associada ao Algoritmo das Projeções Sucessivas (SPA-LDA).

O presente trabalho propõe uma nova metodologia analítica para a classificação de amostras de biodiesel (B100) e de misturas biodiesel/diesel (B5) a partir do seu óleo de origem utilizando imagens digitais e técnicas de reconhecimento de padrões supervisionadas. Para tanto, a Modelagem Flexível Independente por Analogia de Classes (SIMCA), a Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e a Análise Discriminante Linear associada com seleção de variáveis pelo Algoritmo das Projeções Sucessivas (SPA) foram avaliadas.

2. REVISÃO DA LITERATURA

2.1. Biodiesel

O biodiesel é um combustível biodegradável oriundo de matriz renovável, menos tóxico se comparado a combustíveis fósseis, tais como o diesel e a gasolina, podendo ser produzido a partir de uma reação de transesterificação de óleo vegetal ou gordura animal. É um combustível para uso em motores a combustão por compressão, podendo substituir parcial ou totalmente o óleo diesel (FRANCESCHINI e MACCHIETTO, 2008; SENGO *et al.*, 2010).

Uma gama de matérias-primas vem sendo utilizada para a produção de biodiesel, dentre elas soja, girassol, amendoim, mamona, dendê e canola (GERIS *et al.*, 2007; BERGMANN *et al.*, 2013). Existem também espécies menos conhecidas sendo considerado para a produção de biodiesel, incluindo macaúba, buriti e licuri (BERGMANN *et al.*, 2013).

Físico-quimicamente semelhante ao óleo diesel, o biodiesel é perfeitamente miscível e pode ser utilizado puro ou misturado ao diesel em quaisquer proporções em motores do ciclo diesel. As misturas binárias de biodiesel e óleo diesel de petróleo são designadas pela abreviação BX, onde X é a proporção em volume de biodiesel na mistura (COSTA NETO *et al.*, 2000).

No cenário ambiental, o biodiesel apresenta vantagens claras sobre os combustíveis convencionais por serem provenientes de fontes renováveis, biodegradáveis, não tóxicos e reduzem significativamente a emissão de CO₂ e SO_x durante o processo de combustão (CORDEIRO *et al.*, 2011).

2.2. Misturas biodiesel/diesel

Em decorrência da crescente demanda por combustíveis de fontes renováveis e do potencial brasileiro em atender parte expressiva dessas necessidades (BALABIN e SAFIEVA, 2011), o

Governo Federal lançou em 2004 o Programa Nacional de Produção e Uso do Biodiesel. Este programa interministerial teve como objetivo a implantação, de forma sustentável, da cadeia de produção e uso do biodiesel, visando à geração de emprego e renda na agricultura familiar.

Em 2005, com a publicação da Lei Federal N° 11.097 foi autorizada a adição de 2% (em volume) de biodiesel ao diesel mineral, mistura conhecida como B2, que passou a ser obrigatória em 2008. Segundo esta lei, o percentual de biodiesel adicionado ao diesel aumentaria para 5% apenas em 2013 (BAKEAS *et al.*, 2011). Entretanto, essa obrigatoriedade foi antecipada e, desde janeiro de 2010 a mistura B5 vem sendo comercializada em todo o território brasileiro. Com a expansão da capacidade produtiva do biodiesel no Brasil, há uma tendência ao longo dos próximos anos da comercialização de misturas com teores cada vez mais elevados de biodiesel, mediante a autorização da ANP.

2.3. Imagens digitais

A área de processamento digital de imagens tem crescido consideravelmente nas últimas duas décadas, com a utilização de imagens e gráficos em uma grande variedade de aplicações, aliado ao fato de que a tecnologia de construção de computadores também tem se aprimorado, possibilitando a utilização de sistemas mais eficientes e baratos. Muitas áreas vêm utilizando sistemas de processamento digital de imagens, tais como: reconhecimento de padrões (indústria), medicina, agricultura, pesquisas espaciais, meteorologia, química, etc. (TRAINA e OLIVEIRA, 2004).

As imagens digitais podem ser produzidas por diversos aparelhos, incluindo câmeras digitais, scanners, aparelhos de raios-X, microscópios eletrônicos, aparelhos de ultrassom, radares, celulares, webcam, dentre outros dispositivos (MACHADO e SOUKI, 2004).

Como um computador não é capaz de processar imagens contínuas, pois este trabalha com números discretos, se faz necessário que as imagens sejam, primeiramente, convertidas em números na forma de matriz como arranjos bidimensionais de pontos (MARQUES FILHO e VIEIRA NETO, 1999). Para fazer a conversão de imagem em números, a imagem é subdividida em uma grade contendo milhões de quadrados de igual tamanho, sendo cada um deles associado a um valor numérico da intensidade da cor naquele ponto. Cada quadrado na imagem é denominado *pixel* (elemento de imagem) (ALBUQUERQUE, 2002). Os valores desses *pixels* representam uma discretização da intensidade luminosa captada por uma determinada área, variando em valores ou porcentagem, dependendo do espaço de cor utilizado (GONZALEZ e WOODS, 2000).

2.3.1. Pixel

Em uma imagem digitalizada, cada pixel tem um único valor numérico correspondente ao nível de sinal naquele ponto (SOLOMON e BRECKON, 2011).

A conversão do conjunto numérico formado pelos pixels a uma imagem é realizada por meio de um mapa de cores, onde é designado um nível (ou tom) específico de cor para cada ponto numérico da imagem. Os níveis das cores variam num intervalo de 0 a um valor máximo, de acordo com o nível do sinal (MARQUES FILHO e VIEIRA NETO, 1999; SOLOMON e BRECKON, 2011).

O conceito de qualidade de uma imagem digital está relacionado a dois parâmetros principais: a resolução espacial e a quantidade de bits utilizada para quantificação da informação da imagem (MARQUES FILHO e VIEIRA NETO, 1999). A resolução espacial expressa o tamanho da matriz correspondente à imagem. De forma simplificada, quantas linhas e colunas, ou seja, quantos *pixels* são usados para cobrir o espaço visual captado pela imagem. Normalmente é representada por $M \times N$, como por exemplo, 2880×1620 (SOLOMON e BRECKON, 2011).

O número de bits define o número possível de valores de intensidades ou cores que um *pixel* pode assumir (SOLOMON e BRECKON, 2011). Por exemplo, em 1 *bit* por *pixel* ($2^1 = 2$), apenas duas cores podem ser assumidas: preto e branco. No caso das imagens no sistema RGB, a utilização de 24 bits permite alocar 8 bits para cada componente da cor, totalizando 256 níveis de cores por componente e 16,7 milhões de cores totais na imagem.

2.3.2. Histograma

O histograma de uma imagem é a distribuição de frequência dos valores que um *pixel* pode assumir. Trata-se do percentual de pixels naquela imagem que tem um dado valor de intensidade ou de cor. No caso de uma imagem em Escala de Cinza, por exemplo, o histograma revela o número de vezes que cada valor de cinza (0 – 255) ocorre dentro da imagem (MARQUES FILHO e VIEIRA NETO, 1999; QUEIROZ e GOMES, 2001; SOLOMON e BRECKON, 2011).

Segundo Solomon *et al.* (2011), após a normalização o histograma da imagem pode ser entendido como uma função de densidade de probabilidade discreta que define a probabilidade de um dado valor de *pixel* ocorrer dentro da imagem.

Na Figura 1 é mostrada uma imagem colorida e os histogramas resultantes de sua decomposição nas componentes de cores Vermelho, Verde, Azul e Escala de Cinza.

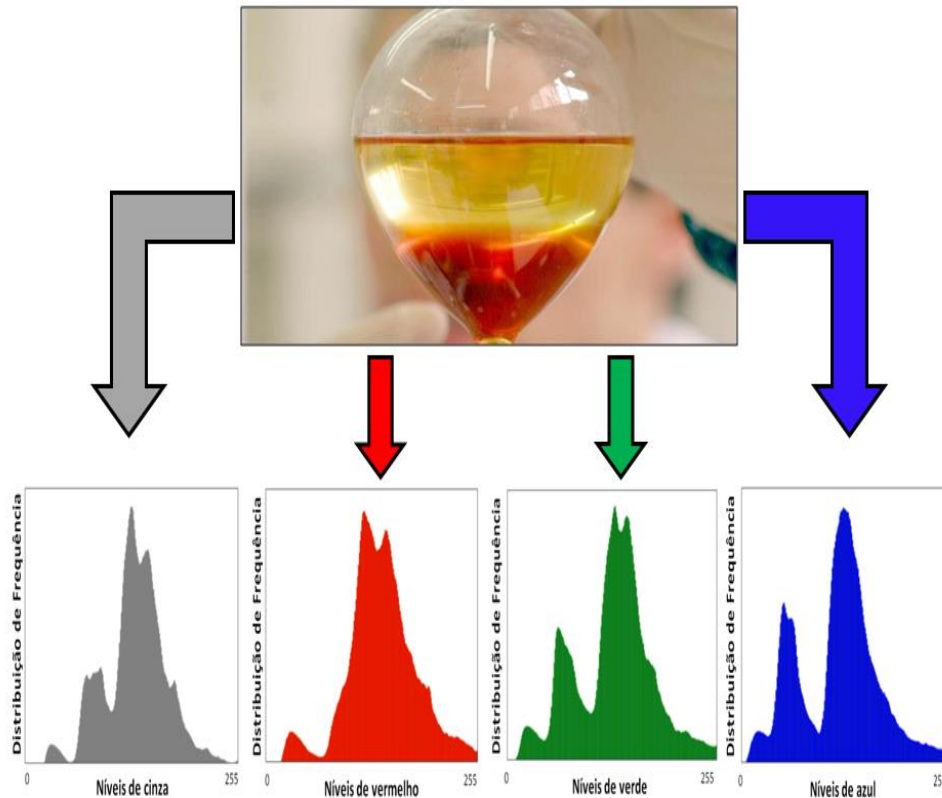


Figura 1. Histogramas referentes à distribuição de frequência de todos os pixels nos níveis de Cinza, Vermelho, Verde e Azul da imagem apresentada.

Fonte: Próprio autor

2.4. Sistemas de cores

Os sistemas de cores foram criados para permitir a especificação das cores em um formato padronizado. Em termos gerais um sistema de cores é uma representação tridimensional na qual cada cor é representada por um ponto no sistema de coordenadas 3D. (GONZALEZ e WOODS, 2000).

Um sistema de cores é um modelo que explica as propriedades ou comportamento das cores num contexto particular. Entretanto, não existe um sistema que explique todos os aspectos relacionados à cor (AZEVEDO e CONCI, 2003). Por isso, são utilizados sistemas diferentes para ajudar a descrever as diferentes características da cor que são percebidas pelo ser humano. Existem vários sistemas de cores, tais como RGB (Vermelho, Verde e Azul, do inglês: *red-green-blue*), CMYK (ciano, magenta, amarelo e preto, do inglês: *cyan-magenta-yellow-black*), HSI (Matiz, Saturação e Intensidade, do inglês: *hue-saturation-intensity*), HSV (Matiz, Saturação e

Luminância, do inglês: *hue-saturation-value*), HSL (Matiz, Saturação e Luminosidade, do inglês: *hue-saturation-lightness*) e Escala de Cinza (do inglês: *grayscale*) (MACHADO, 2006).

O universo de cores que pode ser reproduzido por um sistema é chamado de espaço de cores, sendo este definido como um método formal, necessário para se quantificar as sensações visuais das cores que podem, assim, ser mais precisamente especificadas (AZEVEDO e CONCI, 2003).

2.4.1. Sistemas de cores RGB

A mistura das cores dominantes, também chamadas de primárias aditivas, em diferentes combinações e níveis variados de intensidade pode simular todas as cores existentes no espectro da região do visível. Se a luz refletida contém a máxima intensidade das luzes vermelha, verde e azul, o olho percebe o branco, e se não existe luz, é percebido o preto (BATISTA, 2004). Combinando duas cores aditivas primárias puras será produzida uma cor secundária (ou primária subtrativa), conforme o diagrama mostrado na Figura 2. As cores secundárias, ciano, magenta e amarelo, são as cores opostas ao vermelho, verde e azul, respectivamente.

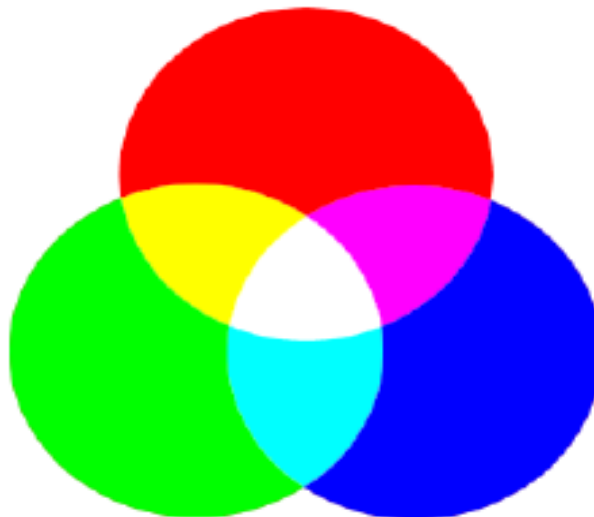


Figura 2. Cores primárias aditivas e as cores resultantes de suas combinações R: Vermelho; G: Verde; B: Azul.

Fonte: <http://processing.org/learning/color/>

Acesso em: dezembro de 2014

Conforme Souto (2000), a luz branca é produzida se os três círculos coincidirem, sendo uma composição equitativa entre as cores primárias. As cores primárias não podem ser produzidas pela mistura de duas delas, por isso são definidas como cores primárias aditivas. Outras cores são produzidas quando duas cores se misturam:

- ✓ vermelho + azul = magenta
- ✓ vermelho + verde = amarelo
- ✓ verde + azul = ciano

Com a variação da quantidade relativa das cores primárias, uma enorme gama de cores secundárias pode ser produzida, se incluir diversos tons a cada uma delas. Usando-se filtros, as cores podem ser subtraídas da luz branca:

- ✓ subtração de vermelho = ciano
- ✓ subtração de verde = magenta
- ✓ subtração de azul = amarelo

O sistema RGB pode ser representado por um sistema de coordenadas cartesianas, a partir do qual é gerado um cubo sólido, conhecido como cubo de cores (Figura 3) (AZEVEDO e CONCI, 2003).

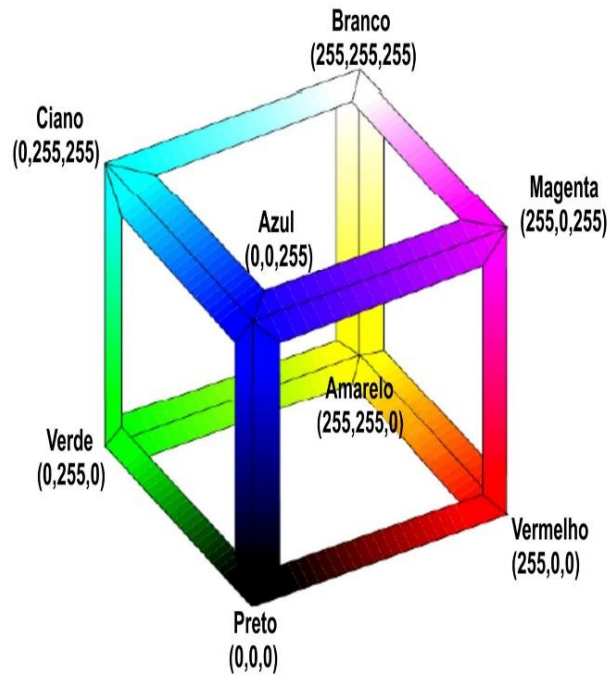


Figura 3. Cubo de cores RGB.

Fonte: Adaptada de Gonzalez, Woods (2000).

O cubo possui os valores de RGB nos três vértices dos eixos do sistema de coordenadas e ciano, magenta e amarelo estão nos outros três vértices. O preto está na origem $(0,0,0)$ e branco está no vértice mais distante da origem $(255,255,255)$. Neste sistema, a Escala de Cinza estende-se do preto até o branco ao longo da linha juntando estes dois pontos, e as cores são pontos sobre ou dentro do cubo definidos por vetores estendendo-se a partir da origem. Assume-se que todos os valores de cor foram normalizados, de modo que o cubo representado é unitário. Portanto, todos os valores de cores são assumidos estar no intervalo de $[0,255]$ (GONZALEZ e WOODS, 2000).

2.4.2. Sistema de cores HSI

Os atributos HSI descrevem a formação de cores de uma maneira muito mais próxima àquela pela qual o sistema visual humano as percebe e, devido a isso, este sistema apresenta vantagens em relação ao sistema RGB. O Matiz de um objeto é a medida do comprimento de onda (λ) médio da luz que ele reflete ou emite, definindo, portanto, a cor do objeto. A saturação, ou pureza, expressa o intervalo de λ 's ao redor do λ médio no qual a energia é transmitida ou refletida. Já a intensidade é a medida de energia total envolvida em todos os comprimentos de onda, sendo, portanto, responsável pela sensação de brilho dessa energia incidente sobre o olho. (GONZALEZ e WOOD, 2000; BATISTA, 2004).

No sistema HSI pode ser representado por um prisma circular como ilustrado na Figura 4, em que os planos horizontais originam círculos. Nos vértices contém as cores primárias e, no centro, há uma combinação das três cores em proporções iguais. Essas combinações podem estar próximo do preto ou do branco dependendo da altura em que o plano foi efetuado, propiciando, diferentes intensidades luminosas claras ou escuras.

A Matiz representa um ângulo de 0° no vermelho, 120° no verde e no azul 240° , a saturação pode ser definida em um valor entre 0 e 1 grau e a intensidade é um valor não negativo. Na Matiz há um movimento circular do espectro partindo de 0° e quando há um giro no sentido horário ou anti-horário, diferentes matizes ou tonalidades e ângulos são obtidos. (BATISTA, 2004). A Saturação, como pode ser observado e é delimitada pela distância da região central até bordas do círculo

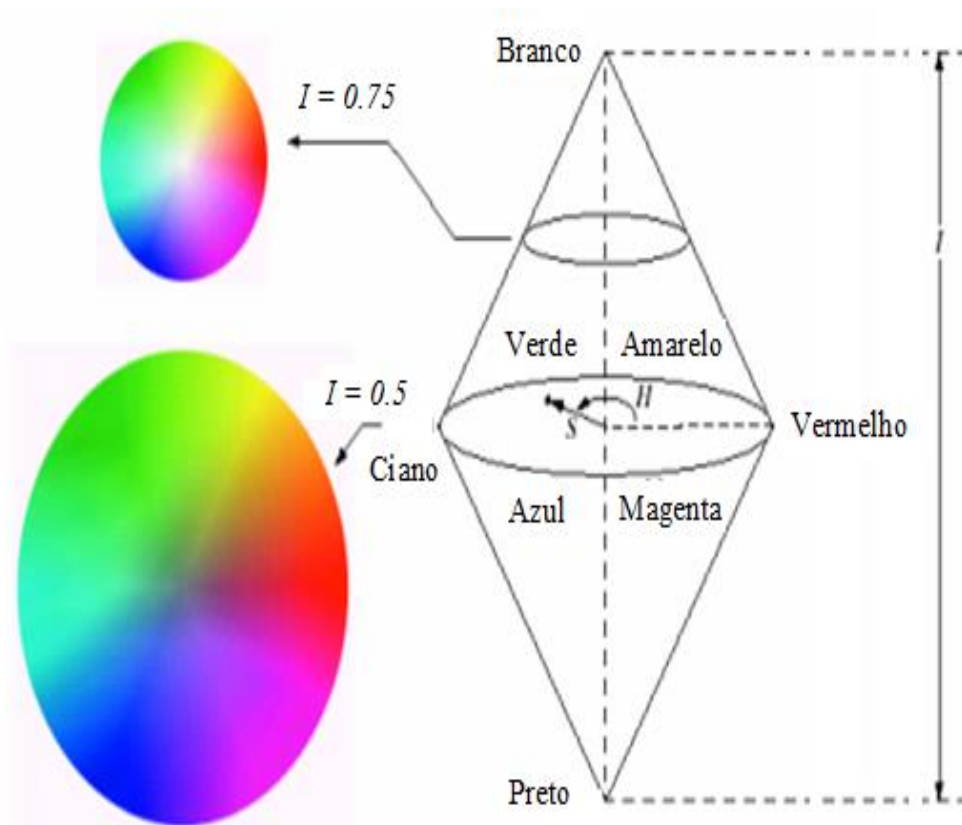


Figura 4. Sistema HSI de cores

Fonte: <http://www.csie.ntpu.edu.tw/~dalton/course/DIP/lecture/Chapter06.pdf>

Acesso em: fevereiro de 2014

2.4.3. Escala de Cinza

Imagens em Escala de Cinza (do inglês: *Gray Scale*) assumem valores dentro de uma escala de 256 níveis onde o valor 0 (zero) representa o preto e o valor 255 representa o branco. Os demais tons de cinza estão entre esses dois valores (GONZALEZ e WOODS, 2000). Por causa da forma de armazenamento no computador essas imagens são, normalmente, chamadas imagens de 8 *bits*. Um exemplo oportuno de imagens em Escala de Cinza são as radiológicas. Para melhor exemplificar essa escala de cores faz-se a multiplicação da resolução de uma pequena imagem radiográfica (320 x 240 *pixels*) para se chegar ao tamanho final requerido para sua representação visual: 76.800 *bytes* onde cada *byte* representa 8 *bits*. A Figura 5 apresenta alguns dos 256 tons da Escala de Cinza.

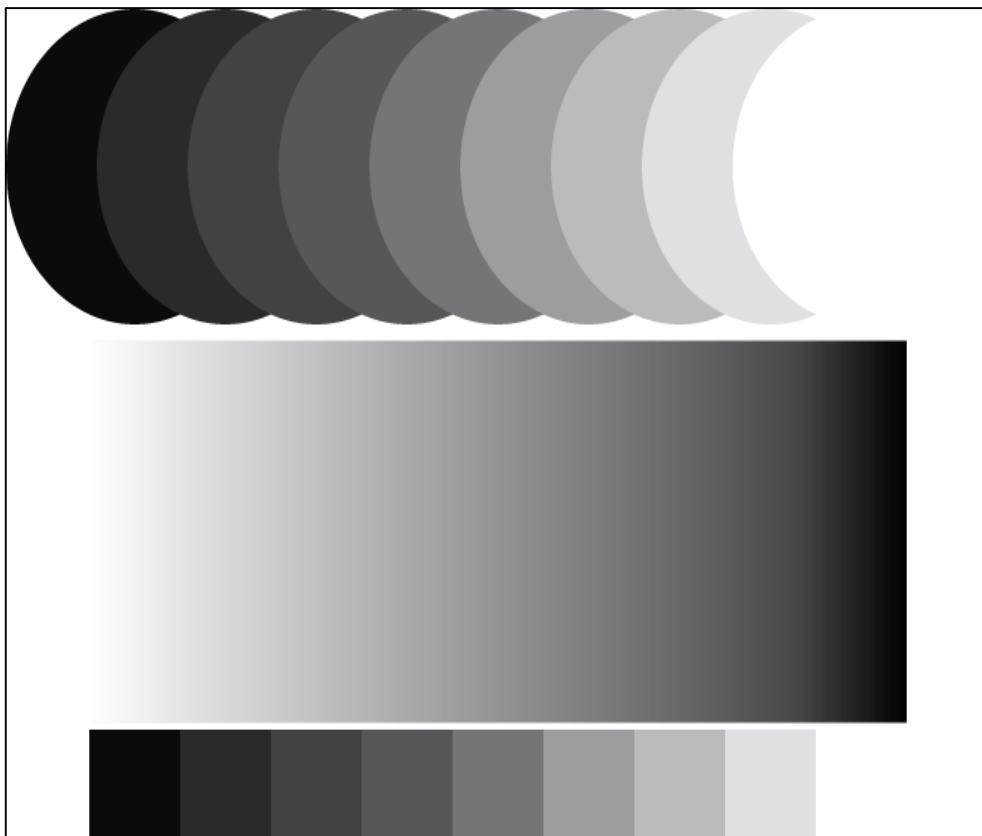


Figura 5. Escala dos níveis de Cinza.

Fonte: <https://www.behance.net/gallery/6097929/Pyramid-Module-Value-Grayscale>

Acesso em: dezembro de 2014

2.5. Quimiometria

A Quimiometria pode ser definida como a utilização de métodos matemáticos e estatísticos para o tratamento de dados químicos de forma a extrair uma maior quantidade de informações e melhores resultados analíticos (BLANCO e ALCALA, 2006). A diferença da Quimiometria para as demais ciências metrológicas reside nas características dos dados serem de origem química, que possuem menor quantidade de variáveis, os quais em sua maioria são controláveis, de fácil repetição nas condições de trabalho desejadas, etc.

A Quimiometria está dividida em algumas áreas principais muito pesquisadas e aplicadas atualmente, tais como processamento de sinais analíticos, planejamento e otimização de experimentos, classificação e reconhecimento de padrões, calibração multivariada, monitoramento e modelagem de processos multivariados e métodos de inteligência artificial (PIMENTEL *et al.*, 1998). A Quimiometria pode ser aplicada em todas as etapas de uma análise no sentido de melhorar o rendimento e dar melhor qualidade nos resultados.

2.5.1. Técnicas de reconhecimento de padrões

São técnicas responsáveis por identificar as semelhanças e diferenças entre as amostras, a partir do grau de similaridade entre suas propriedades físicas e químicas, comparando-as entre si. As técnicas quimiométricas de reconhecimento de padrões (RP) promovem os agrupamentos dos objetos, possibilitando, inclusive, a identificação de amostras anômalas (do inglês: *outliers*). Dentre as técnicas de RP estão as não supervisionadas e as supervisionadas (MASSART *et al.*, 1997).

2.5.1.1. Técnicas de reconhecimento de padrões não supervisionadas

Os métodos de reconhecimento de padrões não supervisionados são utilizados quando não se tem informação anterior sobre as classes as quais pertencem os objetos, ou seja, uma avaliação prévia dos dados sem sofrer influência do rótulo das amostras é realizada. Dentre estes, os mais conhecidos são: Análise por Componentes Principais (PCA, do inglês: *Principal Component Analysis*) e Análise Hierárquica de Agrupamentos (HCA, do inglês: *Hierarchical Cluster Analysis*) (CASTRO e PRADO 2002; CORREIA e FERREIRA, 2007).

2.5.1.1.1. Análise de Componentes Principais (PCA)

A Análise por Componentes Principais (PCA) é um dos métodos mais comumente empregados na análise de informações qualitativas e é frequentemente usada para detectar padrões nos dados medidos (FERREIRA, 2002).

A PCA é realizada sob uma matriz multivariada de dados $\mathbf{X}_{n \times m}$, seja esta original ou pré-processada, onde os n objetos (amostras) encontram-se nas linhas e as m variáveis nas colunas (MARTENS, NAES, 1996). A matriz \mathbf{X} é decomposta em dois subespaços, o de escores (t ; do inglês: *scores*) e o de pesos (l ; do inglês: *loadings*) (SABIN *et al.*, 2004).

O processo de decomposição se repete de modo que a primeira decomposição recebe o nome de “primeira componente principal”, $PC1(t_1, l_1)$, e a segunda decomposição de “segunda componente principal”, $PC2(t_2, l_2)$, de forma que diferentes componentes são obtidas sucessivamente (Figura 6) (THOMAS, 1994), as quais são representadas pelas equações 1 e 2:

$$X = T_1 l_1^T + T_2 l_2^T + T_A l_A^T + E \quad (1)$$

$$X = TL^T + E \quad (2)$$

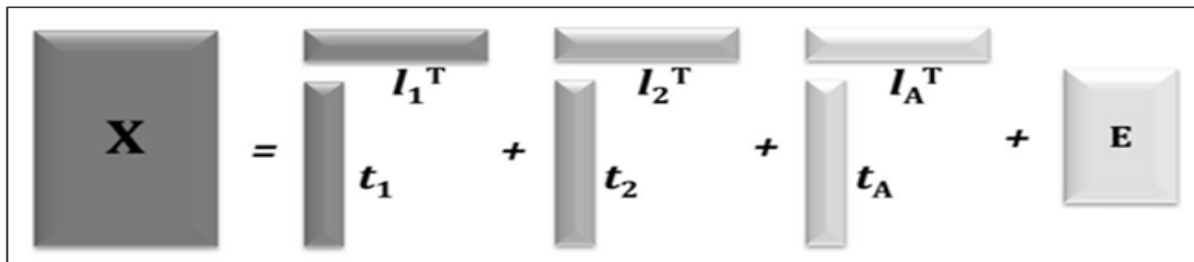


Figura 6. Decomposição da matriz X em Componentes Principais

Fonte: Adaptada de ESBENSEN (2009).

Com o método PCA é possível (BEEBE *et al.*, 1998):

- ✓ demonstrar as diferenças entre as formas das amostras;
- ✓ definir quais variáveis contribuem mais para estas diferenças;
- ✓ identificar quais variáveis contribuem de forma correlacionada ou independentemente uma da outra;
- ✓ detectar padrão de amostras (agrupamentos);
- ✓ diferenciar informações úteis de ruídos, etc.

Neste sentido, a natureza multivariada dos dados pode ser visualizada em poucas dimensões. A primeira componente principal (PC1) representa a máxima variância associada a todas as combinações lineares e cada componente sucessiva (PC2, PC3, PC4, ..., PCn) representa a variância não representada por PC1 e assim sucessivamente (Figura 7) (MARTENS, NAES, 1996).

A Análise de Componentes Principais (PCA) é uma técnica muito utilizada e serve de base para a maioria das demais técnicas quimiométricas multivariadas, inclusive as técnicas de reconhecimento de padrões supervisionado como o SIMCA (MARTENS, NAES, 1996).

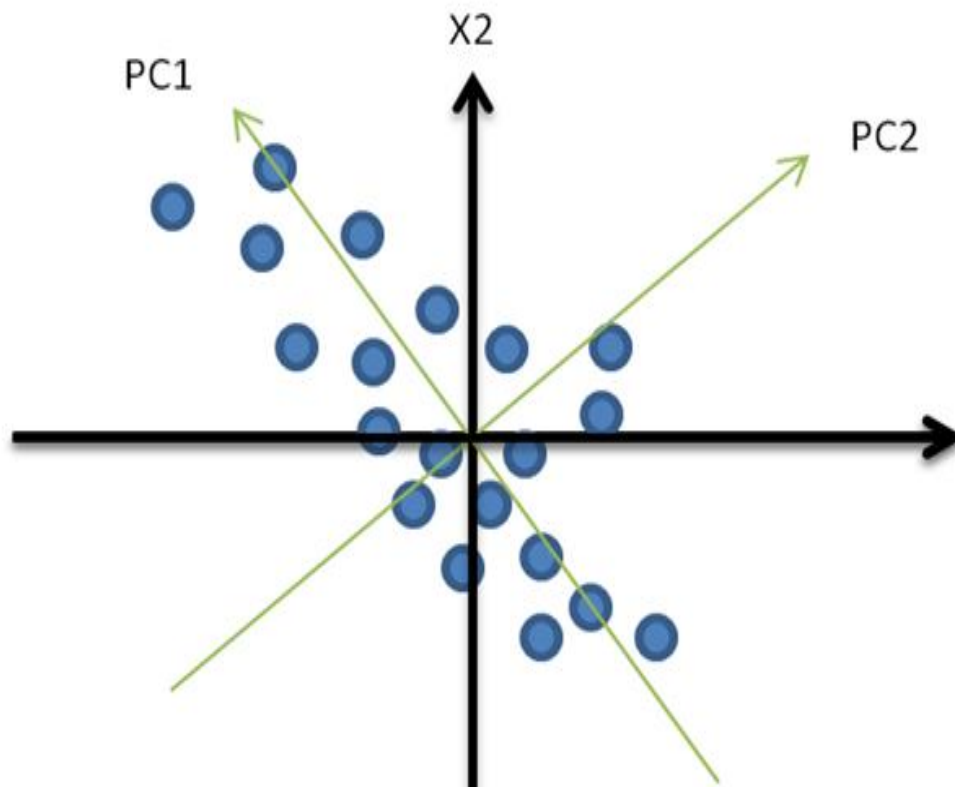


Figura 7. Posicionamento dos parâmetros da PCA entre duas variáveis originais.

Fonte: Adaptada de BEEBE (1998).

2.5.1.2. Técnicas de reconhecimento de padrões supervisionadas

Os métodos supervisionados são utilizados quando se tem alguma informação prévia sobre as classes e esta é usada para a construção do modelo. Dentre estas técnicas destacam-se: a

Modelagem Independente e Flexível por Analogia de Classes (SIMCA), o K-ésimo vizinho mais próximo (KNN) e a Análise Discriminante Linear (LDA) (BEEBE *et al.*, 1998).

2.5.1.2.1. Modelagem Independente e Flexível por Analogia de Classe (SIMCA)

A Modelagem Independente e Flexível por Analogia de Classes (SIMCA) é um método de reconhecimento de padrões baseado na similaridade (BEEBE *et al.*, 1998). Ela considera, para a classificação, informações da distribuição da população, estima um grau de confiança desta classificação e pode prever novas amostras como pertencentes a uma ou mais classes ou ainda a nenhuma das classes (FERREIRA *et al.*, 1999).

O SIMCA objetiva criar um espaço limitado para cada classe. Isto pode ser mais bem compreendido para uma classe descrita por duas componentes principais (THOMAS, 1994). O poder de discriminação dos modelos SIMCA em distinguir as classes é dado pela distância e pelos resíduos entre classes (CALAS *et al.*, 1997). A atribuição de uma amostra teste a uma dada classe é baseada na projeção da mesma no espaço dos escores e sua distância das fronteiras da classe em questão (SRIVASTAVA e CARTER, 1983). Isto é repetido para todas as classes e, no final, se os modelos não tiverem poder de discriminação suficiente, a amostra pode ser membro de mais de uma classe (HAIR *et al.*, 1995). O outro extremo também pode acontecer, quando a amostra é anômala (suficientemente diferente) em relação ao conjunto de treinamento e, neste caso, ela não pertence às classes definidas. Neste sentido, na construção do modelo podem ocorrer dois tipos de erros: erros tipo I, onde uma dada amostra não é incluída na sua própria classe, e erros tipo II, onde uma amostra é incluída em uma classe incorreta (FERREIRA, 2002).

2.5.1.2.2. Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA)

A Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA) é um método de reconhecimento de padrões supervisionado baseado na regressão PLS para construção de modelos preditivos que buscam relacionar as respostas instrumentais (X) com as propriedades de interesse, ou variáveis de resposta (y), utilizando os rótulos das classes como um vetor dependente y (PONTES *et al.*, 2011; ALMEIDA *et al.*, 2013). Para problemas de classificação envolvendo mais de duas classes, o algoritmo PLS2 é utilizado e o número de variáveis de resposta é correspondente ao número de classes envolvidas (Figura 9) (SILVA *et al.*, 2012).

Matriz X	Matriz Y			
	y ₁	y ₂	y ₃	y ₄
Classe 1	1	0	0	0
Classe 2	0	1	0	0
Classe 3	0	0	1	0
Classe 4	0	0	0	1

Figura 8. Representação de um modelo preditivo ideal de PLS-DA para 4 classes distintas.

Fonte: Próprio autor

Uma sequência de operações matemáticas decompõe a matriz X , formada pelas variáveis de resposta instrumental, em variáveis latentes, a fim de reduzir a dimensionalidade do conjunto de dados. A escolha do número ideal de variáveis latentes é realizada com base no método de validação cruzada das amostras de calibração (ALMEIDA *et al.*, 2013).

2.5.1.2.3. Análise Discriminante Linear (LDA)

A Análise Discriminante Linear (LDA) é um dos métodos de reconhecimento de padrões mais utilizados e consiste na determinação de funções discriminantes (DF, do inglês *discriminant function*) lineares que maximizam a variância entre as classes e minimizam a variância dentro das classes (BRUNS e FAIGLE, 1985; MASSART *et al.*, 1997).

A Figura 8 mostra um esquema de obtenção de uma DF linear e como opera a classificação de um dado elemento. Os centros de gravidade (os asteriscos) dos grupos formados por círculos e quadrados são inicialmente determinados. Em seguida traça-se um plano de decisão equidistante e perpendicular à linha que une os centros de gravidade dos grupos, formando as regiões R1 e R2. Se um ponto qualquer estiver presente na região R1, o mesmo é classificado como pertencente ao grupo 1. Caso contrário, obrigatoriamente pertencerá à categoria 2 (GAMBARRA NETO *et al.*, 2009).

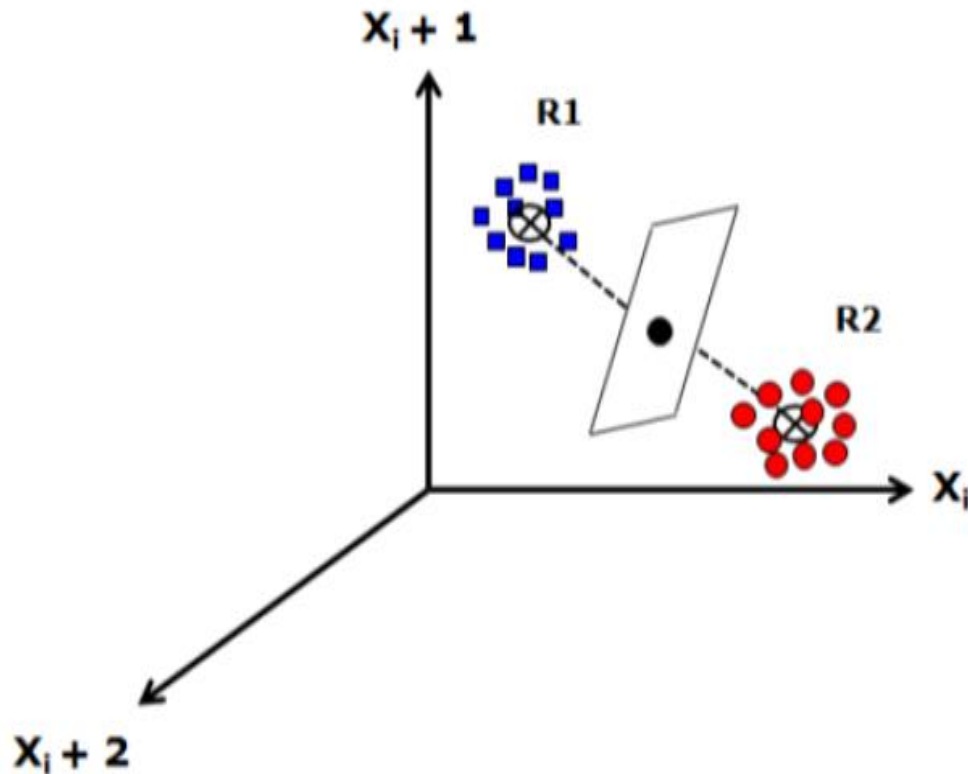


Figura 9. Representação esquemática da Análise Discriminante Linear.

Fonte: Adaptada de Gambarra-Neto (2008).

Ao contrário do SIMCA, método de classificação consolidado para trabalhar em altas dimensões (MELLINGER, 1987), a Análise Discriminante Linear (LDA) limita-se ao uso de dados com baixa dimensionalidade. Dessa forma, torna-se necessário a utilização de métodos de seleção de variáveis que são úteis para aumentar a capacidade preditiva dos modelos, eliminar variáveis não informativas e reter aquelas que, provavelmente, resultem em uma boa exatidão (BRUNS, FAIGLE, 1985).

2.5.2. Seleção de variáveis

As técnicas de seleção de variáveis (GESTAL *et al.*, 2004) envolvem a utilização de métodos computacionais cuja finalidade é encontrar um subconjunto de variáveis capazes de melhorar a eficiência dos resultados. Os métodos de seleção de variáveis buscam, ainda, produzir modelos mais simples ou parcimoniosos, por meio da remoção de variáveis não informativas e da minimização da multicolinearidade entre as variáveis (OLIVEIRA *et al.*, 2004; SOARES *et al.*, 2010; GOMES *et al.*, 2012).

Vários algoritmos têm sido propostos para seleção de variáveis (CANECA *et al.*, 2006), dentre os quais podemos destacar o Algoritmo das Projeções Sucessivas, inicialmente desenvolvido para seleção de variáveis no contexto de calibração multivariada (ARAÚJO *et al.*, 2001) e posteriormente adaptado para o contexto de reconhecimento de padrões (PONTES *et al.*, 2006).

2.5.2.1. Algoritmo das Projeções Sucessivas (SPA)

O SPA (SOARES, *et al.*, 2013), proposto inicialmente por Araújo *et al.* (2001), tem como objetivo buscar subconjuntos de variáveis que contenham informação minimamente redundante, minimizando a colinearidade dos dados. Para tanto, o SPA baseia-se num método de “inclusão” (método *forward*), que se inicia com uma variável e incorpora-se outra variável a cada iteração, até que a quantidade de N (número de variáveis necessárias) seja atingida. A essência do SPA consiste em operações de projeção efetuadas na matriz de dados, cujas linhas e colunas correspondem, respectivamente, às amostras de treinamento e variáveis (número de onda, potenciais, etc.). As cadeias de variáveis são geradas e avaliadas através do risco médio G (Equação 3) de uma classificação correta por LDA, calculada com base nas amostras do conjunto de validação. O subconjunto de variáveis que contém o menor valor de G será, então, selecionado (PONTES, *et al.*, 2005).

$$G = \frac{1}{K_v} \sum_{K=1}^{K_v} g_k \quad (3)$$

Na equação 3, K_v é o número de amostras do conjunto de validação e g_k representa o risco de uma classificação incorreta do objeto x_k da k -ésima amostra de validação, sendo definido como:

$$g_k = \frac{r^2(X_k, \mu_{ik})}{\min_{I_j \neq I_k} r^2(X_k, \mu_{I_j})} \quad (4)$$

Na equação 4, o numerador $r^2(X_k, \mu_{IK})$ corresponde ao quadrado da distância de Mahalanobis entre o objeto x_k (de índice de classe IK) e a média de sua classe (μ_{IK}). O denominador da equação é o quadrado da distância de Mahalanobis entre o objeto x_k e o centro da classe incorreta mais próxima. Dessa forma, o menor risco será obtido quanto mais distante o objeto estiver da classe incorreta e mais próximo estiver de sua classe verdadeira.

Para problemas de classificação com um número reduzido de amostras, onde apenas os conjuntos de treinamento e teste são utilizados, o cálculo da função de custo, J_{cost} , é realizado como mostrado na equação 5.

$$J_{cost} = \frac{1}{N_{train} - L - C} \sum_{n=1}^{N_{train}} g_n \quad (5)$$

onde L é o número de variáveis no subconjunto que está sendo avaliado e C representa o número de classes de objetos investigadas.

2.6. Aplicações de ferramentas quimiométricas e imagens digitais em problemas de classificação

As imagens digitais têm sido amplamente utilizadas em diversas áreas, inclusive na química analítica. As aplicações variam desde a classificação de refrigerantes, chás, méis, bactérias, controle de qualidade de óleos comestíveis, sementes de mamona e cafés (GODINHO *et al.*; 2008; DINIZ *et al.*, 2012; DOMINGUEZ *et al.*, 2014; MILANEZ *et al.*, 2014; ALMEIDA *et al.*, 2014; VILAR *et al.*, 2014; SOUTO *et al.*, 2015).

Godinho *et al.* (2008) classificaram refrigerantes de diferentes tipos utilizando análise de componentes principais (PCA) das imagens dos refrigerantes capturadas por um scanner. Além disso, realizou-se uma correlação entre os resultados de análises físico-químicas com os resultados da PCA. Neste estudo, a cor foi empregada como um indicativo de falsificação de refrigerantes e os histogramas dos canais RGB foram utilizados qualitativamente.

Diniz *et al.* (2012) fizeram uso de imagens obtidas por uma câmera digital para classificação de chás naturais não processados. A matriz de dados era formada por níveis de cores

das componentes de cores dos sistemas RGB, HSI e Escala de Cinza. Dois métodos quimiométricos foram comparados: SIMCA e LDA com seleção de variáveis pelo SPA. Os melhores resultados foram obtidos pelo SPA-LDA, onde todas as amostras foram corretamente classificadas de acordo com suas respectivas regiões geográficas.

Domínguez *et al.* (2014) classificaram origem geográficas de amostras de mel da Argentina utilizando imagens digitais obtidas com uma webcam. Os estudos foram levados a cabo usando histogramas de cores RGB, HSV e Escala de Cinza e as ferramentas quimiométricas SIMCA, PLS-DA e SPA-LDA.

Milanez *et al.* (2014) classificaram amostras de óleo vegetal de acordo com o tipo e estado de conservação utilizando imagens obtidas a partir de uma webcam e da distribuição de frequência dos índices de cor em RGB, HSI e Escala de Cinza. A técnica de LDA foi aplicada ao histograma a fim de construir modelos de classificação com base na redução de um subconjunto de variáveis.

Vilar *et al.* (2014) classificaram sementes de mamona com respeito ao tipo de cultivar e genótipo, utilizando imagens digitais a partir de uma webcam e histogramas de cor nos canais RGB, HSI e Escala de Cinza. PLS-DA e LDA foram empregadas como classificadores multivariados.

Almeida *et al.* (2014) propuseram uma metodologia para a diferenciação de cinco diferentes tipos de bactérias (*Escherichia coli*, *Enterococcus faecalis*, *Streptococcus salivarius*, *Streptococcus oralis* e *Staphylococcus aureus*) utilizando histogramas de cor obtidos a partir imagens digitais capturadas com uma webcam e SPA-LDA.

Souto *et al.* (2014) propuseram uma metodologia para identificar adulterações em amostras de cafés torrado (devido a presença de cascas e paus) utilizando imagens digitais e SPA-LDA, SIMCA e PLS-DA.

Percebe-se que as imagens digitais têm uma ampla aplicação e podem ser obtidas e tratadas de diferentes formas. É preciso enfatizar ainda que a totalidade dos trabalhos mencionados faz uso de ferramentas quimiométricas para tratamento das imagens digitais. Até o presente momento, não foram encontrados na literatura trabalhos voltados à aplicação de métodos de classificação multivariada a dados obtidos de imagens digitais para propósito de classificação de biodiesel (B100) e misturas biodiesel/diesel (B5) a partir do seu óleo de origem.

3. OBJETIVOS

3.1. Objetivo geral

Desenvolvimento de uma metodologia simples, rápida, de baixo custo e não destrutiva baseada na utilização de imagens digitais e técnicas de reconhecimento de padrões supervisionadas para a classificação de amostras de biodiesel (B100) e de misturas biodiesel/diesel (B5) em termos do óleo de origem.

3.2. Objetivos específicos

- ✓ Utilizar histogramas de cor (RGB, HSI e Escala de Cinza), gerados a partir das imagens digitais capturadas com uma webcam, como informação analítica.
- ✓ Utilizar técnicas de reconhecimento de padrões supervisionados (SIMCA, PLS-DA e SPA-LDA) para fins de classificação multivariada.
- ✓ Classificar amostras de biodiesel B100 e B5 a partir do óleo de origem utilizando histogramas de cor e técnicas de reconhecimento de padrões supervisionados.

4. MATERIAL E MÉTODOS

4.1. Localização do experimento

Os ensaios foram conduzidos no Laboratório de Química Analítica e Quimiometria (LQAQ) pertencente à Universidade Estadual da Paraíba (UEPB), Campus I, Campina Grande - PB.

4.2. Amostras de biodiesel

Para a obtenção do biodiesel foram utilizadas diferentes marcas e lotes de óleos refinados de algodão, girassol, milho e soja, perfazendo 120 amostras. Álcool metílico foi utilizado para promover a reação de transesterificação, empregando hidróxido de potássio como catalisador. Após a reação de transesterificação, as amostras de biodiesel foram lavadas com água destilada e solução de ácido clorídrico $0,1 \text{ mol L}^{-1}$, até alcançar pH próximo de 7,0 e continuaram até não ser perceptível a presença de resíduos, tais como excesso de álcool, glicerina, sais dos ácidos graxos e excesso do catalisador (LÔBO *et al.*, 2009). Todos os reagentes utilizados foram de grau analítico.

O processo de obtenção dos biodieseis de algodão, girassol, milho e soja por rota metílica seguiu as seguintes etapas:

Etapa 1: O volume de álcool metílico foi de 25,28 mL aferido com o auxílio de uma proveta, pesou-se 1 g da massa do catalisador hidróxido de potássio (KOH), transferiu-se para um balão de duas bocas acoplado a um sistema de refluxo com agitação constante para formação do metóxido de potássio, conforme ilustrado na Figura 10.



Figura 10. Sistema de refluxo e processo de formação do metóxido de potássio.

Fonte: Próprio autor.

Etapa 2: Pesou-se 100 g da massa de óleo e transferiu-se para o balão de duas bocas. O sistema ficou sob agitação durante uma hora (Figura 11).



Figura 11. Adição do óleo ao balão de duas bocas acoplado ao sistema de refluxo para reação de transesterificação.

Fonte: Próprio autor.

Etapa 3: Transferiu-se a mistura do balão de duas bocas para um funil de separação. Deixou-se em repouso por 24 horas para que ocorresse separação da glicerina, um dos produtos da reação, através da decantação, como pode ser observado na Figura 12.

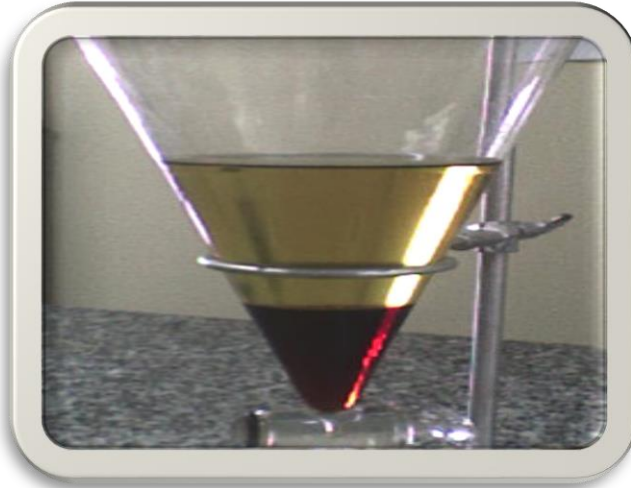


Figura 12. Separação da glicerina após 24 horas de repouso da mistura.

Fonte: Próprio autor

Etapa 4: Após a separação da glicerina, lavou-se o biodiesel com água destilada a 100°C até a remoção do sabão e alcançar um pH próximo de 7,0. Na Figura 13 é possível verificar o aspecto inicial e final da lavagem do biodiesel.

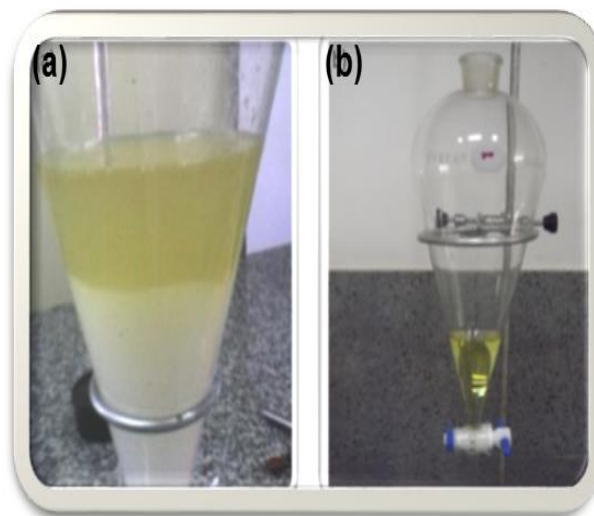


Figura 13. (a) Processo de purificação do biodiesel; (b) biodiesel purificado.

Fonte: Próprio autor

Etapa 5: Após a lavagem, colocaram-se as amostras de biodiesel na estufa. A amostra ficou a 100° C em estufa com circulação de ar por 3 horas. Deixou-se esfriar e armazenou-se a amostra (Figura 14).

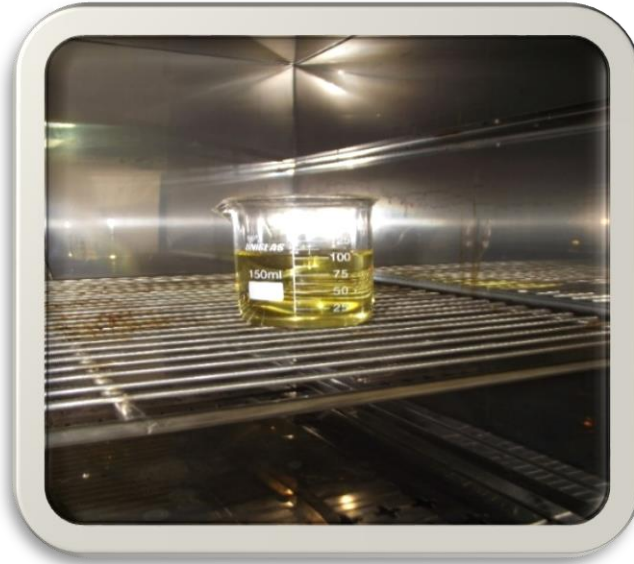


Figura 14. Processo de tratamento térmico do biodiesel.
Fonte: Próprio autor

Por fim, o biodiesel puro foi obtido separando a água por decantação e os traços de álcool foram eliminados através de um aquecimento na estufa a 100 °C durante 1 hora.

4.3. Amostras de misturas biodiesel/diesel (B5)

Foram preparadas 120 amostras de misturas biodiesel/diesel B5, sendo 30 amostras para cada classe (algodão, girassol, milho e soja). As misturas foram efetuadas na proporção de 5% v/v de biodiesel metílico para 95% v/v de diesel, denominada B5, conforme ilustrado na Figura 15.

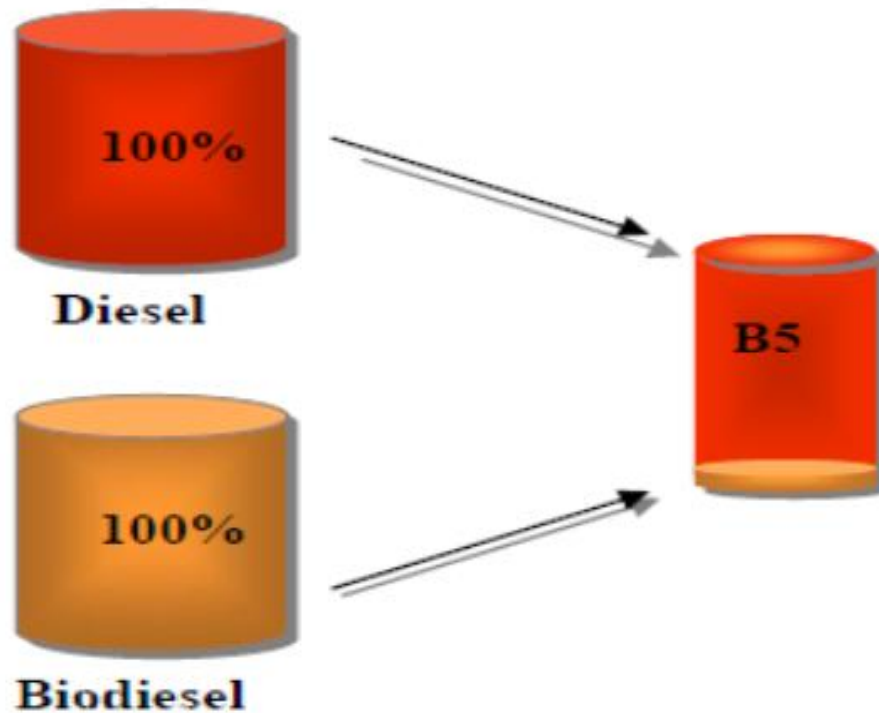


Figura 15. Esquema ilustrativo para mistura do biodiesel ao diesel.

Fonte: Próprio autor

4.4. Aparato

Para garantir a reprodutibilidade no processo de captura das imagens foi construído um compartimento de madeira de dimensões 34 cm x 21 cm x 23 cm com o intuito de isolar as amostras de biodiesel B100 e B5 da luz externa e, assim, facilitar o controle da luz incidente sobre as mesmas, garantindo a qualidade e, principalmente, a uniformidade das imagens registradas. Para esse fim, o compartimento foi revestido no seu interior com papel de cor branca, seguindo os princípios ópticos da reflexão da luz, para que todas as cores presentes na imagem fossem devidamente detectadas pela webcam. Dentro do compartimento foram colocadas uma placa de Petri que contém 5 mL da amostra e uma webcam modelo Microsoft Lifecam Cinema, com 7 megapixels de resolução, disposta estrategicamente para a aquisição das imagens. Uma lâmpada fluorescente comercial de cor branca de 8 W foi colocada fora do compartimento a aproximadamente 28 cm acima da amostra com o propósito de uniformizar a iluminação interna (Figura 16).

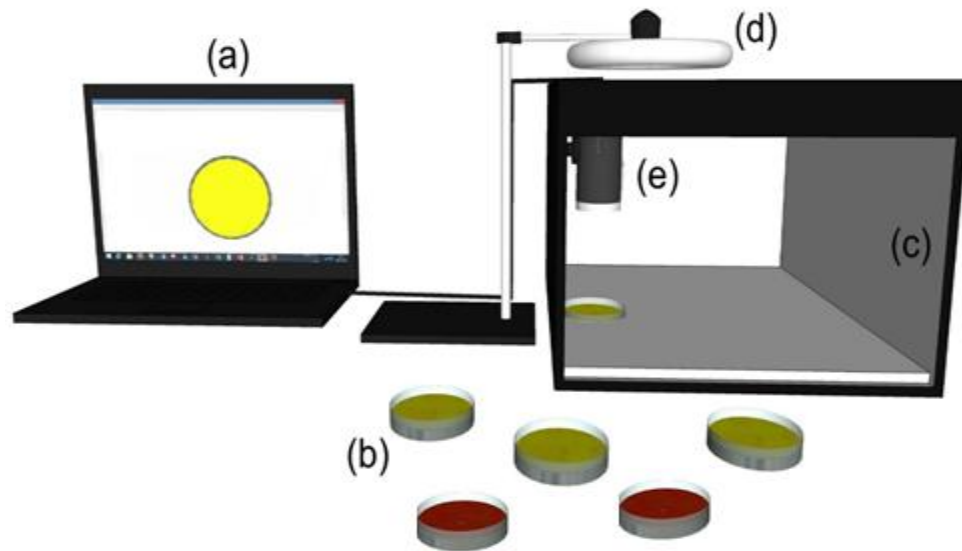


Figure 16. Sistema utilizado na aquisição das imagens das amostras de biodiesel B100 e B5. (a) notebook; (b) placa de Petri; (c) caixa; (d) lâmpada fluorescente e (e) webcam.

Fonte: Próprio autor

4.5. Aquisição das imagens

O método de aquisição de imagens leva em consideração as características visuais globais da superfície da amostra. Foram capturadas 5 imagens sequenciais para cada amostra e, em seguida, cada imagem foi decomposta em histogramas contendo a frequência de pixels correspondente a uma região específica da imagem, chamada de região de interesse (ROI, do inglês *region of interest*). Posteriormente, calculou-se a média destes histogramas que, por sua vez, foi utilizada como resposta instrumental referente àquela amostra. As imagens são compostas de 16,7 milhões de cores (24 bits) e 2880×1620 de resolução espacial, armazenados no formato JPEG (jpg).

A matriz de dados é formada por amostras localizadas em filas, enquanto que as colunas são constituídas de variáveis correspondentes aos níveis de cor obtidos para cada componente de cor (Figura 17).

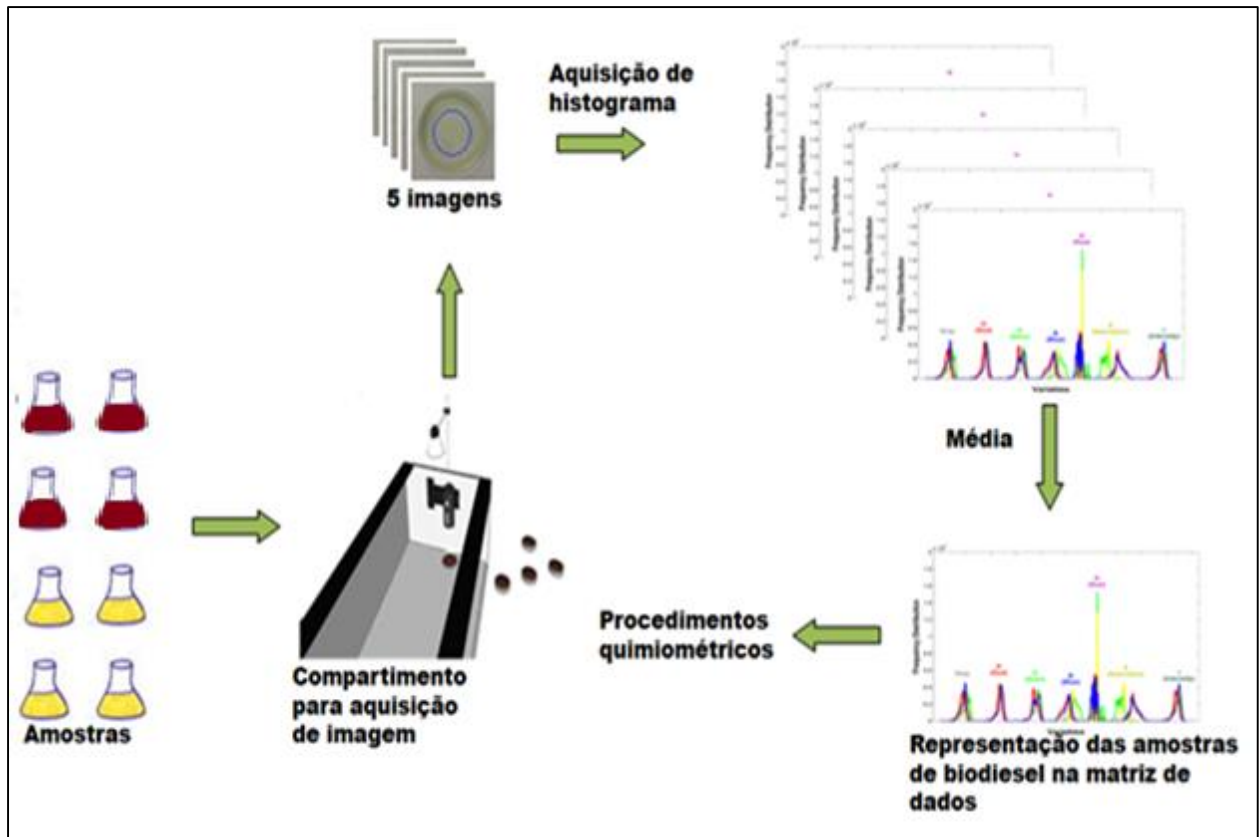


Figura 17. Representação de amostras na matriz de dados.

Fonte: Próprio autor

4.6. Histograma de cores e análise dos dados

Histogramas de cor foram obtidos a partir da captura de cada imagem digital usando uma rotina escrita em ambiente de Matlab R2011b.

Para o tratamento de dados a ROI foi definida por uma área elíptica a partir do centro de cada imagem, que representa aproximadamente 35% da área total da imagem. Usando apenas a região selecionada das imagens foram construídos histogramas de Escala de Cinza, Vermelho, Verde, Azul, Matiz, Saturação e Intensidade. Cada componente de RGB e Escala de Cinza varia de 0 a 255 (256 níveis); S e I de 0 a 1, e H varia de 0 a 360 graus. Estas componentes foram posteriormente utilizadas como informação analítica.

A informação analítica extraída dos histogramas supracitados foi utilizada para a construção de modelos quimiométricos de classificação utilizando SIMCA, PLS-DA e SPA-LDA. Foi avaliada a classificação das 120 amostras de biodiesel B100 e 120 amostras de biodiesel B5 em quatro diferentes classes: (a) algodão (b) girassol, (c) milho e (d) soja. Os dados obtidos a partir

de cada histograma foram separados em conjuntos de treinamento (80) e teste (40) usando o algoritmo Kennard-Stone (KS), que foi aplicado separadamente para cada classe.

PCA foi calculado usando o programa The Unscrambler 9.8 e o KS, SIMCA, PLS-DA, e SPA-LDA foram executados usando o programa Matlab 2011b.

5. RESULTADOS E DISCUSSÃO

5.1. Classificação de biodiesel B100 usando imagens digitais e quimiometria.

A Figura 18 mostra os histogramas médios de cada classe de amostra de biodiesel: algodão, girassol, milho e soja. Como pode ser visto, os histogramas médios exibiram uma notável tendência de separação das amostras de algodão e milho das amostras de soja e girassol nos canais Verde, Azul e Saturação. Com efeito, a cor vista é a cor complementar do comprimento de onda de absorção; portanto, os comprimentos de onda de verde e azul são absorvidos, enquanto os outros sofrem reflexão. Por estas razões, a cor amarelo avermelhada dos biodieseis é percebido, sendo relacionada com a presença de carotenoides nos óleos. Carotenoides ocorrem naturalmente nos óleos e gorduras e consistem principalmente de carotenos como o licopeno e xantofilas como a luteína (TCIS e ISEO, 2006). Além disso, uma combinação da intensidade da luz e quanto ele é distribuído em todo o espectro de comprimentos de onda diferentes determina a saturação de uma cor. Uma vez que os tons nos canais Verde e Azul mudam amplamente de acordo com o tipo de biodiesel, o canal de Saturação também varia fortemente, porque uma cor mais saturada (mais pura) seria obtida apenas pela utilização de um único comprimento de onda a uma alta intensidade. Para corroborar essas premissas, foi realizada uma análise exploratória de dados, como discutido na próxima seção.

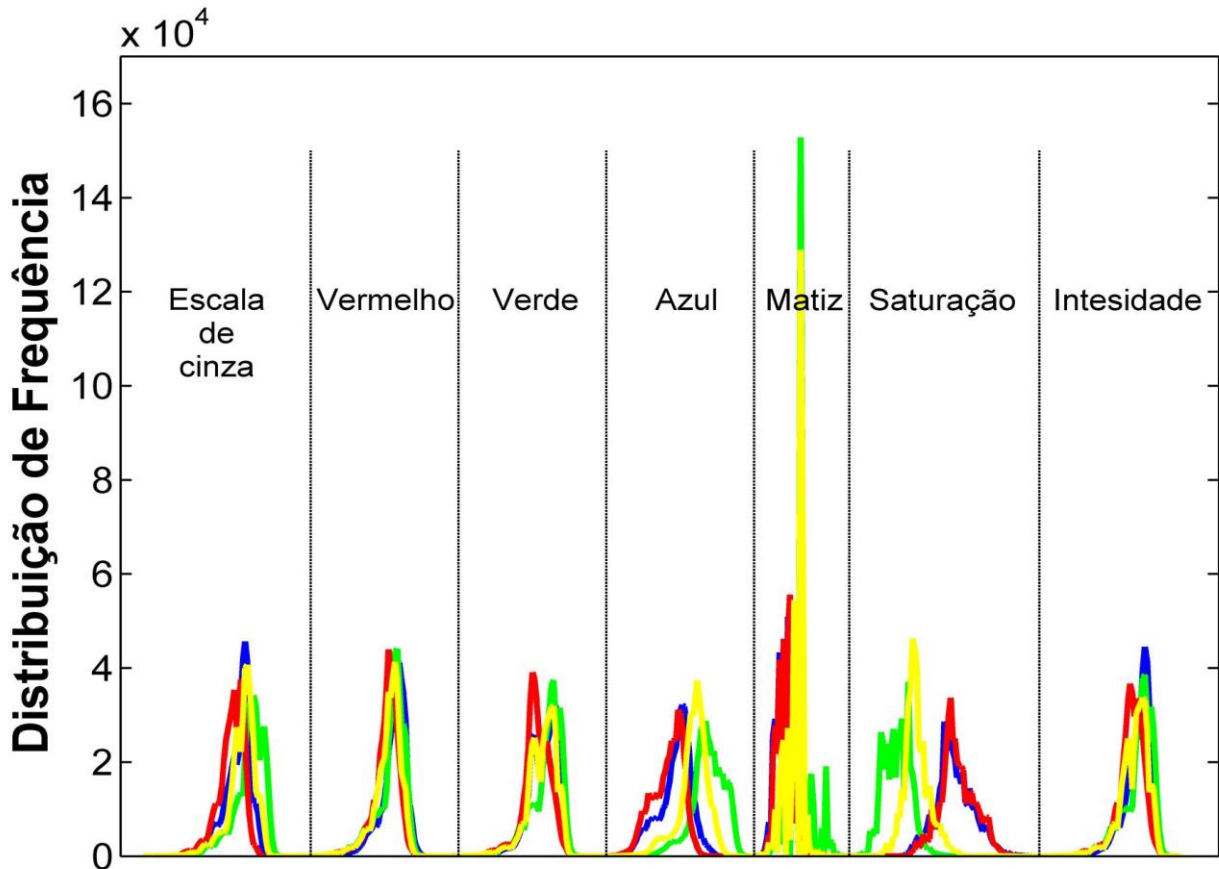


Figura 18. Histogramas médios nos canais de Escala de Cinza, Vermelho, Verde, Azul, Matiz, Saturação e Intesidade para cada classe de amostra biodiesel B100: algodão (linha azul), milho (linha vermelha), girassol (linha verde) e soja (linha amarela).

Fonte: Próprio autor.

5.1.1. Análise exploratória - PCA

Uma análise exploratória de dados foi realizada por meio da Análise de Componentes Principais para os histogramas de diferentes canais de cores no RGB, HSI, Escala de Cinza e suas combinações. Para ilustração, a Figura 19 mostra o PC1 x PC2 para as seguintes combinações: (a) Escala de Cinza + RGB + HSI, (b) Escala de Cinza + RGB, (c) Intesidade e (d) RGB. As outras abordagens estudadas apresentaram resultados muito semelhantes.

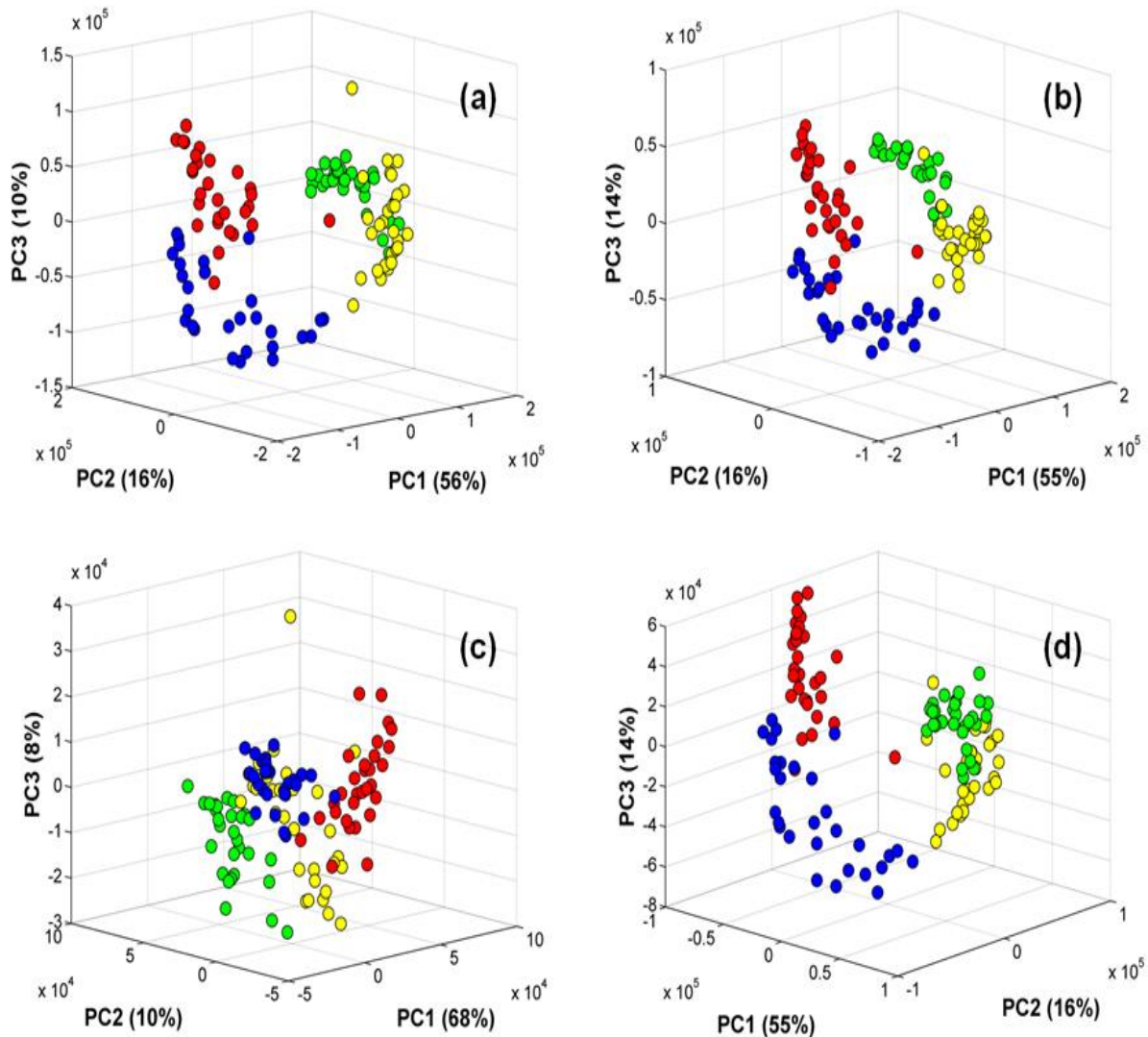


Figura 19. PCA gráficos dos escores PC1 x PC2, obtidos para as componentes de cores (a) RGB + HSI + Escala de Cinza (b) RGB e (c) Escala de Cinza + HSI para as quatro classes das amostras de biodiesel B100: (●) algodão (●) girassol, (●) milho, (●) soja.

Fonte: Próprio autor

Como pode ser visto, PCA forneceu resultados promissores quanto à discriminação das amostras de biodiesel claramente distinguidas em dois grandes grupos: um deles contendo amostras de algodão e milho e o outro com amostras de girassol e soja. Isto pode ser devido à composição de ácidos graxos de óleos vegetais utilizados como matéria-prima do biodiesel (Tabela 1). Embora a presença de carotenoides e outros pigmentos nas fontes do óleo, as amostras de biodiesel a partir de sementes de algodão, girassol, milho e soja apresentaram uma coloração amarelo claro. Esta redução na cor das amostras de biodiesel pode ser explicada pelas sucessivas lavagens das amostras com uma solução de ácido e água quente durante a etapa de purificação.

Em seguida, foram usadas técnicas de reconhecimento de padrões supervisionadas a fim de melhorar a discriminação entre as quatro classes diferentes. Para isso, modelos de classificação SIMCA, PLS-DA e SPA-LDA foram construídos com validação cruzada usando os canais individuais e suas combinações. As amostras de teste foram utilizadas apenas para a avaliação final dos dados e comparação dos modelos de classificação assinalando novos objetos (amostras desconhecidas).

Tabela 1- Composição de ácidos graxos (% m/m) de alguns óleos vegetais comuns usados como fonte de biodiesel^a.

Ácidos graxos	Algodão	Milho	Soja	Girassol
Mirístico (C14:0)	0.6 – 1.0	nd – 0.3	nd – 0.2	nd – 0.2
Palmítico (C16:0)	21.4 – 26.4	8.6 – 16.5	8.0 – 13.5	5.0 – 7.6
Palmitoleico (C16:1)	nd – 1.2	nd – 0.5	nd – 0.2	nd – 0.3
Estearico (C18:0)	2.1 – 3.3	nd – 3.3	2.0 – 5.4	2.7 – 6.5
Oleico (C18:1)	14.7 – 21.7	20.0 – 42.2	17.0 – 30.0	14.0 – 39.4
Linoleico (C18:2)	46.7 – 58.2	34.0 – 65.6	48.0 – 59.0	48.3 – 74.0
Linolênico (C18:3)	nd – 0.4	nd – 2.0	4.5 – 11.0	nd – 0.3
Araquídico (C20:0)	0.2 – 0.5	0.3 – 1.0	0.1 – 0.6	0.1 – 0.5
Beênico (C22:0)	nd – 0.6	nd – 0.5	nd – 0.7	0.3 – 1.5

^aDados obtidos da referência CODEX, 2001.

nd = não detectável, definido como $\leq 0,05\%$ m/m.

5.1.2. Classificação

A Tabela 2 mostra a precisão da classificação obtida para os diferentes histogramas de cor nos canais RGB, HSI, Escala de Cinza e suas combinações usando SIMCA, PLS-DA e SPA-LDA.

Tabela 2 - Taxa de classificação correta obtida com os modelos SIMCA, PLS-DA e SPA-LDA para a classificação de amostras de biodiesel B100.

Histogramas	Classificação correta (%)					
	SIMCA		PLS-DA		SPA-LDA	
	Treinamento	Teste	Treinamento ^a	Teste	Treinamento ^a	Teste
Vermelho	81,25	100	(9) 91,25	95,00	(13) 96,25	100
Verde	80,00	100	(14) 93,75	97,50	(15) 98,75	100
Azul	82,50	92,50	(9) 90,00	97,50	(13) 97,50	95,00
Matiz	71,25	90,00	(15) 82,50	85,00	(21) 98,75	90,00
Saturação	83,75	87,50	(6) 91,25	97,50	(15) 97,50	100
Intensidade	82,50	92,50	(5) 95,00	97,50	(12) 98,75	100
Escala de Cinza	92,50	100	(4) 92,50	100	(9) 97,50	100
RGB	93,75	95,00	(3) 92,50	100	(20) 100	100
HSI	93,75	95,00	(5) 93,75	95,00	(25) 98,75	95,00
Escala de Cinza + RGB	93,75	100	(6) 95,00	95,00	(7) 98,75	95,00
Escala de Cinza + HSI	93,75	95,00	(5) 93,50	95,00	(14) 97,50	97,50
RGB + HSI	93,75	95,00	(5) 93,75	95,00	(35) 100	97,50
Escala de Cinza+RGB+HSI	93,75	95,00	(5) 92,50	95,00	(16) 97,50	97,50

^a O número de variáveis ou variáveis latentes selecionado é indicado entre parêntesis.

Como pode ser visto na Tabela 2, várias abordagens apresentaram um bom desempenho dos modelos de classificação, atingindo mais de 90% de precisão de classificação em ambos os conjuntos de treinamento e teste. No entanto, considerando os conjuntos de treinamento e teste conjuntamente, os melhores resultados (ou seja, a maior precisão na classificação) foram obtidos para a modelagem SIMCA usando uma combinação de Escala de Cinza + RGB, para a modelagem PLS-DA usando apenas o canal Intensidade e para a modelagem SPA-LDA usando uma combinação de Escala de Cinza + HSI. Estes resultados são discutidos nas seções 5.1.2.1., 5.1.2.2 e 5.1.2.3.

5.1.2.1. Classificação SIMCA

O melhor resultado obtido para o SIMCA usou uma combinação de Escala de Cinza + RGB, atingindo uma taxa de precisão na classificação de 93,75 e 100% nos conjuntos de treinamento e teste, respectivamente. Para alcançar este resultado, o número ideal de componentes principais utilizado para cada classe foi: 1 PC tanto para algodão quanto para milho, 2 PCs para girassol e 4 PCs para soja, como pode ser visto na Tabela 3.

Tabela 3 – Matriz de confusão da classificação SIMCA para as amostras de biodiesel B100 usando o histograma de Escala de Cinza + RGB.

Conjunto de Treinamento						
Classe Verdadeira	N	SIMCA - Classe Prevista				TCC
		Algodão*(1)	Girassol*(2)	Milho*(1)	Soja*(4)	
1	20	19	-	1	-	93,75%
2	20	-	18	-	2	
3	20	1	-	19	-	
4	20	-	1	-	19	
Seletividade		100	90	85	95	
Especificidade		97	98	100	95	

Conjunto de Teste						
Classe Verdadeira	N	SIMCA - Classe Prevista				TCC
		Algodão*(1)	Girassol*(2)	Milho*(1)	Soja*(4)	
1	10	10	-	-	-	100%
2	10	-	10	-	-	
3	10	-	-	10	-	
4	10	-	-	-	10	
Seletividade		100	100	100	100	
Especificidade		100	100	100	100	

*Variáveis latentes; TCC = taxa de classificação correta; N = Número de amostras. O número ótimo de PCs para cada classe é indicado entre parênteses.

Os erros de classificação foram obtidos apenas no conjunto de treinamento: uma amostra de biodiesel de algodão foi classificada incorretamente como milho, uma amostra de biodiesel de milho como algodão, duas amostras de biodiesel de girassol como soja e uma amostra de biodiesel de soja como girassol. Uma vez que SIMCA baseia-se em PCA, os resultados estão próximos uns aos outros, com uma sobreposição entre amostras de biodiesel de algodão e milho e também entre as amostras de biodiesel de girassol e soja. Para ilustração, a Figura 20 mostra os escores de PC1 x PC2 de SIMCA em relação à classe de soja.

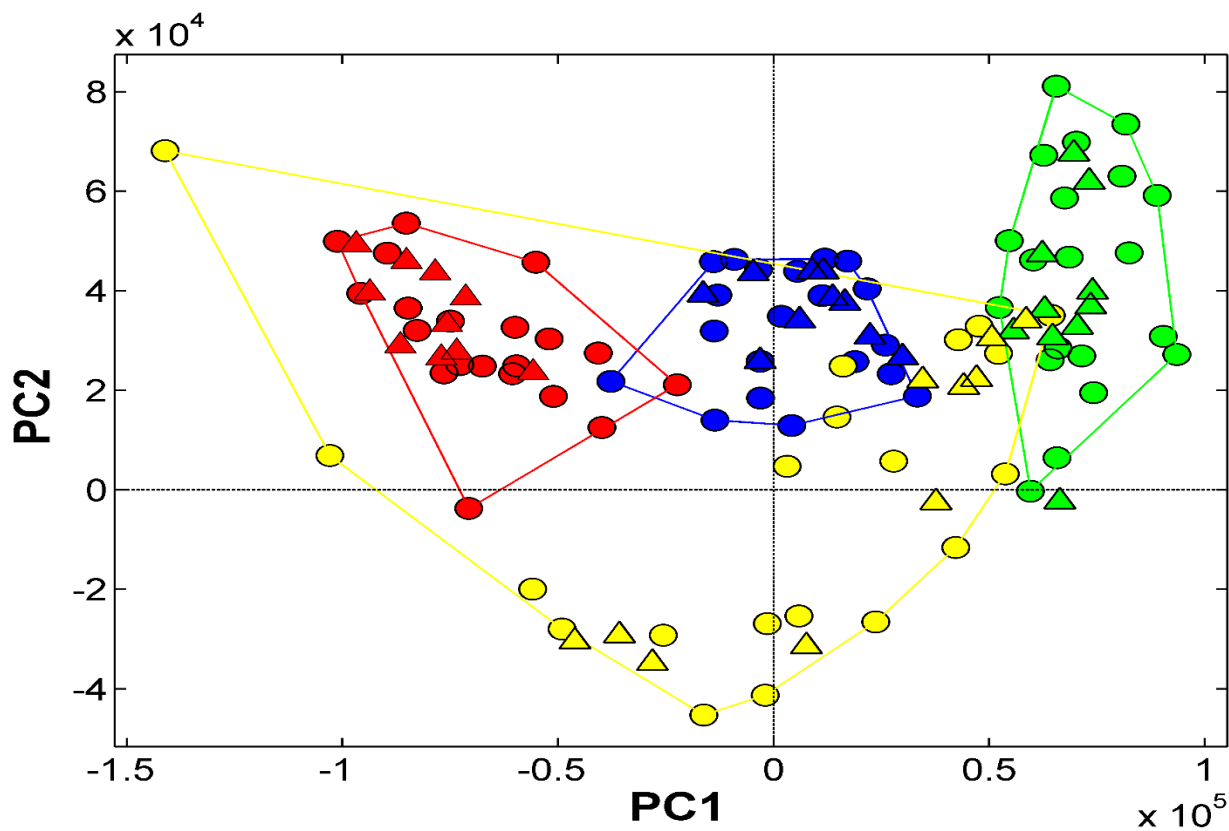


Figura 20. Gráfico de escores de PC1 x PC2 de SIMCA em relação à classe de soja. Amostras de treinamento (círculo) e teste (triângulo) para os biodieseis B100 de algodão (azul), milho (vermelho), girassol (verde) e soja (amarelo).

Fonte: Próprio autor

5.1.2.2. Classificação PLS-DA

A melhor abordagem PLS-DA foi obtida usando cinco variáveis latentes para o canal de Intensidade, chegando a 95,00 e 97,50% de precisão de classificação em ambos os conjuntos de

treinamento e teste, respectivamente (Tabela 4). Os limites de classe foram determinados por aplicação da curva ROC (do inglês *Receiver Operating Characteristics*), a fim de encontrar o ponto em que a sensibilidade e a especificidade têm o mesmo valor numérico (BALLABIO e CONSONNI, 2013).

Tabela 4 - Matriz de confusão da classificação PLS-DA para as amostras de biodiesel B100 usando o histograma de Intensidade.

		Conjunto de Treinamento				
Classe Verdadeira	N	PLS-DA - Classe Prevista *(5)				TCC
		Algodão	Girassol	Milho	Soja	
1	20	20	-	1	-	95,00%
2	20	-	19	-	1	
3	20	2	-	18	-	
4	20	-	1	-	19	
Seletividade		95	95	95	95	
Especificidade		98	98	98	98	

		Conjunto de Teste				
Classe Verdadeira	N	PLS-DA - Classe Prevista *(5)				TCC
		Algodão	Girassol	Milho	Soja	
1	10	10	-	-	-	97,50%
2	10	-	9	-	1	
3	10	-	-	10	-	
4	10	-	-	-	10	
Seletividade		100	90	100	100	
Especificidade		100	100	100	97	

*Variáveis latentes; TCC = taxa de classificação correta; N = Número de amostras.

De acordo com os resultados acima foram obtidos quatro e um erro de classificação para os conjuntos de treinamento e teste, respectivamente. No primeiro caso, uma amostra de biodiesel de girassol foi incorretamente classificada como soja, uma amostra de biodiesel de soja como girassol, e duas amostras de biodiesel de milho como algodão, enquanto que para o conjunto de teste apenas uma amostra de soja foi erroneamente classificada como girassol. Para ilustração, a Figura 21 mostra o gráfico de escores de PLS-DA utilizando o canal Intensidade. Aqui, dois grupos principais (algodão/milho e girassol/soja) também podem ser visualizados.

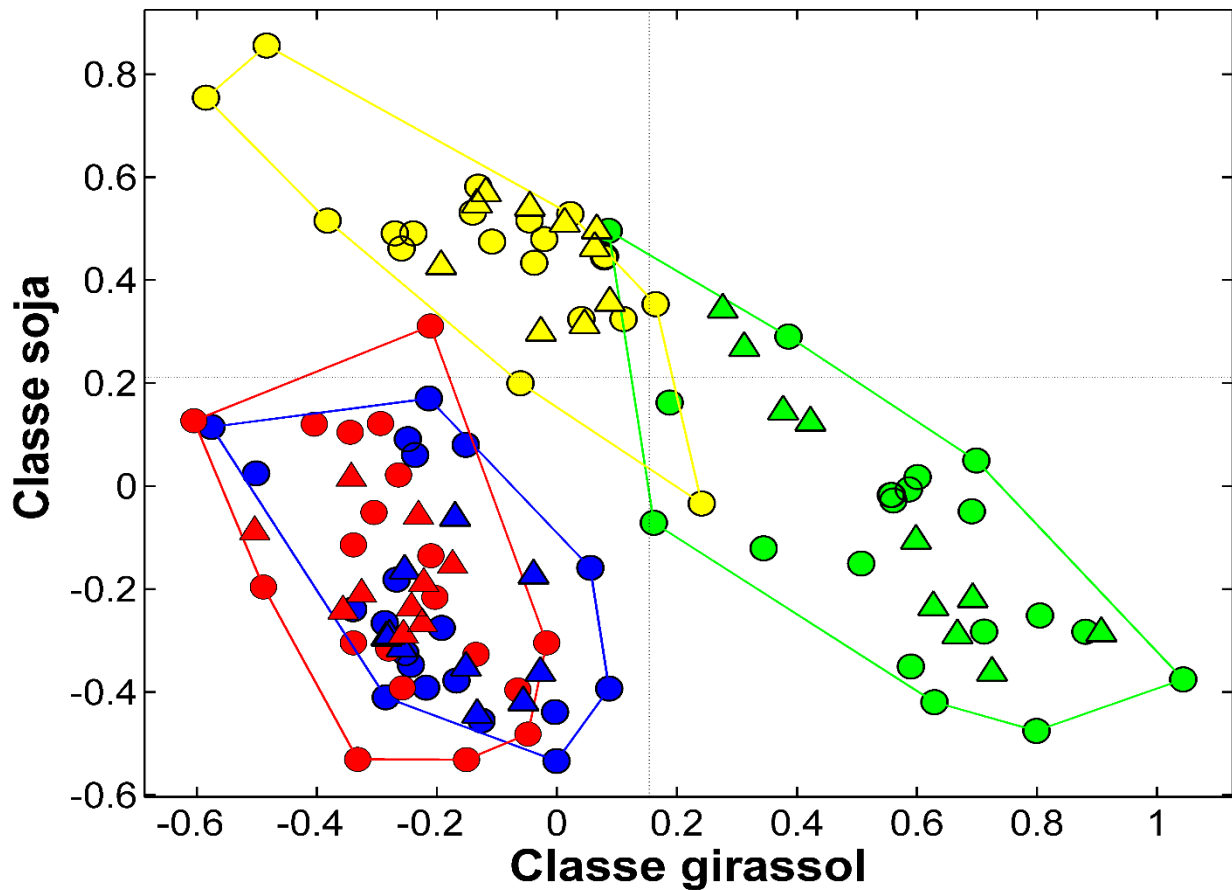


Figura 21. Gráfico de escores de PLS-DA utilizando o canal Intensidade. Amostras de treinamento (círculo) e teste (triângulo) para os biodieséis B100 de algodão (azul), milho (vermelho), girassol (verde) e soja (amarelo).

Fonte: Próprio autor

5.1.2.3. Classificação SPA-LDA

Tomando o melhor resultado, o menor risco de erros de classificação correspondente à função de custo G para a modelagem SPA-LDA foi obtido por seleção de vinte variáveis para o sistema de cor RGB. Nas Figuras 22 são mostradas (a) um histograma médio RGB com as vinte variáveis selecionadas pelo SPA e (b) um histograma completo evidenciando os canais utilizados.

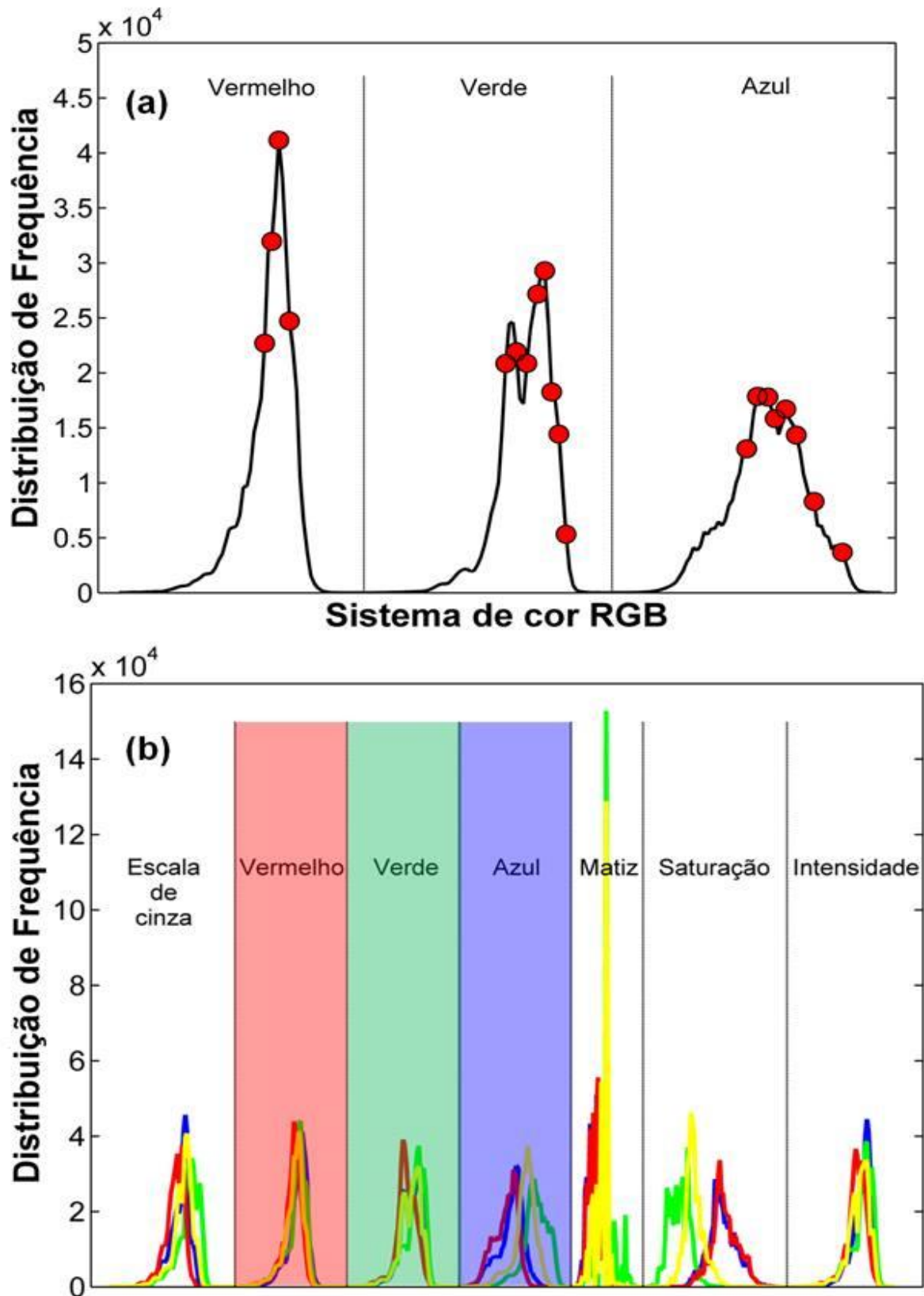


Figura 22. (a) As 20 variáveis selecionadas pelo SPA para o sistema RGB e (b) distribuição de frequências das cores predominantes para a classificação de amostras de biodiesel B100.

Fonte: Próprio autor.

Como pode ser observado, aproximadamente 80% das variáveis foram selecionadas nos canais verde e azul, porque a coloração amarela das amostras de biodiesel percebidos pelo olho humano é a cor complementar daqueles comprimentos de onda absorvidos. As demais variáveis selecionadas foram distribuídas no canal Vermelho e contribui para melhorar a discriminação entre as amostras de biodiesel B100. Para este caso, a taxa de classificação correta bem como a especificidade e a seletividade alcançaram 100% para ambos os conjuntos de treinamento e teste, respectivamente (Tabela 5).

Tabela 5 - Matriz de confusão da classificação SPA-LDA para as amostras de biodiesel B100 usando o histograma RGB.

Conjunto de Treinamento						
Classe Verdadeira	N	SPA-LDA - Classe Prevista				TCC
		Algodão	Girassol	Milho	Soja	
1	20	20	-	-	-	100%
2	20	-	20	-	-	
3	20	-	-	20	-	
4	20	-	-	-	20	
Seletividade		100	100	100	100	
Especificidade		100	100	100	100	

Conjunto de Teste						
Classe Verdadeira	N	PLS-DA - Classe Prevista				TCC
		Algodão	Girassol	Milho	Soja	
1	10	10	-	-	-	100%
2	10	-	10	-	-	
3	10	-	-	10	-	
4	10	-	-	-	10	
Seletividade		100	100	100	100	
Especificidade		100	100	100	100	

TCC = taxa de classificação correta; N = Número de amostras.

Na Figura 23 é observada a discriminação entre os agrupamentos de biodiesel de soja/girassol e as amostras de biodiesel de milho/algodão em DF1, enquanto DF2 discrimina os biodieseis de milho dos de algodão. Analogamente, as amostras de biodiesel de soja são discriminadas dos biodieseis de girassol em DF3.

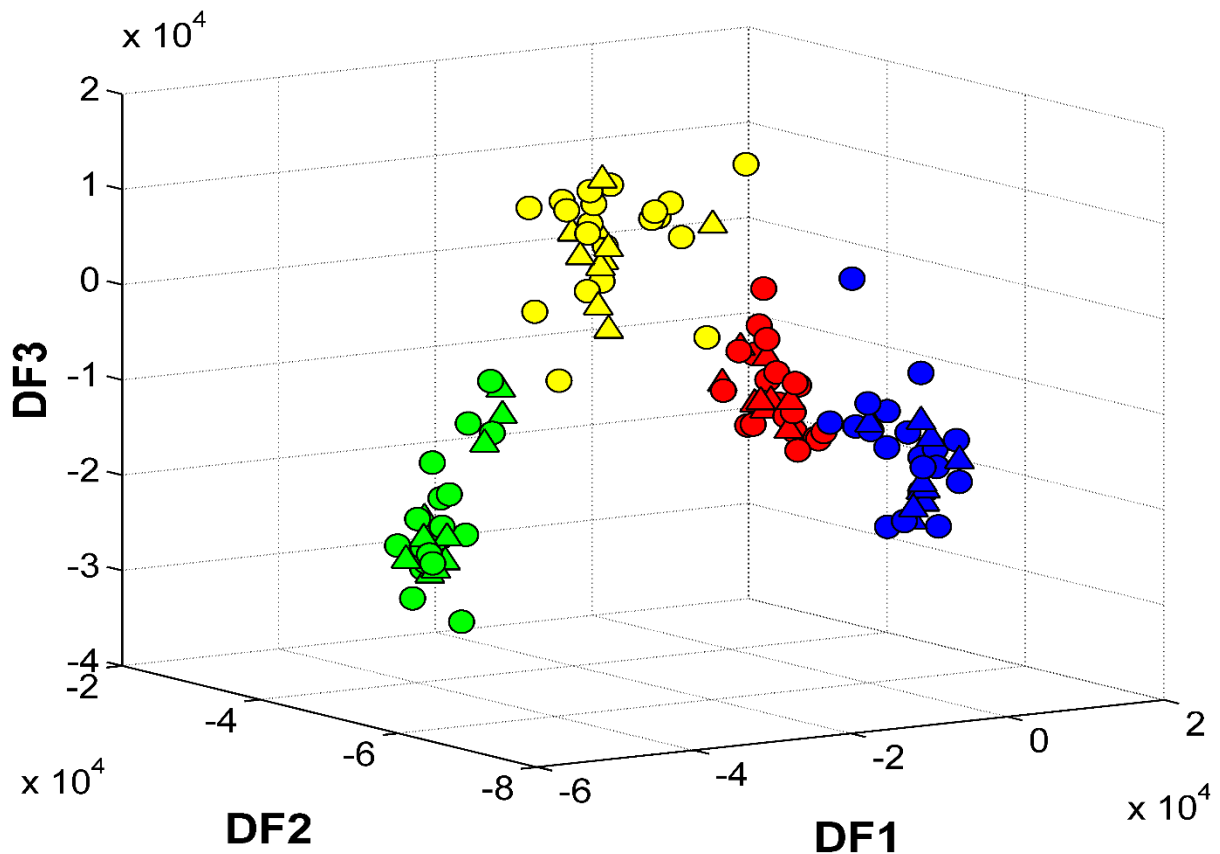


Figura 23. Funções Discriminantes obtidas pelo SPA-LDA na classificação das 120 amostras de biodiesel B100: (●) algodão (●) girassol (●) milho, (●) soja.

Fonte: Próprio autor.

5.2. Classificação de misturas biodiesel/diesel (B5) por imagem digital e quimiometria

A Figura 24 mostra os histogramas médios de cada classe de amostra de misturas biodiesel/diesel (B5): algodão, girassol, milho e soja. Como pode ser visto, os histogramas médios exibiram uma notável tendência de separação entre as amostras contendo biodiesel de algodão e milho daquelas que contém soja e girassol nos canais Verde, Azul e Saturação, conforme descrito na seção 5.1. Para corroborar essas premissas, foi realizada uma análise exploratória de dados, como discutido na próxima seção.

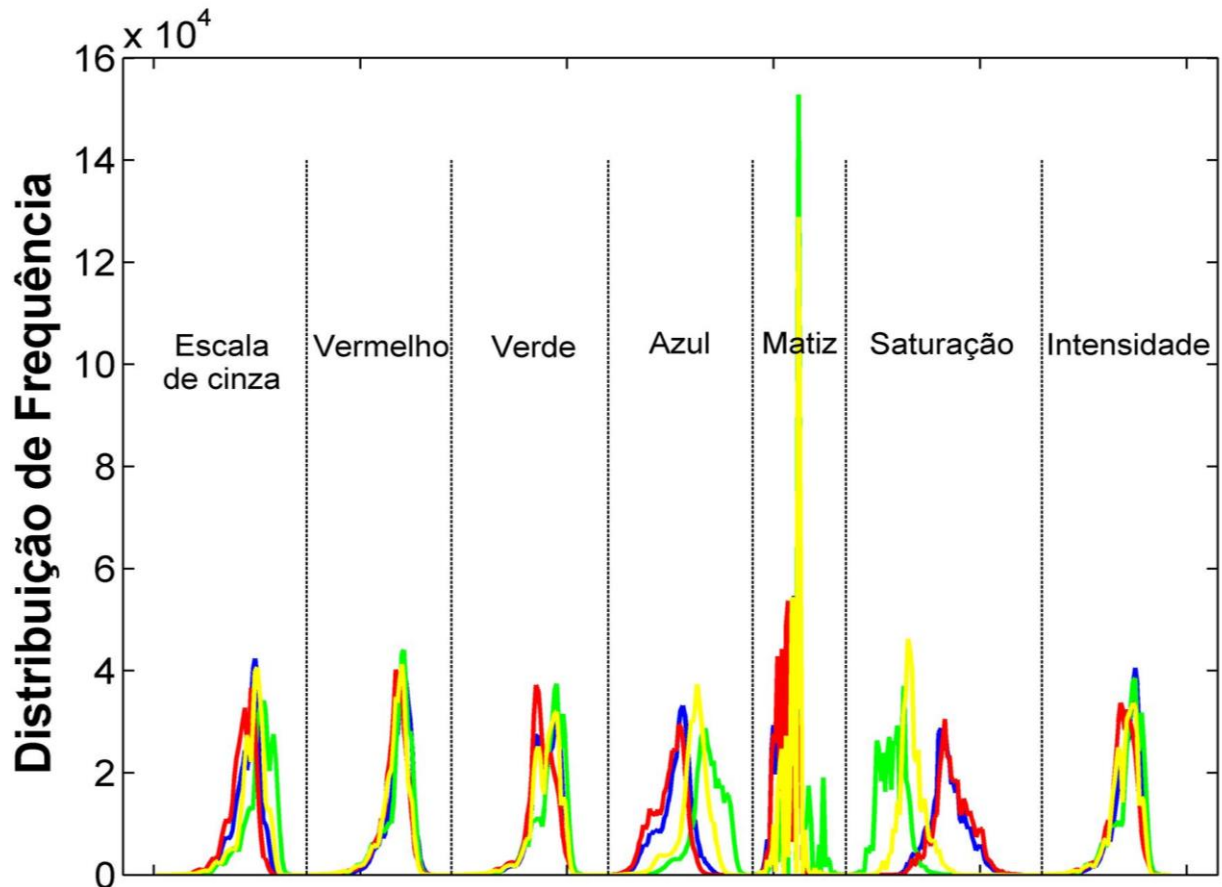


Figure 24. Histogramas médios nos canais de Escala de Cinza, Vermelho, Verde, Azul, Matiz, Saturação e Intensidade para cada classe de amostras de misturas biodiesel/diesel B5 após a remoção dos zeros: algodão (linha azul), milho (linha vermelha), girassol (linha verde) e soja (linha amarela).
Fonte: Próprio autor

5.2.1. Análise exploratória - PCA

Com o intuito de verificar o comportamento natural dos dados em termos de variância para as amostras de misturas biodiesel/diesel (B5) uma Análise de Componentes Principais foi aplicada analogamente ao descrito na seção 5.1.1. A Figura 25 apresenta os gráficos de escores das duas primeiras PCs para (a) Escala de Cinza + RGB + HSI, (b) RGB, (c) HSI e (d) Intensidade. As demais combinações apresentaram resultados similares.

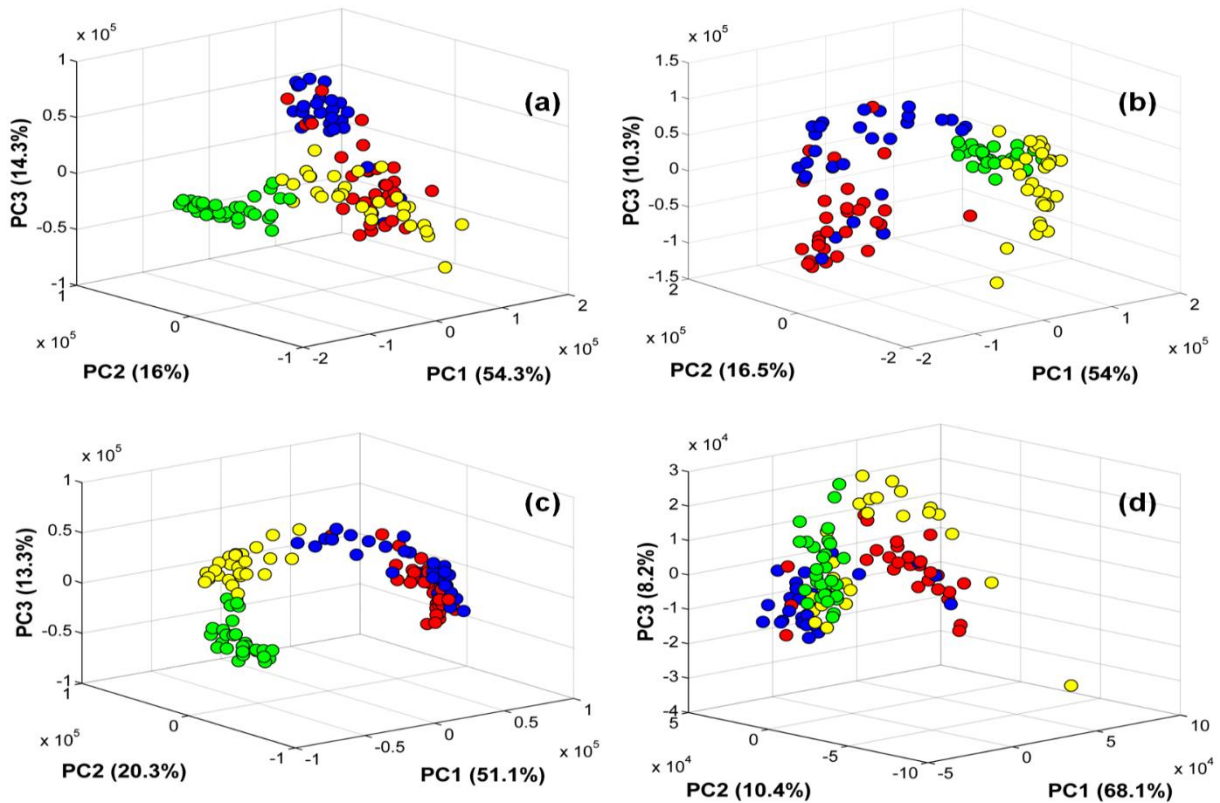


Figura 25. Gráficos dos escores PC1 x PC2 de PCA obtidos para (a) Escala de Cinza + RGB + HSI, (b) RGB, (c) HSI e (d) Intensidade para as quatro classes das amostras de misturas biodiesel/diesel B5: (●) algodão (●) girassol (●) milho, (●) soja.

Fonte: Próprio autor

A informação extraída na PCA do conjunto de dados forneceu resultados promissores quanto à discriminação das classes. Tais resultados sugerem uma investigação com base em técnicas de reconhecimento de padrões supervisionadas e seleção de variáveis com o intuito de melhorar a discriminação entre as diferentes amostras, bem como construir modelos para realizar previsões futuras. Para isso, foram construídos modelos de classificação SIMCA, PLS-DA e SPA-LDA utilizando as componentes de cores individuais e suas combinações.

5.2.2. Classificação

A Tabela 6 mostra a precisão da classificação das amostras de misturas biodiesel/diesel B5 obtida para os histogramas dos diferentes componentes de RGB, HSI e Escala de Cinza e suas combinações usando SIMCA, PLS-DA e SPA-LDA.

Tabela 6 - Taxa de classificação correta obtida com os modelos SIMCA, PLS-DA e SPA-LDA para a classificação das amostras de misturas biodiesel/diesel B5.

Histogramas	Classificação correta (%)					
	SIMCA		PLS-DA		SPA-LDA	
	Treinamento	Teste	Treinamento ^a	Teste	Treinamento ^a	Teste
Vermelho	60,00	87,50	(10) 80,00	90,00	(13) 90,00	92,50
Verde	71,25	95,00	(7) 83,75	90,00	(5) 82,50	92,50
Azul	76,25	90,00	(10) 76,25	90,00	(10) 90,00	90,00
Matiz	73,75	75,00	(13) 72,50	77,50	(4) 75,00	70,00
Saturação	76,25	80,00	(7) 78,75	95,00	(7) 83,75	90,00
Intensidade	78,75	95,00	(6) 85,00	95,00	(12) 87,50	97,50
Escala de Cinza	78,75	90,00	(4) 83,75	95,00	(8) 83,75	92,50
RGB	81,25	87,50	(3) 83,75	97,50	(5) 82,50	92,50
HIS	82,50	90,00	(4) 81,25	90,00	(3) 77,50	90,00
Escala de Cinza + RGB	78,75	95,00	(3) 81,25	97,50	(7) 82,50	92,50
Escala de Cinza + HSI	82,50	90,00	(3) 80,00	80,00	(4) 81,25	97,50
RGB + HSI	81,25	90,00	(3) 81,25	92,50	(3) 80,00	80,00
Escala de Cinza +RGB+HSI	85,00	90,00	(3) 82,50	90,00	(3) 81,25	92,50

^a O número de variáveis / variáveis latentes selecionados é indicada entre parêntesis.

5.2.2.1. Classificação SIMCA

Para o modelo SIMCA o melhor resultado foi obtido utilizando Escala de Cinza + RGB + HSI, atingindo uma taxa de classificação correta de 85 e 90% para os conjuntos de treinamento e teste, respectivamente. Para alcançar este resultado, o número ideal de componentes principais utilizadas para cada classe foi de 1 PC, como pode ser visto na Tabela 7.

Tabela 7 – Matriz de confusão da classificação SIMCA para as amostras de misturas B5 usando o histograma de Escala de Cinza + RGB + HSI.

		Conjunto de Treinamento				TCC
Classe	N	SIMCA - Classe Prevista				
Verdadeira		Algodão*(1)	Girassol*(1)	Milho*(1)	Soja*(1)	
1	20	12	-	8	-	85%
2	20	-	18	0	2	
3	20	5	-	15	-	
4	20	1	-	1	18	
Seletividade		60,0	90,0	75,0	90,0	
Especificidade		90,0	100	85,0	96,6	

		Conjunto de Teste				TCC
Classe	N	SIMCA - Classe Prevista				
Verdadeira		Algodão*(1)	Girassol*(1)	Milho*(1)	Soja*(1)	
1	10	8	-	2	-	90%
2	10	-	10	-	-	
3	10	-	-	10	-	
4	10	-	-	-	10	
Seletividade		80,0	100	100	100	
Especificidade		100	100	93,3	100	

* Variáveis latentes; TCC = taxa de classificação correta; N = Número de amostras.

De acordo com resultado mostrado na Tabela acima, os erros de classificação no conjunto de treinamento consistem de oito amostras de misturas B5 de algodão classificadas incorretamente como milho, duas amostras de misturas B5 de girassol como soja, cinco amostras de misturas B5 de milho como algodão, uma amostra de misturas B5 de soja como milho e uma amostra de misturas B5 de soja como algodão. Já para o conjunto de teste ocorreram apenas dois erros de classificação, sendo duas amostras de misturas B5 de algodão incorretamente classificadas na classe de milho.

5.2.2.2. Classificação PLS-DA

Para o modelo PLS-DA o número ótimo de variáveis latentes foi obtido usando três VLs para o sistema RGB. O limite das classes foi determinado aplicando a curva ROC a fim de encontrar o ponto onde sensibilidade e especificidade possuem o mesmo valor numérico (BALLABIO e CONSONNI, 2013). Os resultados detalhados para o modelo PLS-DA podem ser encontrados na Tabela 8.

Tabela 8 – Matriz de confusão da classificação PLS-DA para as amostras de misturas B5 usando o histograma de RGB.

Conjunto de Treinamento						
Classe		(3)* PLS-DA - Classe Prevista				TCC
Verdadeira	N	Algodão	Girassol	Milho	Soja	
1	20	15	-	4	1	83,75%
2	20	-	19	-	1	
3	20	5	-	14	1	
4	20	-	1	-	19	
Seletividade		75,0	95,0	70,0	95,0	
Especificidade		91,6	98,3	93,3	95,0	

Conjunto de Teste						
Classe		(3)* PLS-DA - Classe Prevista				TCC
Verdadeira	N	Algodão	Girassol	Milho	Soja	
1	10	9	-	1	-	97,50%
2	10	-	10	-	-	
3	10	-	-	10	-	
4	10	-	-	-	10	
Seletividade		90,0	100	100	100	
Especificidade		100	100	96,6	100	

* Variáveis latentes; TCC = taxa de classificação correta; N = Número de amostras.

Usando o sistema RGB, as taxas de classificação correta para os conjuntos de treinamento e teste foram 83,75 e 97,50 %, respectivamente. Observou-se que ocorreram treze erros de classificação no conjunto de treinamento e apenas um erro no conjunto de teste. No primeiro caso, cinco amostras de misturas B5 de algodão foram classificadas erroneamente como milho (quatro erros) e soja (um erro), uma amostra de misturas B5 de girassol como soja, cinco amostras de misturas B5 de milho como algodão, uma amostra de misturas B5 de milho como soja

e uma amostra de misturas B5 de soja como girassol. Já para o teste uma amostra de misturas B5 de algodão foi classificada erroneamente como milho. Para ilustração, a Figura 26 mostra o gráfico de escores PLS-DA utilizando o sistema de cores RGB. Aqui, dois grupos principais (algodão/milho e girassol/soja) também podem ser visualizados.

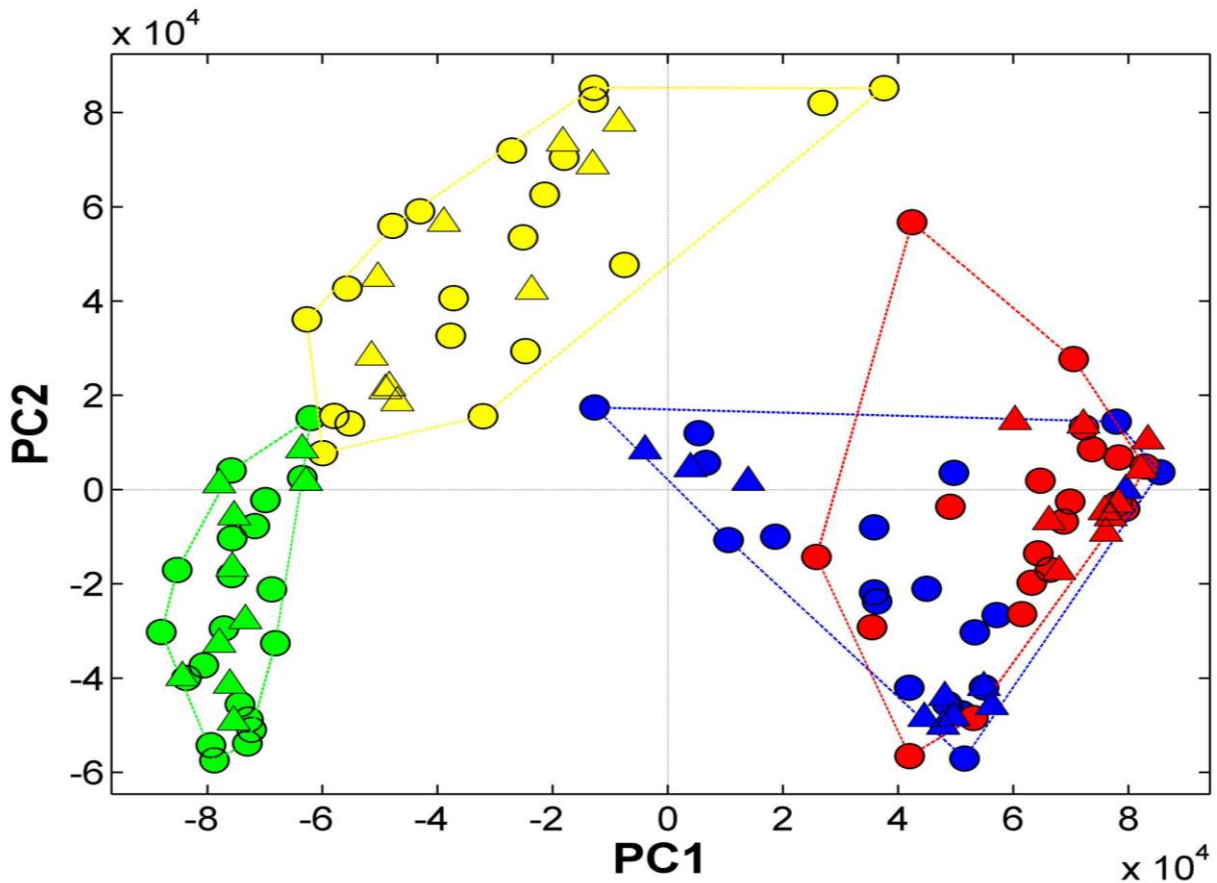


Figura 26. Gráfico de escores de PLS-DA utilizando o sistema de cores RGB. Amostras de treinamento (círculo) e teste (triângulo) para as amostras de misturas B5 de algodão (azul), milho (vermelho), girassol (verde) e soja (amarelo).

Fonte: Próprio autor.

5.2.2.3. Classificação SPA-LDA

Para a modelagem SPA-LDA o critério utilizado para seleção da cadeia de variáveis foi o gráfico de custo da validação, como resultado, a curva exibe um mínimo em doze variáveis. A Figura 27 é mostrado um histograma médio do canal de Intensidade com as doze variáveis selecionadas pelo SPA.

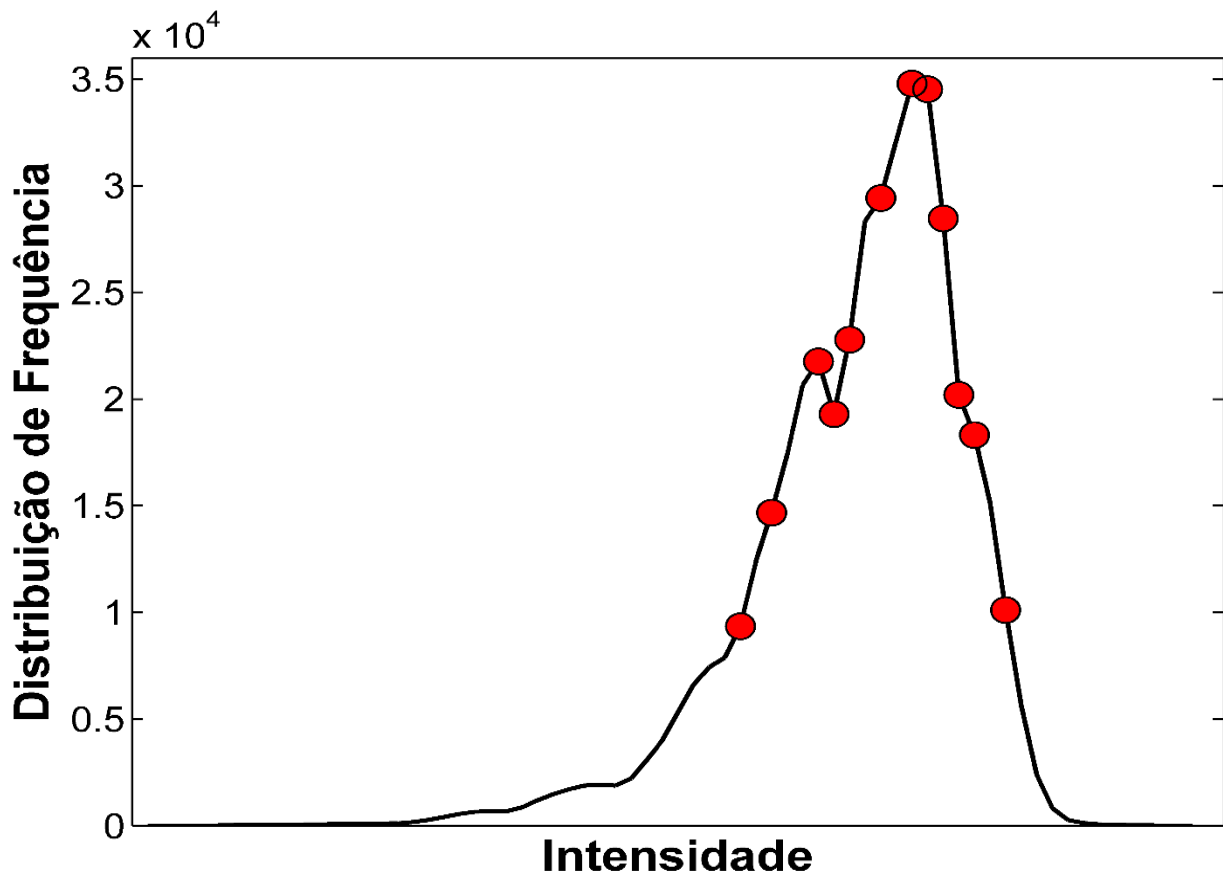


Figura 27. Histograma médio do canal de Intensidade com as doze variáveis selecionadas (●) pelo SPA.

Fonte: Próprio autor.

O modelo LDA obtido com as doze variáveis selecionadas por SPA foi utilizada para a classificação das amostras de misturas B5, apresentando 87,5 e 97,5% de classificação correta para os conjuntos de treinamento e teste, respectivamente. Para este caso, as especificidades alcançaram 91,6 e 100% e as seletividades de 70,0 e 100% valores mínimo e máximo para o conjunto de treinamento, respectivamente. Para o conjunto de teste as especificidades alcançaram 96,6 e 100% e as seletividades de 90,0 e 100% valores mínimo e máximo, respectivamente (Tabela 9).

Tabela 9 – Matriz de confusão da classificação SPA-LDA para as amostras de misturas B5 usando o histograma de I (intensidade).

Conjunto de Treinamento						
Classe		SPA-LDA - Classe Prevista				TCC
Verdadeira	N	Algodão	Girassol	Milho	Soja	
1	20	17	-	3	-	
2	20	-	19	-	1	
3	20	5	-	14	1	87,5%
4	20	-	-	-	20	
Seletividade		85,0	95,0	70,0	100	
Especificidade		91,6	100	95,0	96,6	

Conjunto de Teste						
Classe		SPA-LDA - Classe Prevista				TCC
Verdadeira	N	Algodão	Girassol	Milho	Soja	
1	10	9	-	1	-	
2	10	-	10	-	-	
3	10	-	-	10	-	97,5%
4	10	-	-	-	10	
Seletividade		90,0	100	100	100	
Especificidade		100	100	96,6	100	

TCC = taxa de classificação correta; N = Número de amostras.

O bom desempenho da classificação para o modelo SPA-LDA é demonstrado na Figura 28. Em DF1 pode ser observada a discriminação entre dois grandes grupos (soja/girassol e milho/algodão), enquanto DF2 apresenta uma sobreposição das amostras de misturas B5 de milho e algodão. As amostras de misturas B5 de soja são discriminadas das amostras de girassol em DF3.

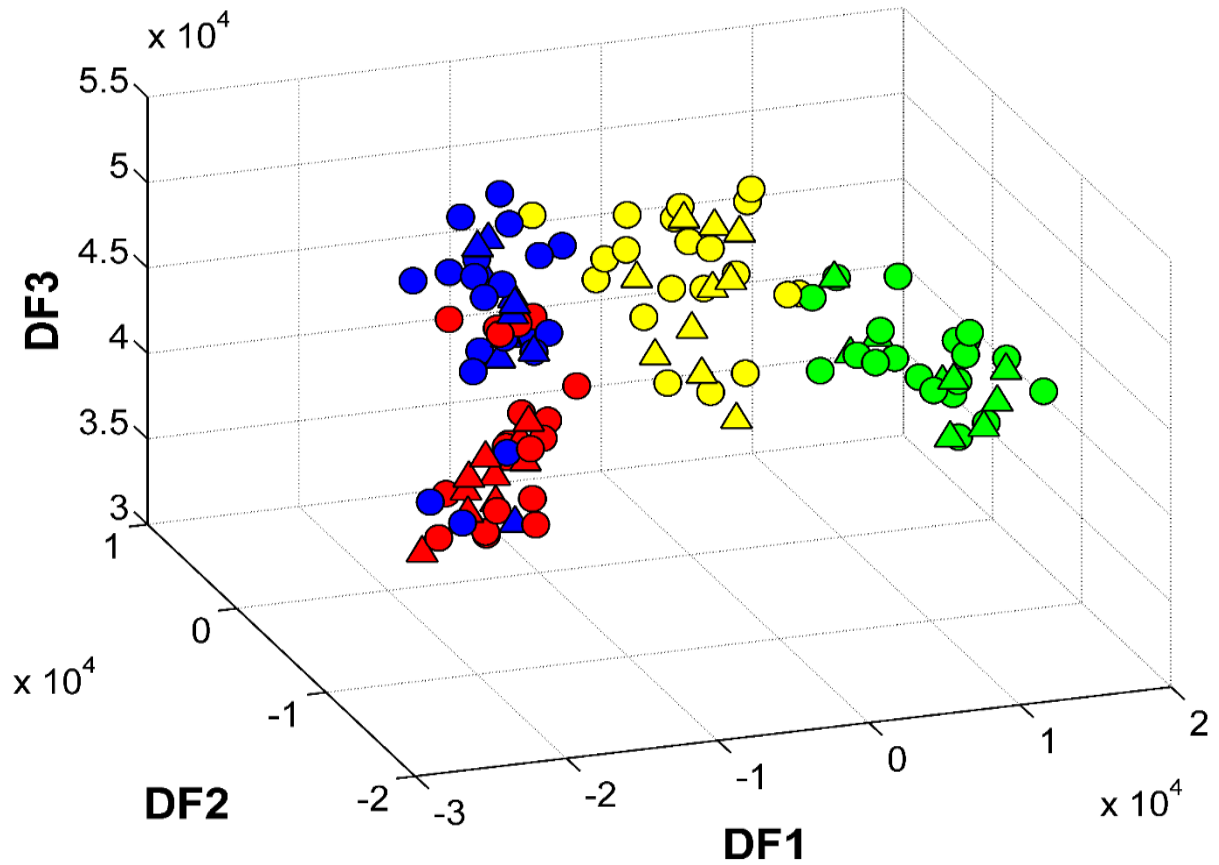


Figura 28. Funções Discriminantes obtidas pelo SPA-LDA na classificação das 120 amostras de misturas biodiesel/diesel B5: (●) algodão (●) girassol (●) milho, (●) soja.

Fonte: Próprio autor

6. CONCLUSÃO

Neste trabalho foi apresentada uma metodologia baseada na utilização de imagens digitais e técnicas de reconhecimento de padrões para classificação de biodiesel B100 e B5 a partir do seu óleo de origem.

Para este fim, um sistema portátil foi projetado para aquisição das imagens. Nesse sistema, foram considerados todos os elementos necessários para garantir o padrão de qualidade das imagens registradas que, satisfatoriamente, possibilitaram a extração de informações pertinentes a respeito das cores das amostras. Os histogramas obtidos foram analisados via SIMCA, PLS-DA e SPA-LDA.

A partir da análise empreendida ao longo do estudo foi possível perceber que as imagens digitais associadas às ferramentas quimiométricas empregadas obtiveram taxa de classificação correta elevada (acima de 90%) para a classificação de amostras de biodiesel B100 segundo o óleo de origem: algodão, girassol, milho e soja. O melhor resultado foi para o modelo SPA-LDA apresentou uma taxa de classificação correta, especificidade e seletividade de 100% para ambos os conjuntos de treinamento e teste com a utilização do sistema de cor RGB.

Quanto ao segundo problema de classificação de misturas biodiesel/diesel B5 a partir do óleo origem, o melhor resultado também foi com o modelo SPA-LDA, que apresentou uma taxa de classificação correta de 87,5 e 97,5% para os conjuntos de treinamento e teste, respectivamente, utilizando o canal Intensidade.

Apesar do bom desempenho dos modelos de classificação SIMCA e PLS-DA para os dois problemas de classificação abordados no presente estudo, os modelos SPA-LDA forneceram melhores resultados em termos de taxa de classificação correta (precisão), sensibilidade e especificidade para ambos os conjuntos de treinamento e teste, respectivamente, o que indica que as variáveis selecionadas pelo algoritmo contém informação adequada para classificação de biodiesel (B100) e suas misturas com diesel (B5) de acordo com o óleo de origem. Isto é importante porque o produto final pode ter diferentes propriedades dependendo da variedade da

matéria-prima utilizada, o que influencia diretamente na sua qualidade e, conseqüentemente, nos preços repassados ao consumidor. Além disso, outras vantagens intrínsecas são obtidas porque a análise é rápida, não requer uso de reagentes e evita caracterização química das amostras, isto é, a geração de resíduos é notadamente reduzida, contribuindo para um objetivo primário da Química Verde.

Em outras palavras, a característica mais importante para implementar o trabalho proposto como análise de rotina foi alcançada por causa de sua imediata disponibilidade a um baixo custo para o pesquisador, sendo útil para avaliação inicial e também fornecendo uma base de dados sobre fontes de biodiesel. Contudo, mais variedades de fontes de óleo devem ser implementadas a fim de generalizar a metodologia proposta.

REFERÊNCIAS

ANTONELLI, A.; COCCHI, M.; FAVA, P.; FOCA, G.; FRANCHINI, G.C.; MANZINI, D.; ULRICI, A. Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm. **Analytica Chimica Acta**, v. 515, n.1, p.3-13, 2004.

ALBUQUERQUE, M. P.; *Processamento de Imagens: Métodos e Análises*. **Centro Brasileiro de Pesquisas Físicas – CBPF/MCT-RJ**, 2002.

ALMEIDA, M. R.; CORREA, D. N.; ROCHA, W. F.C.; SCAFI, F. J.O.; POPPI, R. J. Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation. **Microchemical Journal**, v.109, n.1, p.170–177, 2013.

ALMEIDA, V.E.; COSTA, G.B.; FERNANDES, D.D.S.; DINIZ, P.H.G.D.; BRANDÃO, D.; MEDEIROS, A.C.D.; VÉRAS, G. Using color histograms and SPA-LDA to classify bacteria. **Analytical and Bioanalytical Chemistry**, v.406, n.24, p.5989-5995, 2014.

ANDERSEN, O.; WEINBACH J.E. Residual animal fat and fish for biodiesel production Potentials in Norway. **Biomass Bioenergy**, v.1, n.34, p. 1183–1188. 2010.

ARAUJO, M. C. U. The Successive Projections Algorithm for variable selection in spectroscopy multicomponent analysis. **Chemometrics and Intelligent Laboratory Systems**, v.57, n.1, p.65-67, 2001.

AZEVEDO, E.; CONCI, A. *Computação Gráfica: Teoria e Prática*. Rio de Janeiro: **Elsevier**, 2003.

BAKEAS, E.; KARAVALAKIS, G.; STOURNAS, S. Biodiesel emissions profile in modern diesel vehicles. Part 1: effect of biodiesel origin on the criteria emissions. **Science of the Total Environment**, v.409, n.1, p.1670–1676, 2011.

BALABIN, R. M.; SAFIEVA, R. Z.; LOMAKINA, E.I; Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. **Fuel**, v. 90, n.1, p. 2007–2015, 2011.

BALABIN, R.M.; SAFIEVA, R.Z. Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data, **Analytica Chimica Acta**, v.89, n.1, p. 190–197, 2011.

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

BATISTA, L.V. **Notas de aula da disciplina - Introdução ao Processamento Digital de Imagens**. 2004. Disponível em: <<http://www.di.ufpb.br/leonardo/pdi/PDI2005.pdf>>. Acesso em: 10 Jan. 2015.

BEEBE, K.R.; PELL, R.J.; SEASHOLTZ, M.B. Chemometrics: a practical guide. **Wiley-Interscience**, v.1, n.1, p.348-349, 1998.

BERGMANN, J.C.; TUPINAMBÁ, D.D.; COSTA, O.Y.A.; ALMEIDA, J.R.M.; BARRETO, C.C.; QUIRINO, B.F. Biodiesel production in Brazil and alternative biomass feedstocks. **Renewable and Sustainable Energy Reviews**, v.21, n.1, p.411-420, 2013.

BLANCO, M.; ALCALA, M. Simultaneous quantitation of five active principles in a pharmaceutical preparation: Development and validation of a near infrared spectroscopic method. **European Journal of Pharmaceutical Sciences**, v.27, n.1, p.280-284, 2006.

BRIONES, V.; AGUILERA, J.M. Image analysis of changes in surface color of chocolate. **Food Research International**, v.38, n.1, p.87–94, 2005.

BRUNS, R. E.; FAIGLE, J.F.G. Quimiometria. **Química Nova**, v.8, n.1, p.84-87, 1985.

CALAS, M.; CORDINA, G.; BOMPART, J.; BARI, M.B.; JEI, T.; ANCELIN, M.L.; VIAL H. Antimalarial Activity of Molecules Interfering with Plasmodium falciparum Phospholipid Metabolism. **Journal of Medicinal Chemistry**, v.40, n.1, p.3557-3566, 1997.

CANECA, A.R.; PIMENTEL, M.F.; GALVÃO, R.K.H.; MATTA, C.E.; CARVALHO, F.R.; RAIMUNDO JR, I.M.; PASQUINI, C.; ROHWEDDER, J.J.R. Assessment of infrared spectroscopy and multivariate techniques for monitoring the service condition of diesel-engine lubricating oils. **Talanta**, v.70, n.1, p.344-348, 2006.

CASTRO, A.A.M.; PRADO, P.P.L. Algoritmos para reconhecimento de padrões, **Revista Ciências Exatas**, v.5, n.2, 1999 – 2002, 2002.

CODEX ALIMENTARIUS COMMISSION, CODEX STANDARDS FOR NAMED VEGETABLE OILS, CODEX STAN 210-1999, FAO/WHO, **Roma**, 2001.

CORDEIRO, C. S.; SILVA, F. R.; WYPYCH, F.; RAMOS, L. P. Catalisadores heterogêneos para a produção de monoésteres graxos (Biodiesel). **Química Nova**, v.3, n.1, p. 477-486, 2011.

CORREIA, P. R. M.; FERREIRA, M. M. C. Reconhecimento de padrões por métodos não supervisionados: Explorando procedimentos quimiométricos para tratamento de dados analíticos. **Química Nova**, v.30, n.2, p.481-487, 2007.

COSTA, A. O.; OLIVEIRA, L. B.; LINS, M. P. E.; SILVA, A. C. M.; ARAUJO, M. S. M.; PERAIRA JR., A. O.; ROSA, L. P. Sustainability analysis of biodiesel production: A review on different resources in Brazil. **Renewable and Sustainable Energy Reviews**. v.27, n.1, p. 407-412, 2013.

COSTA NETO, P. R.; ROSSI, L. F. S.; ZAGONEL, G. F.; RAMOS L. P. Produção de Biocombustível Alternativo ao Óleo Diesel através da Transesterificação de Óleo de Soja Usado em Frituras. **Química Nova**, v.23, n.1, p.532-537, 2000.

DINIZ, P. H. G. D.; DANTAS, H. V.; MELO, K. D. T.; BARBOSA, M. F.; HARDING, D. P.; NASCIMENTO, E. C. L.; PISTONESI, M. F.; BAND, B. S. F.; ARAÚJO, M. C.U. Using a simple digital camera and SPA-LDA modeling to screenteas. **Analytical Methods**, v.4, n.1, p. 2648–2652, 2013.

DOMÍNGUES, M.A.; DINIZ, P.H.G.D.; DI NEZIO, M.S.; ARAUJO, M.C.U.; CENTURIÓN, M.E. Geographical origin classification of Argentinean honeys using a digital image-based flow-batch system. **Microchemical Journal**, v.112, n.1, p.104–108, 2014.

DU, C.J.; SUN, D.W. Recent developments in the applications of image processing techniques for food quality evaluation. **Trends in Food Science & Technology**, v.15, n.1, p. 230-249, 2004

EL-HAMDY, A.H.; EL-FIZGA, N.K. Detection of olive oil adulteration by measuring its authenticity factor using reversed-phase high-performance liquid chromatography. **Journal of Chromatography A**, v.708, n.1. p.351-355, 1995.

ESBENSEN, K.H.; GELADI, P. Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice Comprehensive Chemometrics, Editors-in-Chief: Stephen D. Brown, Romà Tauler, and Beata Walczak. **Oxford**, v.13, n.2, p.211–226, 2009.

FERREIRA, M.C. Multivariate QSAR. **Journal Brazilian Chemical Society**, v.13, n.6, p.742-753, 2002.

FERREIRA, M.M.C.; ANTUNES, A.M.; MELGO, M.S.; VOLPE, P.L.O. Quimiometria I: Calibração multivariada, um tutorial. **Química Nova**, v.22, n.5, p. 724-31, 1999.

FORE S.R.; LAZARUS, W.; PORTER, P.; JORDAN, N. Economics of small-scale on-farm use of canola and soybean for biodiesel and straight vegetable oil biofuels. **Biomass and Bioenergy**, v.35, n.1, p.193–202, 2011.

FRANCESCHINI, G; MACCHIETTO, S. Anti-Correlation Approach to Model-Based Experiment Design: Application to a Biodiesel Production Process. **Industrial & Engineering Chemistry Research**, v.47, n.1, p. 2331-2348, 2008.

GAMBARRA-NETO, F.F.; MARINO, G.; ARAUJO, M.C.U.; GALVÃO, R.K.H.; PONTES, M.J.C. Classification of edible vegetable oils square wave voltammetry with multivariate data analysis. **Talanta**, v.77, n.1, p. 1660-1666, 2009.

GARRIDO-NOVELL, C.; PEREZ-MARIN, D., AMIGO, J. M.; FERNANDEZ-NOVALES, J.; GUERRERO, J. E.; GARRIDO, A.V. 2012. Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras. **Journal of Food Engineering**, v.113, n.2, p.281-288, 2012.

GERIS, R.; SANTOS, N. A. C.; AMARAL, B. A.; MAIA, I. S.; CASTRO, V. D; CARVALHO, J.R.M. Biodiesel de soja: Reação de transesterificação para aulas práticas de química orgânica. **Química Nova**, v.30, n.3, p. 1369-1373, 2007.

GESTAL, M. Classification of apple beverages using artificial neural networks with previous variable selection. **Analytica Chimica Acta**, v.524, n.1, p. 225-228, 2004.

GODINHO, M.S.; PEREIRA, R.O.; RIBEIRO, K.O., SCHIMIDT, F.; OLIVEIRA, A.E.; OLIVEIRA, S.B. Classificação de refrigerantes através de análise de imagens e análise de componentes principais (PCA), **Química Nova**, v.31, n.1, p.1485-1489, 2008.

GOMES, A. A. **Algoritmo das Projeções Sucessivas aplicado à seleção de variáveis em regressão PLS**. 2012. 121f. Dissertação (Mestrado em Química) – Universidade Federal da Paraíba.

GONZALEZ, R.C.; WOODS, R.E. **Processamento de imagens digitais**. Tradução Roberto Marcondes Cesar Júnior, Luciano da Fontoura Costa. 1ª Ed. São Paulo: Editora Blucher, 2000.

HAIR JUNIOR, J.F.; ANDERSON, R.E; TATHAM, R. L.; BLACK, W.C. Multivariate data analysis with readings. **Prentice-Hall**, v.1, n.1, p.197-200, 1995.

INSAUSTI, M.; GOMES, A.A.; CRUZ, F.V.; PISTONESI, M.F.; ARAUJO, M.C.U.; GALVÃO, R.K.H.; PEREIRA, C.F.; BAND, B.S.F. Screening analysis of biodiesel feedstock using UV–vis, NIR and synchronous fluorescence spectrometries and the successive projections algorithm **Talanta**, v.97, n.1, p.579-583, 2012.

KNOTHE, G. Analyzing biodiesel: standards and other methods. **Journal of the American Oil Chemists' Society**, v.83, n.1, p.823–833, 2006.

KNOTHE, G. Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. **Fuel Processing Technology**, v.86, n.1, p.1059-1070, 2005.

KNOTHE, G. Rapid monitoring of transesterification and assessing biodiesel fuel quality by near-infrared spectroscopy using a fiber-optic probe. **Journal of the American Oil Chemists' Society**, v.76, n.1, p.795–800, 1999.

KONG, F.; TAN, J. DietCam: Automatic dietary assessment with mobile camera phones. **Pervasive and Mobile Computing**, v.8, n.1, p.147-163, 2012.

LEE, D.-S.; NOH, B.-S.; BAE, S.-Y.; KIM, K. Characterization of fatty acids composition in vegetable oils by gas chromatography and chemometrics **Analytica Chimica Acta**, v.358, n.2, p.163 - 175, 1998.

LÔBO, I. P.; FERREIRA, S. L. C.; CRUZ, R. S. Biodiesel: parâmetros de qualidade e métodos analíticos. **Química Nova**, v.32, n.1, p.1596-1608, 2009.

MACHADO, A. M. L. **Extração automática de contornos de edificações utilizando imagem gerada por câmara digital de pequeno formato e dados lidar**. 2006. 213f. Tese (Doutorado em Ciências Geodésicas) –Universidade Federal do Paraná.

MACHADO, A. W.; LEITE, E. B.; SOUKI, B. Q. Fotografia Digital em Ortodontia: Parte II – Sistema Digital X Sistema Analógico. **Jornal Brasileiro de Ortodontia & Ortopedia Facial**, v. 9, n. 50, p. 146-153, 2004.

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

MARTENS H.; TORMOD N. **Multivariate Calibration**, John Wiley & Sons, New York, 1996

MASSART, D.L.; VANDEGINSTE, B.G.M.; BUYDENS, S.J.; LEWI, P.J.; SMEYERS-VERBEKE, J. Handbook of Chemometrics and Qualimetrics: *Parte B*, Elsevier, 1997.

MILANEZ, K. D. T. M.; PONTES, M. J. C. Classification of edible vegetable oil using digital image and pattern recognition techniques. **Microchemical Journal**, v.113, n.1, -10-16, 2014.

OLIVEIRA, F. C.; SOUZA, A. T. P. C.; DIAS, J. A.; DIAS, S. C. L.; RUBIM, J. C. A. Escolha da faixa espectral no uso combinado de métodos espectroscópicos e quimiométricos. **Química Nova**, v.27, n.1, p.218-225, 2004.

PIMENTEL, M. F.; SALDANHA, T. C. B.; ARAÚJO, M. C. U. Effects of experimental design on calibration curve precision in routine analysis, **Journal of Automatic Chemistry**, v.20, n.1, p.9-11, 1998.

PONTES, M.J.C.; SANTOS, S.R.B.; ARAÚJO, M.C.U.; ALMEIDA, L.F.; LIMA, R.A.C.; GAIÃO, E.N.; SOUTO, U.T.C.P. Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectroscopy. **Food Research International**, v.39, n.1, p.182-189, 2006.

PONTES, M.J.C.; GALVÃO, R.K.H.; ARAUJO, M.C.U.; MOREIRA, P.N.T.; PESSOA NETO, O.D.; JOSÉ, G.E.; SALDANHA, T.C.B. The successive projections algorithm for spectral variables election in classification problems. **Chemometrics and Intelligent Laboratory Systems**, v.78, n.1, p.11-18, 2005.

PONTES, M.J.C.; PEREIRA, C.F.; PIMENTEL, M.F.; VASCONCELOS, F.V.C.; SILVA, A.G.B. Screening analysis to detect adulteration in diesel/biodiesel blends using near infrared spectrometry and multivariate classification. **Talanta**, v.85, n.1, p.2159-2165. 2011.

QUEIROZ, J.E.R.; GOMES, H.M. **Introdução ao Processamento Digital de Imagens**, Revista RITA: instruções para preparação de documentos em Word, 8 (1): 1, 2001.

SABIN, J.G.; FERRÃO, M.F.; FURTADO, J.C. Análise multivariada aplicada na identificação de fármacos antidepressivos. Parte II: Análise por componentes principais (PCA) e o método de classificação SIMCA. **Revista Brasileira de Ciências Farmacêuticas**, v.40, n.3, p.387-96, 2004.

SACCHI, R.; PATUMI, M.; FONTANAZZA, G.; BARONE, P.; FIORDIPONTI, P.; MANNINA, L.; ROSSI, E.; SEGRE, A.L. A High-Field ¹H Nuclear Magnetic Resonance Study of the Minor Components in Virgin Olive. **Journal of the American Oil Chemists' Society**, v.73, n.6, p 747-758, 1996.

SANTOS, P.M.; PEREIRA-FILHO, E.R. Digital image analysis - an alternative tool for monitoring milk authenticity. **Analytical Methods**, v.5, p. 3669-3674, 2013.

SENGO, I.; GOMINHO, J.; D'OREY, L.; MARTINS, M.; DUARTE, E. A.; PEREIRA, H.; DIAS, S. F. Response surface modeling and optimization of biodiesel production from *Cynara cardunculus* oil. **European Journal of Lipid Science and Technology**, v.112, n.3, p.310–320, 2010.

SILVA, A.C.; PONTES, L.F.B.L.; PIMENTEL, M.F.; PONTES, M.J.C. Detection of adulteration in hydrated ethyl alcohol fuel using infrared spectroscopy and supervised pattern recognition methods. **Talanta**, v.93, n.1, p.129-134, 2012.

SILVA, G.W.B; GOMES, A.A; SILVA, P; COSTA, G.B; FERNANDES, D.D.S; FONTES, M.M; VERA, S.G. Biodiesel/Diesel Blends Classification with Respect to Base Oil Using NIR Spectrometry and Chemometrics Tools. **Journal of the American Oil Chemists' Society**, v.89, n.1, p.1165-1171, 2012.

SOARES, S. F. C. **Um novo critério para seleção de variáveis usando o Algoritmo das Projeções**. 2010. 107f. Dissertação (Mestrado em Química) – Universidade Federal da Paraíba.

SOARES, S. F. C., GOMES, A. A., GALVÃO FILHO, A. R., ARAÚJO, M. C. U. The successive projections algorithm. **Trends in Analytical Chemistry**, v.42, n.1, p. 84–98, 2013.

SOLOMOM, C.; BRECKON, T. **Fundamentals of Digital Image Processing** - A Practical Approach with Examples in Matlab. 1st Edition. United States of America: John Wiley & Sons Ltd, 2011.

SOUTO, R.P. **Segmentação de imagem multiespectral utilizando-se o atributo matiz**. 2000. Disponível em: <<http://www.obt.inpe.br/pgsere/Souto-R-P-2000/publicacao.pdf>>. Acesso em: 15 Dez. 2014.

SOUTO, U.T.C.P.; BARBOSA, M.F.; DANTAS, H.V.; PONTES, A.S.; LYRA, W.S.; DINIZ, P.H.G.D.; ARAÚJO, M.C.U.; SILVA, E.C. Screening for Coffee Adulteration Using Digital Images and SPA-LDA. **Food Analytical Methods**, In Press doi 10.1007/s12161-014-0020-7.

SRIVASTAVA, M.S.; CARTER, E.M.; Na Introduction to applied multivariate statistics. **Elsevier**, v.1, n.1, p.243-245, 1983.

SUAREZ, P.A.Z.; SANTOS, A.L.F.; RODRIGUES, J.P.; ALVES M.B. Biocombustíveis a partir de óleos e gorduras: desafios tecnológicos para viabilizá-los. **Química Nova**, v.32, n.1, p.768-775, 2009.

Technical Committee of the Institute of Shortening and Edible Oils, Food Fats and Oils, Institute of Shortening and Edible Oils, ISEO, Washington, 2006.

THOMAS, E. A primer on multi variate calibration. **Analytical Chemistry** , v.66, n.1, p.795- 804, 1994.

TRAINA, A.J.M.; OLIVEIRA, M.C.F. **Apostila de Computação Gráfica**. 2004. Disponível em: <http://gbdi.icmc.usp.br/documentacao/apostilas/cg>. Acesso em: 10 Dez. 2014.

ULRICI, A.; FOCA, G.; IELO, M.C.; VOLPELLI, L.A.; LO FIEGO, D.P. Automated identification and visualization of food defects using RGB imaging: Application to the detection of red skin defect of raw hams. **Innovative Food Science and Emerging Technologies**, v.16, n.1, p. 417- 426, 2012.

VAN DALEN, G. (2004). Determination of the size distribution and percentage of broken kernels of rice using flatbed scanning and image analysis. **Food Research International**, v.37, n.1, p. 51-57, 2004.

VERAS, G.; BRITO, A.L.B.; SILVA, A.C.; COSTA, G.B.; FELIX, C.N.; FERNANDES, D.D.S.; FONTES, M.M. Classificação de biodiesel na região do visível. **Química Nova**, v.35, n. 2, p. 315–318, 2012.

VÉRAS, G.; COSTA, G.B.; FREITAS, J.P.; MEDEIROS, A.C.; ARAUJO, A.A.; FERNANDES, D.D.S. Screening analysis of biodiesel mixture feedstock using near infrared spectrometry. **Nir news**, v.24, n.1, p.6-10, 2013.

VERAS, G.; GOMES, A. A.; SILVA, A. C.; BRITO, A. L. B.; ALMEIDA, P. B. A.; MEDEIROS, E.P. Classification of biodiesel using NIR spectrometry and multivariate techniques. **Talanta**, v.83, n.1, p. 565-568, 2010.

VILAR, W.T. **Classificação individual de sementes de mamona usando espectroscopia de reflectância no visível, imagens digitais e análises multivariadas**. 2014. 113f. Dissertação (Mestrado em Química) – Universidade Federal da Paraíba.