



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



DISSERTAÇÃO DE MESTRADO

**Classificação de óleos vegetais comestíveis usando imagens
digitais e técnicas de reconhecimento de padrões**

Karla Danielle Tavares de Melo Milanez

João Pessoa – PB – Brasil

Agosto/2013



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

DISSERTAÇÃO DE MESTRADO

Classificação de óleos vegetais comestíveis usando imagens digitais e técnicas de reconhecimento de padrões

Karla Danielle Tavares de Melo Milanez*

Dissertação apresentada como requisito para obtenção de título de Mestre em Química pela Universidade Federal da Paraíba.

Orientador: Prof. Dr. Márcio José Coelho de Pontes

*Bolsista CAPES

João Pessoa – PB – Brasil

Agosto/2013

M637c Milanez, Karla Danielle Tavares de Melo.
Classificação de óleos vegetais comestíveis usando imagens digitais e técnicas de reconhecimento de padrões / Karla Danielle Tavares de Melo Milanez.-- João Pessoa, 2013.
102f. : il.

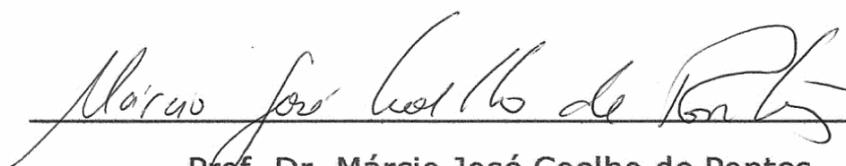
Orientador: Márcio José Coelho de Pontes
Dissertação (Mestrado) – UFPB/CCEN
1. Química. 2. Óleos vegetais comestíveis. 3. Imagens digitais. 4. Análise discriminante linear. 5. Seleção de variáveis.

UFPB/BC

CDU: 54(043)

Dissertação de Mestrado submetida ao Corpo Docente do Programa de Pós-Graduação em Química do Departamento de Química do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba como parte dos requisitos para a obtenção do grau de Mestre em Química, área de concentração Química Analítica.

Aprovada pela banca examinadora:



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



Profa. Dra. Fernanda Araújo Honorato
Examinadora



Profa. Dra. Cláudia de Oliveira Cunha
Examinadora

Aos meus amados pais, Luis Carlos e Iricelia, como forma de amor e gratidão.

A Lucas Milanez, pelo carinho, paciência e compreensão.

À minha querida irmã, Anna Laura.

Com carinho, dedico.

AGRADECIMENTOS

- A Deus;
- Aos meus queridos pais, Luis Carlos e Iricelia, que com tanto amor e dedicação me guiaram até aqui;
- À minha irmã Laura, pela tolerância a tantos momentos estressantes;
- A Lucas Milanez, pelo amor, incentivo, compreensão e, principalmente, pela paciência nos momentos mais difíceis;
- À toda minha família pela torcida, mesmo de longe;
- Ao Professor Márcio José Coelho Pontes, pela orientação, dedicação e confiança a longo deste trabalho;
- Ao Professor Mário César Ugulino Araújo, pela acolhida acadêmica;
- Aos professores Pablo Nogueira e Ricardo Alexandre pela ajuda nos momentos decisivos deste trabalho;
- A Sófacles, Renato, Adenilton, Hebertty e Ambrosina, pela ajuda importantíssima e por toda contribuição que me ajudou a finalizar este trabalho;
- Às minhas queridas amigas Danielle, Gabriela, Clarissa e ao meu amigo Felipe, pela amizade, carinho, paciência, respeito, confiança e tantos momentos de estresse e diversão compartilhados;
- A Universidade Federal da Paraíba;
- À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, pela bolsa concedida;
- A todos que contribuíram indiretamente para a realização deste trabalho.

SUMÁRIO

LISTA DE FIGURAS	iii
LISTA DE TABELAS	vi
LISTA DE ABREVIATURAS	viii
RESUMO	ix
ABSTRACT	x

1 INTRODUÇÃO	2
1.1 CARACTERIZAÇÃO GERAL DA PROBLEMÁTICA	2
1.2 OBJETIVOS	3
1.2.1 <i>Objetivo geral</i>	3
1.2.2 <i>Objetivos específicos</i>	3
2 FUNDAMENTAÇÃO TEÓRICA	6
2.1 ÓLEOS VEGETAIS	6
2.1.1 <i>Composição dos óleos vegetais</i>	6
2.1.1.1 Os não glicerídeos e a cor dos óleos vegetais	7
2.1.1.2 Ácidos graxos	8
2.1.2 <i>Fontes de óleos vegetais</i>	9
2.1.2.1 Óleo de canola	10
2.1.2.2 Óleo de milho	10
2.1.2.3 Óleo de girassol	11
2.1.2.4 Óleo de soja	11
2.1.3 <i>Principais processos degradativos</i>	12
2.1.3.1 Oxidação enzimática	12
2.1.3.2 Fotoxidação	13
2.1.3.3 Autoxidação	13
2.1.4 <i>Estabilidade oxidativa</i>	14
2.1.5 <i>Óleos vegetais e sua importância no metabolismo humano</i>	15
2.1.6 <i>Química analítica no controle de qualidade dos óleos vegetais</i>	15
2.2 A NATUREZA DA LUZ E O ESPECTRO ELETROMAGNÉTICO	19
2.2.1 <i>Interação luz-objeto e seus efeitos na percepção das cores</i>	20
2.2.1.1 Reflexão da luz	21
2.2.1.2 Transmissão da luz	21
2.3 O SISTEMA VISUAL HUMANO	22
2.4 MODELOS DE CORES	24
2.4.1 <i>Modelo RGB de cores</i>	25
2.4.2 <i>Modelo HSI de cores</i>	27
2.5 DAS IMAGENS ANALÓGICAS ÀS IMAGENS DIGITAIS	29
2.5.1 <i>Imagens digitais</i>	30
2.5.1.1 Pixel	32
2.5.1.2 Histogramas	33
2.6 FERRAMENTAS QUIMIOMÉTRICAS	35
2.6.1 <i>Técnicas de reconhecimento de padrões não supervisionadas</i>	35
2.6.1.1 Análise de Componentes Principais (PCA)	36
2.6.2 <i>Técnicas de reconhecimento de padrões supervisionadas</i>	37

2.6.2.1	Análise Discriminante Linear (LDA)	38
2.6.2.2	Algoritmo das Projeções Sucessivas (SPA)	39
2.6.2.3	Stepwise (SW)	40
2.6.2.4	Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA)	42
3	METODOLOGIA	45
3.1	AMOSTRAS	45
3.2	SISTEMA PARA AQUISIÇÃO DAS IMAGENS	45
3.3	COMPARTIMENTO	46
3.4	ILUMINAÇÃO DO SISTEMA	47
3.5	BOMBA PERISTÁLTICA	47
3.6	NOTEBOOK	47
3.7	AQUISIÇÃO DAS IMAGENS	47
3.8	ANÁLISE QUIMIOMÉTRICA	50
4	RESULTADOS E DISCUSSÃO	54
4.1	HISTOGRAMAS	54
4.2	CLASSIFICAÇÃO COM RESPEITO AO TIPO	55
4.2.1	<i>Análise de Componentes Principais (PCA)</i>	56
4.2.2	<i>Escolha da faixa de trabalho</i>	59
4.2.3	<i>Modelos LDA</i>	60
4.2.4	<i>Classificação PLS-DA</i>	63
4.3	CLASSIFICAÇÃO COM RESPEITO AO ESTADO DE CONSERVAÇÃO	64
4.3.1	<i>Histogramas</i>	64
4.3.2	<i>Análise de Componentes Principais (PCA)</i>	66
4.3.3	<i>Escolha da faixa de trabalho</i>	67
5	CONCLUSÕES	74
5.1	PERSPECTIVAS	75
	REFERÊNCIAS BIBLIOGRÁFICAS	77

LISTA DE FIGURAS

FIGURA 2.1. REAÇÃO DE CONDENSAÇÃO ENTRE UMA MOLÉCULA DE GLICEROL E TRÊS MOLÉCULAS DE ÁCIDOS GRAXOS, ONDE O PRODUTO PRINCIPAL É UMA MOLÉCULA DE TRIGLICERÍDEO.....	7
FIGURA 2.2. ESTRUTURA MOLECULAR DOS ÁCIDOS GRAXOS DE MAIOR OCORRÊNCIA NA NATUREZA.....	9
FIGURA 2.3. AUTOXIDAÇÃO DOS ÓLEOS VEGETAIS, ONDE: RH – ÁCIDO GRAXO INSATURADO, R· – RADICAL LIVRE, ROO· – RADICAL PERÓXIDO E ROOR – HIDROPERÓXIDO.....	14
FIGURA 2.4. ESPECTRO ELETROMAGNÉTICO.....	19
FIGURA 2.5. REFLEXÃO DA LUZ.	21
FIGURA 2.6. TRANSMISSÃO DA LUZ.	22
FIGURA 2.7. REPRESENTAÇÃO DO OLHO HUMANO.....	23
FIGURA 2.8. REPRESENTAÇÃO DE UM BASTONETE E DE UM CONE.....	23
FIGURA 2.9. CORES PRIMÁRIAS ADITIVAS E AS CORES RESULTANTES DE SUAS COMBINAÇÕES.	25
FIGURA 2.10. CUBO DE CORES RGB.	26
FIGURA 2.11. MODELO HSI DE CORES. (A) SÓLIDO DE CORES, (B) TRIÂNGULO DE CORES.....	28
FIGURA 2.12. REPRESENTAÇÃO DOS ATRIBUTOS DA COR.....	29
FIGURA 2.13. IMAGEM MONOCROMÁTICA. OS NÍVEIS DE CINZA CORRESPONDEM À INTENSIDADE LUMINOSA DA IMAGEM F EM CADA COORDENADA ESPACIAL (X,Y) QUE COMPÕEM A IMAGEM.....	30
FIGURA 2.14. IMAGEM MONOCROMÁTICA E SUA REPRESENTAÇÃO EM IMAGEM DIGITAL.	31
FIGURA 2.15. AS TRÊS BANDAS DE UMA IMAGEM, SEPARADAMENTE, E A IMAGEM COLORIDA RESULTANTE DA COMPOSIÇÃO DESSAS TRÊS BANDAS.	32
FIGURA 2.16. HISTOGRAMAS REFERENTES A DISTRIBUIÇÃO DE FREQUÊNCIA DE TODOS OS POSSÍVEIS VALORES DE UM PIXEL NOS NÍVEIS DE CINZA, VERMELHO, VERDE E AZUL DE UMA IMAGEM.	34

- FIGURA 2.17.** PRIMEIRA E SEGUNDA PCS OBTIDAS A PARTIR DA COMBINAÇÃO LINEAR DAS VARIÁVEIS ORIGINAIS. AS LINHAS TRACEJADAS NA COR VERDE CORRESPONDEM AOS ESCORES E AS LINHAS TRACEJADAS-PONTILHADAS NA COR CINZA REPRESENTAM OS PESOS.....37
- FIGURA 2.18.** REPRESENTAÇÃO ESQUEMÁTICA DA ANÁLISE DISCRIMINANTE LINEAR.....38
- FIGURA 2.19.** REPRESENTAÇÃO DO MODELO PREDITIVO PLS-DA.....42
- FIGURA 3.1.** SISTEMA UTILIZADO NA AQUISIÇÃO DAS IMAGENS DAS AMOSTRAS DE ÓLEOS VEGETAIS COMESTÍVEIS. (A) CAIXA; (B) WEBCAM; (C) CUBETA DE FLUXO; (D) BOMBA PERISTÁLTICA (E) LÂMPADA FLUORESCENTE E (F) NOTEBOOK.45
- FIGURA 3.2.** IMAGEM INTERNA DO COMPARTIMENTO.....46
- FIGURA 3.3.** IMAGENS DOS ÓLEOS VEGETAIS OBTIDAS PELA WEBCAM NO INTERIOR DO SISTEMA DESENVOLVIDO.....48
- FIGURA 3.4.** IMAGENS DE ÓLEOS DE CANOLA OBTIDAS PARA UMA AMOSTRA NÃO EXPIRADA E UMA EXPIRADA À 24 MESES.....48
- FIGURA 3.5.** REPRESENTAÇÃO DE UMA AMOSTRA NA MATRIZ DE DADOS.49
- FIGURA 3.6.** REPRESENTAÇÃO DA MATRIZ DE DADOS USADA NO TRATAMENTO DOS DADOS.....50
- FIGURA 4.1.** HISTOGRAMAS DAS 132 AMOSTRAS DE ÓLEOS VEGETAIS COMESTÍVEIS OBTIDOS PARA AS COMPONENTES DE CORES TONS DE CINZA, R: *RED*; G: *GREEN*; B: *BLUE*; H: *HUE*; S: *SATURATION* E I: *INTENSITY*.54
- FIGURA 4.2.** HISTOGRAMAS DAS 132 AMOSTRAS DE ÓLEOS VEGETAIS COMESTÍVEIS APÓS A REMOÇÃO DOS ZEROS. TAIS HISTOGRAMAS FORAM USADOS NO TRATAMENTO QUIMIOMÉTRICO. TONS DE CINZA, R: *RED*; G: *GREEN*; B: *BLUE*; H: *HUE*; S: *SATURATION* E I: *INTENSITY*. OS VALORES APRESENTADOS NA PARTE SUPERIOR DA FIGURA CORRESPONDEM AOS NÍVEIS DE CORES PRESENTES NOS HISTOGRAMAS ORIGINAIS.....55
- FIGURA 4.3.** HISTOGRAMAS MÉDIOS DAS QUATRO CLASSES DE ÓLEOS EM ESTUDO, OBTIDOS PARA AS COMPONENTES TONS DE CINZA, R, G, B, H, S E I, RESPECTIVAMENTE.55
- FIGURA 4.4.** (A) GRÁFICOS DOS ESCORES DE PC1 X PC2, OBTIDOS PARA AS COMPONENTES DE CORES TONS DE CINZA + RGB + HSI E (B) GRÁFICO DOS PESOS DAS VARIÁVEIS EM CADA COMPONENTE DE COR. (●: CANOLA, ■: MILHO, ▲: GIRASSOL E ◆: SOJA).....57
- FIGURA 4.5.** GRÁFICOS DOS ESCORES DE PC1 X PC2, OBTIDOS PARA AS COMPONENTES DE CORES (A) TONS DE CINZA, (B) B E (C) S, PARA OS QUATRO TIPOS DE ÓLEOS VEGETAIS EM ESTUDO. (●: CANOLA, ■: MILHO, ▲: GIRASSOL E ◆: SOJA).58

- FIGURA 4.6.** (A) GRÁFICO DA FUNÇÃO DE CUSTO DO CONJUNTO DE TREINAMENTO *VERSUS* O NÚMERO DE VARIÁVEIS SELECIONADAS PELO ALGORITMO SPA, (B) VARIÁVEIS SELECIONADAS PELO SPA.60
- FIGURA 4.7.** VARIÁVEIS SELECIONADAS PELA FORMULAÇÃO SW EMPREGADAS NA CONSTRUÇÃO DO MODELO SW-LDA.61
- FIGURA 4.8.** GRÁFICOS DOS ESCORES DAS FUNÇÕES DISCRIMINANTES FD2 X FD1 OBTIDOS PELO (A) SPA E (B) SW PARA AS 60 AMOSTRAS DE ÓLEOS VEGETAIS NÃO EXPIRADAS. (•: CANOLA, ■: MILHO, ▲: GIRASSOL E ◆: SOJA).62
- FIGURA 4.9.** HISTOGRAMAS MÉDIOS DAS AMOSTRAS DE ÓLEOS EXPIRADAS E NÃO EXPIRADAS DE (A) CANOLA, (B) MILHO, (C) GIRASSOL E (D) SOJA.65
- FIGURA 4.10.** GRÁFICOS DOS ESCORES OBTIDOS A PARTIR DE TODAS AS COMPONENTES DE CORES PARA AS AMOSTRAS EXPIRADAS E NÃO EXPIRADAS DE (A) CANOLA, (B) MILHO, (C) GIRASSOL E (D) SOJA. EM TODOS OS CASOS, OS SÍMBOLOS PREENCHIDOS NA COR PRETA CORRESPONDEM ÀS AMOSTRAS EXPIRADAS.66
- FIGURA 4.11.** GRÁFICOS DOS ESCORES DA FUNÇÃO DISCRIMINANTE 1 (FD1) PARA AS AMOSTRAS EXPIRADAS E NÃO EXPIRADAS DE: (A) CANOLA, NAS COMPONENTES HSI, (B) MILHO, NAS COMPONENTES HSI, (C) GIRASSOL, NA COMBINAÇÃO DE COMPONENTES TONS DE CINZA + HSI E (D) SOJA, NA COMBINAÇÃO DE COMPONENTES TONS DE CINZA + HSI. EM TODOS OS CASOS, AS AMOSTRAS EXPIRADAS SÃO REPRESENTADAS PELOS SÍMBOLOS PREENCHIDOS NA COR PRETA.71

LISTA DE TABELAS

TABELA 2.1. PRINCIPAIS VEGETAIS OLEAGINOSOS E SEUS RESPECTIVOS CONTEÚDOS DE ÓLEO.	9
TABELA 2.2 TEOR DE ÁCIDOS GRAXOS NOS ÓLEOS VEGETAIS.....	12
TABELA 2.3. PRINCIPAIS CORES DA LUZ VISÍVEL E SUAS FAIXAS DO ESPECTRO.....	20
TABELA 3.1. NÚMERO DE AMOSTRAS DE TREINAMENTO E PREVISÃO EM CADA CLASSE, PARA O PRIMEIRO CASO.	51
TABELA 3.2. NÚMERO DE AMOSTRAS DE TREINAMENTO E PREVISÃO EM CADA CLASSE, PARA O SEGUNDO CASO.	51
TABELA 4.1. TAXA DE CLASSIFICAÇÃO CORRETA OBTIDA COM OS MODELOS DE CLASSIFICAÇÃO LDA/SPA, LDA/SW E PLS-DA PARA O CONJUNTO DE PREVISÃO.....	59
TABELA 4.2. RESULTADOS DA CLASSIFICAÇÃO OBTIDA COM OS MODELOS LDA/SPA, LDA/SW E PLS-DA PARA OS CONJUNTOS DE PREVISÃO E TREINAMENTO. (1): AMOSTRAS DE CANOLA, (2): AMOSTRAS DE MILHO, (3): AMOSTRAS DE GIRASSOL E (4): AMOSTRAS DE SOJA. O NÚMERO DE VARIÁVEIS OU FATORES EMPREGADOS EM CADA MODELO É INDICADO ENTRE PARÊNTESES. N INDICA O NÚMERO DE AMOSTRAS DE PREVISÃO UTILIZADAS NESTE ESTUDO.	63
TABELA 4.3. RESULTADOS DA CLASSIFICAÇÃO OBTIDA COM OS MODELOS LDA/SPA, LDA/SW E PLS-DA PARA OS CONJUNTOS DE TREINAMENTO E PREVISÃO DAS AMOSTRAS DE MILHO. (E): AMOSTRAS EXPIRADAS, (NE): AMOSTRAS NÃO EXPIRADAS. O NÚMERO DE VARIÁVEIS OU FATORES EMPREGADOS EM CADA MODELO É INDICADO ENTRE PARÊNTESES. N INDICA O NÚMERO DE AMOSTRAS DE PREVISÃO UTILIZADAS NESTE ESTUDO.	67
TABELA 4.4. RESULTADOS DA CLASSIFICAÇÃO OBTIDA COM OS MODELOS LDA/SPA, LDA/SW E PLS-DA PARA OS CONJUNTOS DE TREINAMENTO E PREVISÃO DAS AMOSTRAS DE SOJA. (E): AMOSTRAS EXPIRADAS, (NE): AMOSTRAS NÃO EXPIRADAS. O NÚMERO DE VARIÁVEIS OU FATORES EMPREGADOS EM CADA MODELO É INDICADO ENTRE PARÊNTESES. N INDICA O NÚMERO DE AMOSTRAS DE PREVISÃO UTILIZADAS NESTE ESTUDO.	68
TABELA 4.5. RESULTADOS DA CLASSIFICAÇÃO OBTIDA COM OS MODELOS LDA/SPA, LDA/SW E PLS-DA PARA OS CONJUNTOS DE TREINAMENTO E PREVISÃO DAS AMOSTRAS DE CANOLA. (E): AMOSTRAS EXPIRADAS, (NE): AMOSTRAS NÃO EXPIRADAS. O NÚMERO DE VARIÁVEIS OU FATORES EMPREGADOS EM CADA MODELO É INDICADO ENTRE PARÊNTESES. N INDICA O NÚMERO DE AMOSTRAS DE PREVISÃO UTILIZADAS NESTE ESTUDO.	69

TABELA 4.6 RESULTADOS DA CLASSIFICAÇÃO OBTIDA COM OS MODELOS LDA/SPA, LDA/SW E PLS-DA PARA OS CONJUNTOS DE TREINAMENTO E PREVISÃO DAS AMOSTRAS DE GIRASSOL. (E): AMOSTRAS EXPIRADAS, (NE): AMOSTRAS NÃO EXPIRADAS. O NÚMERO DE VARIÁVEIS OU FATORES EMPREGADOS EM CADA MODELO É INDICADO ENTRE PARÊNTESES. N INDICA O NÚMERO DE AMOSTRAS DE PREVISÃO UTILIZADAS NESTE ESTUDO.70

TABELA 4.7. TCC OBTIDA PELO MELHOR MODELO DE CLASSIFICAÇÃO (EM CADA OLEAGINOSA) APLICADO AO CONJUNTO COMPLETO DE AMOSTRAS.70

ABREVIATURAS

ANVISA	Agência Nacional de Vigilância Sanitária
CCD	Dispositivo acoplado por carga
CMY	Ciano, Magenta, Amarelo
E	Expirada
FD	Função discriminante
FTIR	Espectroscopia no infravermelho com transformada de Fourier
HDL	Lipoproteína de alta densidade
HPLC	Cromatografia líquida de alta eficiência
HSI	Matiz, Saturação, Intensidade
LDA	Análise discriminante linear
LDL	Lipoproteína de baixa densidade
MID	Espectroscopia no infravermelho médio
NE	Não expirada
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
SCV	Sinal composto de vídeo
SIMCA	Modelagem independente e flexível por analogia de classe
SPA	Algoritmo das projeções sucessivas
SW	Stepwise
TCC	Taxa de classificação correta
UV-VIS	Ultravioleta e visível

RESUMO

MILANEZ, K.D.T.M. **Classificação de óleos vegetais comestíveis usando imagens digitais e técnicas de reconhecimento de padrões.** 2013. 84 p. Dissertação (Mestrado em Química) – Laboratório de Automação e Instrumentação em Química Analítica, Universidade Federal da Paraíba, João Pessoa, 2013.

Este trabalho apresenta um método simples e de baixo custo, baseado na utilização de imagens digitais e técnicas de reconhecimento de padrões, para a classificação de óleos vegetais comestíveis com relação ao tipo (soja, canola, girassol e milho) e ao estado de conservação (prazo de validade expirado e não expirado). Para isso, imagens das amostras de óleos vegetais foram obtidas a partir de uma webcam e, em seguida, as mesmas foram decompostas em histogramas contendo as distribuições dos níveis de cores permitidos a um pixel. Três modelos para a cor de um pixel foram utilizados: vermelho-verde-azul (RGB), matiz-saturação-intensidade (HSI) e tons de cinza. A análise discriminante linear (LDA) foi utilizada para o desenvolvimento de modelos de classificação com base em um subconjunto reduzido de variáveis. Para fins de seleção de variáveis, duas técnicas foram utilizadas: o algoritmo das projeções sucessivas (SPA) e o stepwise (SW). Modelos baseados na análise discriminante por mínimos quadrados parciais (PLS-DA) aplicados aos histogramas completos (sem seleção de variáveis) também foram utilizados com o propósito de comparação. No estudo envolvendo a classificação com respeito ao tipo, modelos LDA/SPA, LDA/SW e PLS-DA atingiram uma taxa de classificação correta (TCC) de 95%, 90% e 95%, respectivamente. Na identificação de amostras expiradas e não expiradas, o modelo LDA/SPA foi considerado o melhor método para a classificação das amostras de óleos de girassol, soja e canola, atingindo uma TCC de 97%, 94% e 93%, respectivamente, enquanto que o modelo LDA/SW classificou corretamente 100% das amostras de milho. Estes resultados sugerem que o método proposto é uma alternativa promissora para a inspeção de autenticidade e estado de conservação de óleos vegetais comestíveis. Como vantagem, a metodologia não utiliza reagentes, a análise é rápida e procedimentos laboriosos para a caracterização química das amostras não são necessários.

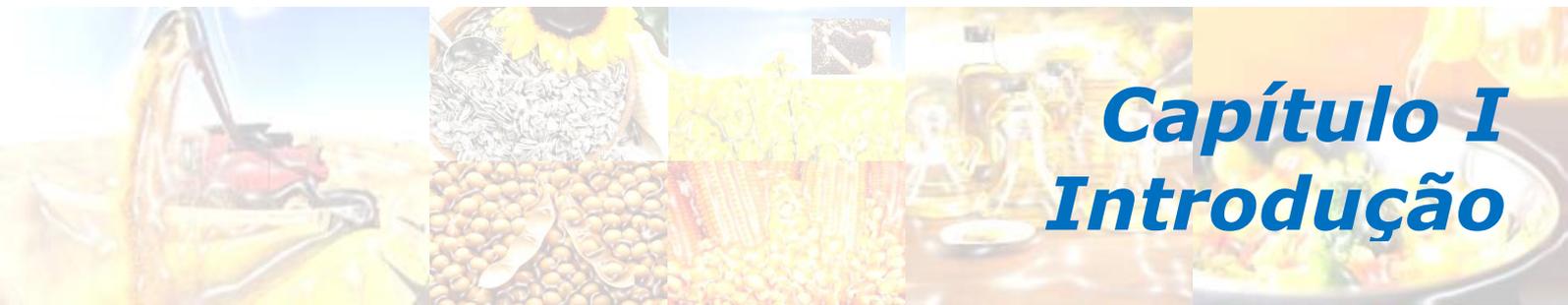
Palavras-chave: Óleos vegetais comestíveis, Imagens digitais, Análise discriminante linear, Seleção de variáveis.

ABSTRACT

MILANEZ, K.D.T.M. **Classification of edible vegetable oil using digital image data and pattern recognition techniques.** 2013. 84 p. Dissertation (Masters thesis in Chemistry) – Laboratório de Automação e Instrumentação em Química Analítica, Universidade Federal da Paraíba, João Pessoa, 2013.

This work presents a simple and non-expensive based on digital image and pattern recognition techniques for the classification of edible vegetable oils with respect to the type (soybean, canola, sunflower and corn) and the conservation state (expired and non-expired period of validity). For this, images of the sample oils were obtained from a webcam, and then, they were decomposed into histograms containing the distribution of color levels allowed for a pixel. Three representations for the color of a pixel were used: red-green-blue (RGB), hue-saturation-intensity (HSI) and grayscale. Linear discriminant analysis (LDA) was employed in order to build classification models on the basis of a reduced subset of variables. For the purpose of variable selection, two techniques were utilized, namely the successive projections algorithm (SPA) and stepwise (SW) formulation. Models based on partial least squares-discriminant analysis and (PLS-DA) applied to full histograms (without variable selection) were also employed for the purpose of comparison. For the study evolving the classification with respect to oil type, LDA/SPA, LDA /SW and PLS-DA models achieved a correct classification rate (CCR) of 95%, 90% and 95%, respectively. For the identification of expired non-expired samples, LDA / SPA models were found to the best method for classifying sunflower, soybean and canola oils, achieving a TCC of 97%, 94% and 93%, respectively, while the model LDA/SW correctly classified 100% of corn oil samples. These results suggest that the proposed method is a promising alternative for inspection of authenticity and the conservation state of edible vegetable oils. As advantages, the method does not use reagents to carry out the analysis and laborious procedures for chemical characterization of the samples are not required.

Keywords: Edible vegetable oil, Digital image, Linear discriminant analysis, Variables selection.



Capítulo I ***Introdução***

1 INTRODUÇÃO

1.1 Caracterização geral da problemática

A qualidade dos produtos oferecidos comercialmente é um fator de extrema importância, considerando-se os aspectos industriais e as necessidades de um mercado consumidor cada vez mais exigente.

Em escala industrial, qualidade é garantia de competitividade e melhorias na cadeia produtiva (SILVA et al., 2006). Do ponto de vista do consumidor, verificar a autenticidade e a qualidade dos produtos comercializados envolve aspectos econômicos e relacionados à saúde. De fato, a escolha de um produto é essencialmente motivada pelo seu valor e pelas propriedades benéficas relacionadas ao mesmo. Nesse contexto, surge a necessidade de metodologias capazes de avaliar a qualidade e a autenticidade de produtos de forma simples, rápida e de baixo custo.

As técnicas baseadas em imagens digitais têm se revelado uma boa alternativa para as análises de controle de qualidade. A disponibilidade e o baixo custo dos computadores pessoais, câmeras digitais e aparelhos de digitalização, além da simplicidade e rapidez com que as análises são realizadas, têm contribuído para o aumento do número de pesquisas tecnológicas empregando-se estes equipamentos (MARQUES FILHO e VIEIRA NETO, 1999).

No setor alimentício, além das vantagens citadas acima, as imagens digitais têm sido muito utilizadas nas análises de rotina dos padrões de qualidade, pois permitem monitorar parâmetros como tamanho, cor e textura, eliminando a subjetividade da análise e a dependência do sistema visual humano, que por sua vez, é profundamente influenciado pelas condições ambientes, podendo levar a uma série de inconsistências (ANTONELLI et al., 2004; DU e SUN, 2004).

Os óleos vegetais são um dos principais produtos extraídos de plantas e, devido às suas propriedades benéficas, foram incorporados à dieta humana (REDA e CARNEIRO, 2007), sendo um dos produtos mais

utilizados no preparo de alimentos, o que torna a verificação de sua autenticidade uma tarefa de extrema importância.

Diversas técnicas analíticas instrumentais vêm sendo empregadas para classificar e/ou autenticar óleos vegetais no que diz respeito ao tipo e integridade do produto, conforme será explicitado na **seção 2.1.6**. Entretanto, esses métodos são, em geral, laboriosos e os equipamentos empregados apresentam um custo alto de aquisição e manutenção.

Diante do exposto, é proposta a utilização de imagens digitais para classificação de óleos vegetais comestíveis quanto ao tipo (*canola, milho, girassol e soja*) e ao estado de conservação (*expirado e não expirado*).

1.2 Objetivos

1.2.1 Objetivo geral

Este trabalho propõe uma metodologia simples, rápida e de baixo custo, baseada na utilização de imagens digitais e técnicas de reconhecimento de padrões, para classificação de óleos vegetais comestíveis com respeito ao tipo de oleaginosa e ao estado de conservação.

1.2.2 Objetivos específicos

- Adaptar um ambiente para a aquisição das imagens dos óleos vegetais com ênfase no rigoroso controle de luminosidade.
- Decompor as imagens digitais obtidas em histogramas contendo a distribuição de pixels em diferentes canais (RGB, HSI e Tons de cinzas).
- Realizar um estudo exploratório dos dados com base na aplicação da análise de componentes principais aos histogramas obtidos.
- Comparar modelos baseados na LDA (com prévia seleção de variáveis pelo SPA e SW) com o PLS-DA em dois problemas de classificação: tipo de oleaginosa empregada (*soja, girassol, canola e*

milho) e estado de conservação (amostras expiradas e não expiradas).



Capítulo II ***Fundamentação Teórica***

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Óleos vegetais

Óleos e gorduras são alguns dos constituintes mais facilmente encontrados na composição dos seres vivos. Pertencem à classe dos lipídios, substâncias lipofílicas que desempenham importantes funções biológicas, podendo agir como hormônios, componentes estruturais das membranas celulares, impermeabilizantes e reserva energética (SILVA JUNIOR e SASSON, 1990).

Segundo a resolução RDC nº 270 (ANVISA, 2005), a diferença entre óleos e gorduras está no estado físico no qual os mesmos se encontram a temperatura de 25 °C. A esta temperatura, os óleos vegetais apresentam-se no estado líquido, enquanto que as gorduras vegetais encontram-se no estado sólido ou pastoso.

Os óleos vegetais são um dos principais produtos extraídos de plantas e, devido às suas propriedades benéficas, foram incorporados à dieta humana. São obtidos na forma bruta a partir das sementes ou grãos das oleaginosas através de processos físicos e químicos e em seguida, são destinados aos processos de refino, cuja finalidade é remover espécies indesejáveis que possam comprometer a qualidade e estabilidade do produto final, tornando-o próprio para o consumo (REDA e CARNEIRO, 2007).

2.1.1 Composição dos óleos vegetais

Os óleos vegetais são constituídos predominantemente por triglicerídeos, triésteres resultantes da reação de condensação entre uma molécula de glicerol e três moléculas de ácidos graxos (MORETTO e FETT, 1998) (**Fig. 2.1**). Os triglicerídeos podem ser simples, se os ácidos graxos que o constituem são iguais, ou compostos, se os ácidos graxos em sua composição são diferentes.

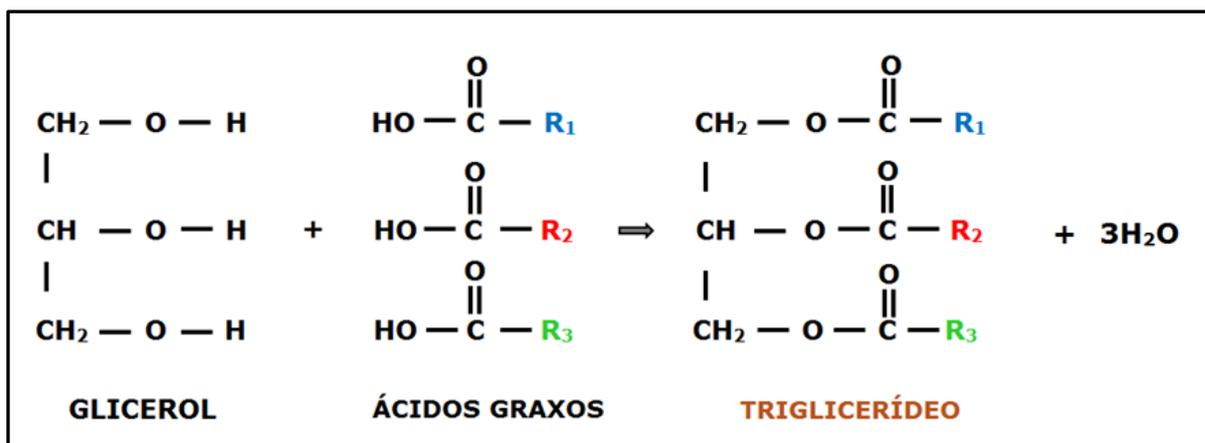


Figura 2.1. Reação de condensação entre uma molécula de glicerol e três moléculas de ácidos graxos, onde o produto principal é uma molécula de triglicerídeo.

Em menores proporções, são encontrados compostos chamados não glicerídeos, constituintes que conferem características organolépticas, afetam a estabilidade oxidativa e estão diretamente relacionadas às propriedades nutricionais dos óleos vegetais. Entre os principais não glicerídeos encontrados nos óleos vegetais estão: fosfatídeos, ceras, esteróis, hidrocarbonetos livres, resinas, tocoferóis e pigmentos como carotenoides, clorofila e gossipol (MORETTO e FETT, 1998; O'BRIEN, 2009).

Ainda na composição dos óleos, são encontrados também os ácidos graxos livres que, embora estejam presentes em pequenas quantidades, exercem um papel fundamental nas propriedades características dos diferentes tipos de óleos vegetais (MORETTO e FETT, 1998).

2.1.1.1 Os não glicerídeos e a cor dos óleos vegetais

Os pigmentos são compostos não glicerídeos responsáveis pelas características aparentes dos óleos vegetais. São eles que atribuem cor aos óleos. A coloração amarelo/vermelho é fornecida pelos carotenoides, enquanto a coloração esverdeada, presente no óleo de oliva e no óleo de soja antes do refino, é devido à clorofila. A maior concentração de carotenoides em óleos vegetais, aproximadamente 0,05 - 0,20%, encontra-se no azeite de dendê, imprimindo-lhe a intensa coloração

laranja/avermelhado. A cor amarelada do óleo de algodão é devido à presença de gossipol, pigmento amarelado presente principalmente nas sementes de algodão (MORETTO e FETT, 1998; O'BRIEN, 2009).

Os óleos brutos apresentam coloração bastante intensa que é removida durante as etapas de processamento. Quando o refino é realizado eficientemente, o óleo obtido é quase incolor, atendendo a uma das principais exigências do consumidor. Dessa forma, além de possibilitar inferências sobre a qualidade do processamento do óleo vegetal, a cor também pode ser utilizada para avaliar seu estado de conservação, uma vez que iniciado o processo de deterioração, alterações organolépticas passam a ocorrer e o aparecimento de cor é uma delas (MORETTO e FETT, 1998; O'BRIEN, 2009; SHAHIDI e ZHONG, 2005).

2.1.1.2 Ácidos graxos

Os ácidos graxos são os constituintes mais importantes das moléculas de glicerídeos e não glicerídeos, representando cerca de 95% do peso total dessas moléculas. As propriedades físicas, químicas e nutricionais dos óleos são profundamente influenciadas pela estrutura molecular dos grupos acila correspondentes aos ácidos graxos (MORETTO e FETT, 1998).

Os ácidos graxos são classificados como saturados ou insaturados, de acordo com ausência ou não de ligações duplas na cadeia. Entre os ácidos graxos saturados de maior ocorrência estão os ácidos láurico, palmítico e esteárico, enquanto que entre os ácidos graxos insaturados estão os ácidos oleico, linoleico e linolênico (MORETTO e FETT, 1998) **(Fig. 2.2)**. Estes últimos podem ainda ser chamados como ácidos graxos ω -9, ω -6 e ω -3, respectivamente, ressaltando a localização da primeira instauração a partir do terminal metil.

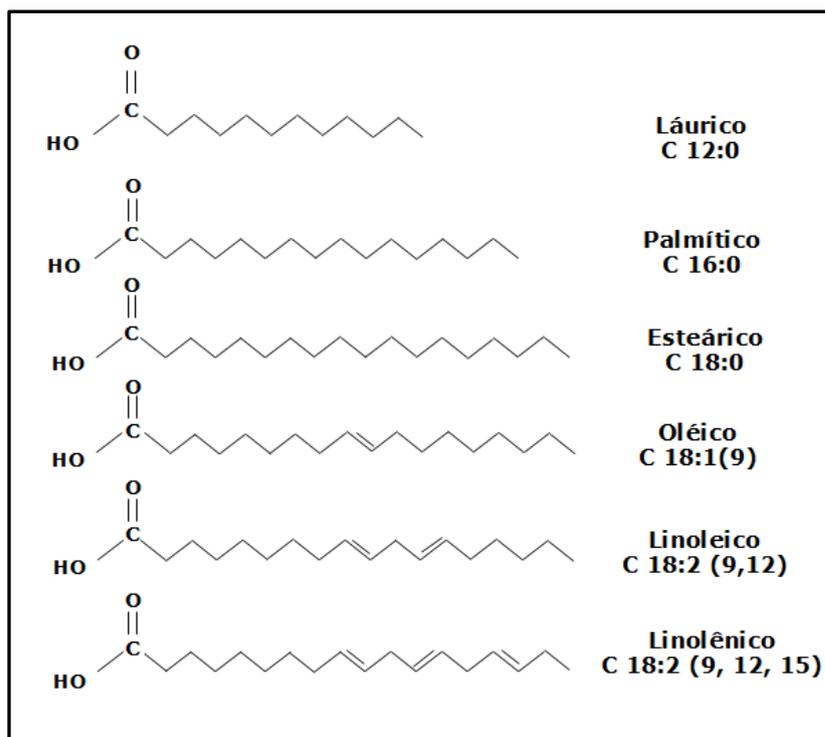


Figura 2.2. Estrutura molecular dos ácidos graxos de maior ocorrência na natureza.

2.1.2 Fontes de óleos vegetais

A **Tabela 2.1** lista os principais vegetais dos quais os óleos podem ser extraídos. Como pode ser observado, é grande a variedade de espécies fornecedoras de óleos, podendo ser sementes, polpas de frutos e alguns grãos.

Tabela 2.1. Principais vegetais oleaginosos e seus respectivos conteúdos de óleo.

Espécie oleaginosa	Conteúdo de óleo (% em massa)
Gergelim	50-55
Polpa de palma (dendê)	45-50
Amendoim	45-50
Colza (canola)	40-45
Girassol	35-45
Oliva	25-30
Algodão	18-20
Soja	18-20
Milho	3,1-5,7

Fonte: Adaptada de Moretto e Fett (1998).

Entre os tipos de óleos comercializados no Brasil com autorização da Agência Nacional de Vigilância Sanitária (ANVISA, 1999), estão: algodão, amendoim, arroz, canola, milho, girassol, soja, oliva, azeite saborizado, coco, palma e gergelim. Na realização deste estudo, foram utilizados os óleos de soja, girassol e canola, por estarem entre os óleos mais produzidos e consumidos em escala mundial (FAO, 2013), e o óleo de milho, por ser um dos óleos mais comumente encontrados no comércio local.

2.1.2.1 Óleo de canola

Embora conhecido desde os primórdios da Era Cristã, o óleo de canola passou a fazer parte da dieta humana após melhoramentos genéticos realizados em laboratórios canadenses por volta de 1960, quando foi possível a redução dos teores de ácido erúico e glucosilatonatos. O nome canola (*Canadian Oil Low Acid*) foi oficialmente aceito pela Canadian Grain Commission em 1987 e seu cultivo se alastrou por vários países (MORETTO e FETT, 1998).

Além do baixo índice de gordura saturada, apenas 6%, para 15% do óleo de soja e 11% do óleo de girassol, muitas das vantagens do óleo de canola estão relacionadas aos altos teores de gorduras insaturadas que auxiliam na prevenção de doenças circulatórias e coronárias diminuindo os índices no sangue do colesterol LDL e mantendo o colesterol HDL dentro da normalidade. Em sua composição, são encontrados ácidos graxos palmítico (4%), esteárico (3%), oleico (56%), linoleico (26%) e linolênico (10%). O óleo de canola apresenta-se na coloração amarelada com sabor e odor característico (MORETTO e FETT, 1998; GUNSTONE, 2005).

2.1.2.2 Óleo de milho

Devido à facilidade de adaptação as mais variadas condições ambientais, o milho, *Zea mays L.*, é um dos cereais mais cultivados em todo mundo (MORETTO e FETT, 1998).

O óleo de milho, extraído do gérmen do milho, é conhecido principalmente pela alta estabilidade oxidativa proporcionada pelo alto índice de insaturações e pela forma como os ácidos graxos estão distribuídos nas moléculas de triglicerídeos. Os principais ácidos graxos encontrados em sua composição são palmítico (9-17%), oleico (20-42%) e linoleico (39-63%). O óleo de milho apresenta-se como um óleo límpido, de cor amarelo claro, odor e sabor suave característico (MORETTO e FETT, 1998; GUNSTONE, 2005; CAMPESTRE, 2013).

2.1.2.3 Óleo de girassol

O óleo de girassol é obtido límpido, na coloração amarelo dourado claro, com odor e sabor suave característico a partir da semente *Helianthus annuus*, com rendimento de até 47%. Os maiores produtores mundiais de girassol são Rússia, Ucrânia, Argentina, União Europeia, China e Estados Unidos (CAMPESTRE, 2013; GUNSTONE, 2005; CONAB, 2011).

Por apresentar alto teor de ácido linoleico (ω -6) e de tocoferóis (vitamina E), e baixa quantidade de ácidos graxos saturados, o óleo de girassol ajuda na redução do nível de colesterol no sangue (ABOISSA, 2013).

2.1.2.4 Óleo de soja

De acordo com a EMBRAPA, o Brasil é o segundo maior produtor mundial de soja ficando atrás apenas dos Estados Unidos (EMBRAPA, 2007). O óleo de soja surgiu como um subproduto do processamento do farelo de soja e lidera o mercado mundial de óleos (MORETTO e FETT, 1998). No Brasil, quase 90% da produção de óleo é derivada dessa leguminosa (FERRARI et al., 2005).

O óleo de soja é rico em ácidos graxos insaturados como ácidos oleico, linoleico e linolênico e apresenta baixo teor de ácidos graxos saturados, o que o torna um importante aliado na prevenção de doenças

coronárias. Após o refino, o óleo de soja apresenta-se como um óleo de cor levemente amarelado, límpido com odor e sabor suave característico (CAMPESTRE, 2013).

Na **Tabela 2.2** são indicadas as proporções de ácidos graxos saturados e insaturados presentes nos óleos de canola, milho, girassol e soja (MORETTO e FETT, 1998).

Tabela 2.2 Teor de ácidos graxos nos óleos vegetais.

Óleos	Ácido graxo saturado	Ácido graxo monoinsaturado	Ácido graxo poliinsaturado	
			Ácido linoleico	Ácido linolênico
Canola	6%	58%	26%	10%
Milho	13%	25%	61%	1%
Girassol	11%	2%	69%	-
Soja	15%	24%	54%	7%

Fonte: Adaptada de Reda e Carneiro (2007).

2.1.3 Principais processos degradativos

A deterioração dos óleos vegetais comestíveis, processo que gera alterações organolépticas e redução do valor nutricional, está relacionada ao fenômeno de rancidez. Na rancidez do tipo oxidativa ocorre a formação de peróxidos e hidroperóxidos, espécies organolepticamente inertes que ao reagir produzem compostos voláteis, aldeídos e cetonas responsáveis pelo odor e ranço do alimento (MORETTO e FETT, 1998). Os processos oxidativos sofridos pelos óleos vegetais podem ocorrer a partir de diferentes rotas: oxidação enzimática, fotooxidação e autooxidação (RAMALHO E JORGE, 2006).

2.1.3.1 Oxidação enzimática

A oxidação enzimática ocorre pela ação das enzimas lipoxigenase, que atuam sobre os ácidos graxos poli-insaturados catalisando a adição de

oxigênio à cadeia hidrocarbonada poli-insaturada. Como resultado, são obtidos os peróxidos e hidroperóxidos responsáveis pela formação de produtos secundários (compostos voláteis, aldeídos e cetonas) (RAMALHO E JORGE, 2006).

2.1.3.2 Fotoxidação

A fotoxidação consiste na transformação do oxigênio normal em estado *tripleto* ($^3\text{O}_2$) em O_2 em estado *singlete* ($^1\text{O}_2$). A reação é essencialmente promovida pela luz em presença de fotossensibilizadores como clorofila em vegetais e compostos de hematina em animais (MORETTO e FETT, 1998; SILVA et al., 1999). O processo envolve reações radiculares e formação de hidroperóxidos que sofrem posterior degradação originando aldeídos, cetonas e hidrocarbonetos (SILVA et al., 1999).

2.1.3.3 Autoxidação

A autoxidação é o principal processo de deterioração dos óleos vegetais e pode ser definida como uma sequência de reações espontâneas, capazes de se auto-propagar, entre o oxigênio da atmosfera com os ácidos graxos insaturados presentes nos óleos. O processo pode ser acelerado por altas temperaturas ocasionando o aumento de ácido graxo livre e de matéria polares, formação de espuma, viscosidade e, sobretudo, mudança na coloração (SHAHIDI e ZHONG, 2005; SILVA et al., 1999).

A autoxidação ocorre em três estágios envolvendo a formação de radicais livres a partir de ácidos graxos insaturados, que sofrem posterior ataque do oxigênio da atmosfera gerando os produtos primários (peróxidos e hidroperóxidos) de oxidação. O processo termina quando dois radicais gerados combinam-se para formar produtos estáveis (produtos secundários de oxidação) obtidos por cisão e rearranjo dos peróxidos (RAMALHO E JORGE, 2006) (**Fig. 2.3**).

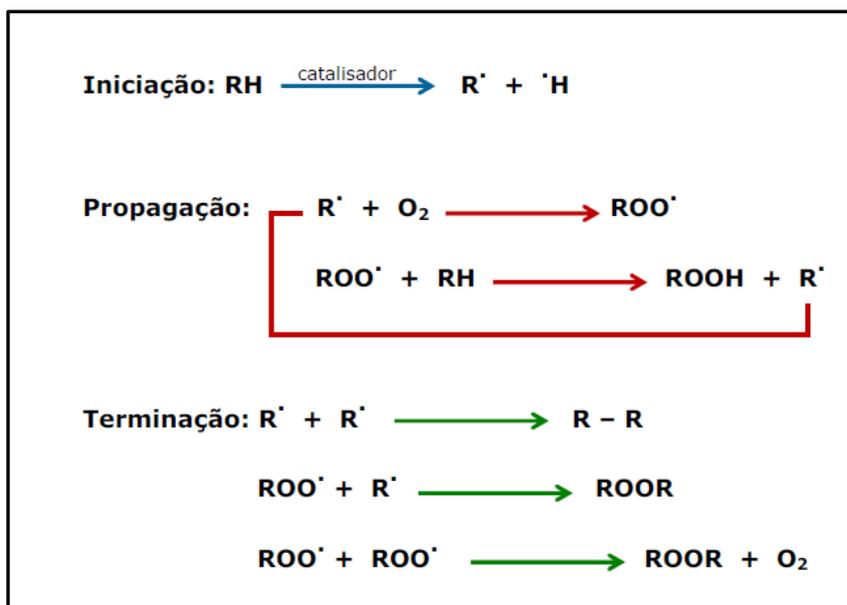


Figura 2.3. Autooxidação dos óleos vegetais, onde: RH – ácido graxo insaturado, R^\cdot – radical livre, ROO^\cdot – radical peróxido e ROOR – hidroperóxido.

2.1.4 Estabilidade oxidativa

A estabilidade dos óleos vegetais frente aos processos oxidativos é diretamente influenciada pela estrutura molecular dos triglicerídeos e pelo meio onde se encontram. Dessa forma, o número e a natureza das duplas ligações presentes, a exposição à luz e ao calor, a presença de pró-oxidantes ou de antioxidantes são fatores determinantes para a estabilidade oxidativa dos óleos vegetais (SILVA et al., 1999).

A ausência de ligações duplas na cadeia dos grupos acila proporciona maior estabilidade oxidativa aos lipídios. Entretanto, quando comparados às gorduras animais, os óleos vegetais mesmo com maior grau de instauração apresentam maior estabilidade à degradação devido à presença de tocoferóis, antioxidantes naturais que retardam a rancidez oxidativa (MORETTO e FETT, 1998; O'BRIEN, 2009). Os oito tocoferóis de ocorrência natural são α -, β -, γ -e δ -tocoferol, e α -, β -, γ - e δ -tocotrienol. A atividade antioxidante dos tocoferóis é principalmente devida à capacidade de doar seus hidrogênios fenólicos aos radicais livres lipídicos interrompendo a propagação em cadeia (BRAMLEY et al., 2000; RAMALHO e JORGE, 2006).

2.1.5 Óleos vegetais e sua importância no metabolismo humano

Os lipídios, como são chamados os óleos e as gorduras, são importantes fontes de energia e estão relacionados a uma série de funções no metabolismo humano, agindo como componentes estruturais das membranas celulares, impermeabilizantes, hormônios e reserva energética (SILVA JUNIOR e SASSON, 1990). Além disso, alguns ácidos graxos insaturados presentes na composição dos óleos vegetais também apresentam efeitos sobre os níveis de colesterol do fígado e do sangue, e auxiliam na proteção contra doenças coronárias e inflamatórias (BRANDÃO et al., 2005). A carência de alguns ácidos graxos essenciais na alimentação humana pode desencadear uma série de alterações imunológicas, neurológicas, no crescimento, na pele e ainda transtornos comportamentais, considerando-se o período perinatal (TINOCO et al., 2007).

2.1.6 Química analítica no controle de qualidade dos óleos vegetais

O controle da qualidade é uma forma de assegurar ao consumidor a conformidade do produto com os padrões estabelecidos pelos órgãos competentes. No que diz respeito aos produtos alimentícios, a qualidade pode ser expressa pela junção de características como valor nutricional, estabilidade e características organolépticas, além de segurança (em relação à microbiologia, contaminantes e toxinas) e autenticidade (KAMM et al., 2001).

Para avaliar a qualidade de um óleo vegetal comestível, parâmetros físico-químicos como, índice de acidez, índice de iodo, matéria insaponificável e índice de refração podem ser analisados. Diversas técnicas analíticas instrumentais têm sido empregadas para classificar e/ou autenticar óleos vegetais comestíveis, entre elas: voltametria (GAMBARRA-NETO et al., 2009), espectrometria de infravermelho com Transformada de Fourier (*Fourier-transform infrared spectroscopy* - FTIR) (LERMA-GARCÍA et al., 2010; JAVIDNIA et al., 2013; YANG et al., 2005),

cromatografia líquida de alta performance (*High Performance Liquid Chromatography* – HPLC) com detecção UV-VIS (CONCHA-HERRERA et al., 2010), espectrometria na região do infravermelho próximo (Near Infrared – NIR) e médio (Mid Infrared – MID) (SINELLI et al., 2010), espectrometria na região do infravermelho médio (Mid Infrared – MID) (LUNA et al., 2013), difração de raio-x (BORTOLETO et al., 2005), espectroscopia de fluorescência (SIKORSKA et al., 2005) e luminescência (POULLI et al., 2005), espectroscopia dielétrica (LIZHI et al., 2010), reflectometria de microonda (CATALDO et al., 2012) e espectrometria UV-VIS (PONTES et al., 2005). Entre as metodologias citadas acima, destacam-se os trabalhos de PONTES et al. (2005) e GAMBARRA-NETO et al. (2009).

PONTES et al. (2005) utilizaram a espectroscopia UV-VIS para classificar óleos vegetais comestíveis de acordo com tipo de oleaginosa (milho, canola, soja e girassol). Para essa finalidade, os autores aplicaram modelos de classificação baseados na Análise Discriminante Linear (LDA) com prévia seleção de variáveis pelo Algoritmo das Projeções Sucessivas (SPA) aos espectros UV-VIS. O objetivo era avaliar o potencial do algoritmo SPA para problemas de classificação. Para efeito de comparação, LDA com seleção de variáveis pelo algoritmo genético (GA) e a Modelagem Independente e Flexível por Analogia de Classe (SIMCA) aplicada ao espectro completo (*full-spectrum*) também foram utilizadas. Segundo os autores, o SPA revelou-se apropriado a partir da redefinição da função de custo, a fim de refletir o risco de erros de classificação para o método em questão.

GAMBARRA-NETO et al. (2009) realizou um estudo parecido com aquele reportado por PONTES et al (2005). Contudo, a voltametria de onda quadrada (SWV) foi empregada como técnica analítica. Além disso, essa metodologia foi capaz de classificar os óleos vegetais com respeito ao estado de conservação, ou seja, modelos de classificação (SPA-LDA e SIMCA) foram empregados tanto para classificar o tipo de óleo (soja, girassol, canola e milho), como para avaliar se as amostras estavam

vencidas ou não-vencidas. Na comparação dos modelos de classificação SIMCA e LDA, o modelo LDA com prévia seleção de variáveis pelo algoritmo SPA, forneceu os melhores resultados.

Nos últimos anos, técnicas baseadas em imagens digitais têm se revelado uma boa alternativa para as análises de controle de qualidade em amostras de alimentos (ZHENG et al., 2006). De fato, a disponibilidade e o baixo custo de microcomputadores, câmeras digitais e aparelhos de digitalização, além da simplicidade e rapidez com que as análises são realizadas, têm contribuído para o aumento do número de pesquisas voltadas para o uso desses equipamentos (MARQUE FILHO e VIEIRA NETO, 1999).

Em particular, no contexto das análises em alimentos, imagens digitais têm sido frequentemente exploradas para o monitoramento de parâmetros como tamanho, cor e textura, entre outros. Tais parâmetros, quando avaliados pelo sistema visual humano, são subjetivos e podem ser influenciados por condições ambientais, acarretando em uma série de inconsistências (ANTONELLI et al., 2004; DU e SUN, 2004).

SILVA e OLIVEIRA (2006) propuseram uma metodologia baseada no parâmetro cor para monitoramento da qualidade da cerveja tipo pilsen. As imagens das amostras obtidas a partir de um scanner foram convertidas em uma matriz de dados RGB e em seguida submetida à análise de componentes principais (PCA), onde ficou constatada a capacidade do método proposto em comparar diferentes fabricantes e o grau de escurecimento ocasionado por processos oxidativos.

MENDOZA et al. (2007) registraram imagens de batatas fritas comerciais e analisaram os parâmetros cor e textura para caracterizar e classificar as amostras de batatas de acordo com as preferências dos consumidores, nas categorias aceitáveis e não aceitáveis. As informações obtidas a cerca dos defeitos ou variações na superfície das amostras foram utilizadas na construção do modelo LDA, onde foram atingidos resultados de classificação de 95,8% de acerto para os dados de

treinamento e 90% para os dados de validação, mostrando que a junção imagens-quimiometria apresenta grande potencial de classificação.

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.

MENDOZA et al. (2009) utilizaram a cor e a textura de amostras de presuntos comerciais fatiados para caracterizá-los nos tipos porco, peru e frango, cozidos ou defumados. As imagens foram obtidas por uma câmera digital e as informações contidas na escala de cinza, RGB, L^* , a^* , b^* , H, S e V foram usadas para a construção de funções discriminantes (FD) que, apesar da complexidade e variabilidade de cor, textura e aparência, capturaram diferenças características entre os tipos de presunto em estudo.

Mais recentemente, 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 modelos RGB, HSI e tons 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 as amostras foram corretamente classificadas de acordo com suas regiões geográficas.

ULRICI et al. (2012) utilizaram imagens RGB obtidas por uma câmera digital para detecção de defeito de pele vermelha em presuntos crus. Os histogramas correspondentes à distribuição de frequência dos índices de cores foram utilizados para construção de modelos PLS-DA e WPTER (*Wavelet Packet Transform-based feature selection/classification*) para classificação das amostras de presunto conforme irregularidade.

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 óleos vegetais comestíveis.

2.2 A natureza da luz e o espectro eletromagnético

A luz sensível ao olho humano, *luz visível*, é um tipo de radiação eletromagnética e, como tal, transporta energia pelo espaço exibindo um comportamento ondulatório proveniente das oscilações periódicas de intensidades de forças eletrônicas e eletromagnéticas associadas com a radiação (BROWN et al., 2005). As principais características ondulatórias da luz são a frequência (f) e o comprimento de onda (λ) (BROWN et al., 2005).

A **Figura 2.4** mostra os vários tipos de radiação eletromagnética apresentadas em ordem crescente de comprimento de onda. Esta representação é conhecida como *espectro eletromagnético*.

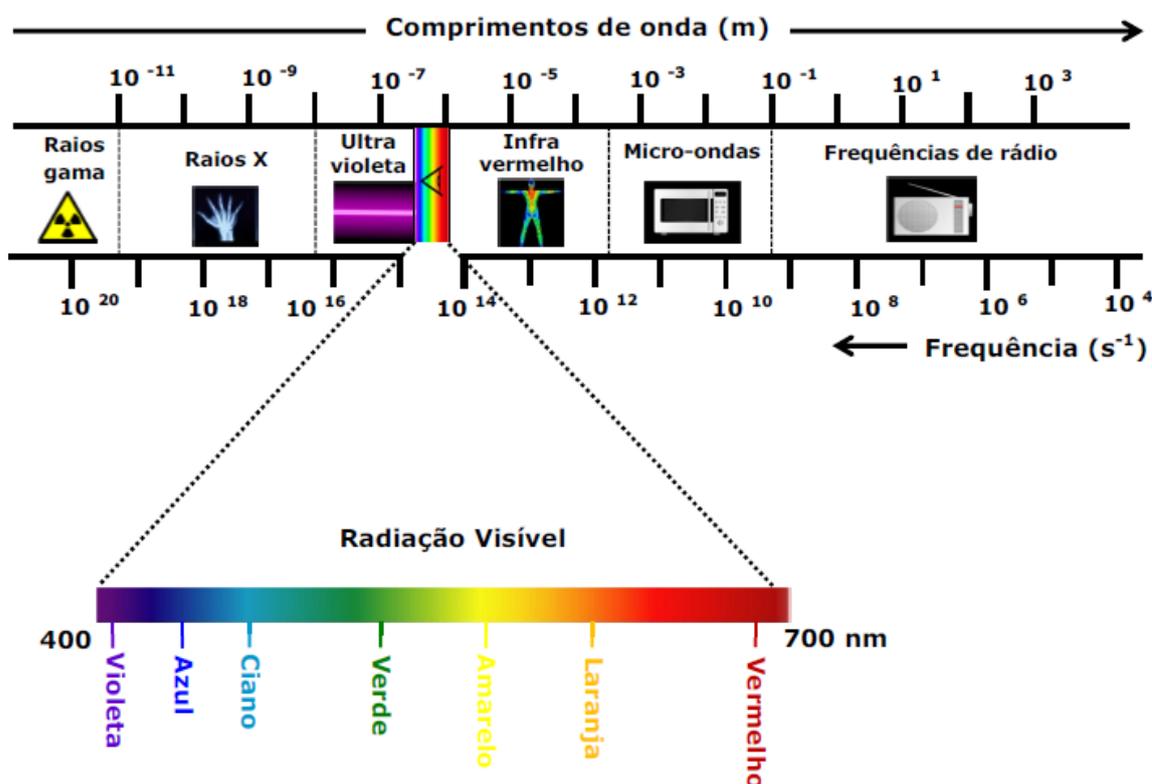


Figura 2.4. Espectro eletromagnético.

Apesar da diversidade de radiações e comprimentos de onda apresentados no espectro eletromagnético, a porção correspondente à luz visível é bastante reduzida, com comprimentos de onda no intervalo de 400 a 700 nm aproximadamente. Quantitativamente, cada cor pode ser expressa pelo comprimento de onda (BROWN et al., 2005). As radiações compostas por um único comprimento de onda são chamadas radiações *monocromáticas*. Entretanto, na maioria dos casos, as fontes de energia radiante emitem um mistura de comprimentos de ondas diferentes (BROWN et al., 2005).

Na **Tabela 2.3** são apresentadas, aproximadamente, as faixas correspondentes às principais cores do espectro.

Tabela 2.3. Principais cores da luz visível e suas faixas do espectro.

Cor	Faixa correspondente no espectro (nm)
Vermelho	625 - 740
Laranja	590 - 625
Amarelo	565 - 590
Verde	500 - 565
Ciano	485 - 500
Azul	440 - 485
Violeta	380 - 440

2.2.1 Interação luz-objeto e seus efeitos na percepção das cores

A natureza da luz refletida por um objeto determina a forma como as cores são percebidas pelo olho humano. A interação entre a radiação luminosa e um objeto pode se dá de diferentes maneiras, o que depende da superfície sobre a qual está ocorrendo a incidência luminosa (FRASER, 2005; GONZALEZ e WOODS, 2000).

Quando a radiação luminosa incide sobre a superfície de um objeto, alguns fenômenos distintos podem ocorrer, dentre os quais: reflexão, transmissão e absorção. Em todos os casos está envolvida a capacidade

do objeto de absorver ou refletir alguns comprimentos de ondas específicos, o que depende das propriedades do material (MELCHIADES e BOSCHI, 1999).

2.2.1.1 Reflexão da luz

A **Figura 2.5** apresenta o fenômeno de reflexão. Quando uma radiação luminosa incide sobre a superfície de um objeto, ocorre uma interação da luz com os átomos ali presentes e, em seguida, a luz reemerge (FRASER, 2005). Durante a interação, o objeto absorve alguns comprimentos de onda e outros são refletidos de tal forma que, a luz refletida difere da luz incidente. Dessa forma, apenas determinadas cores são percebidas. O grau com que alguns comprimentos de onda são refletidos é chamado *reflectância* e representa uma propriedade constante do objeto (FRASER, 2005).

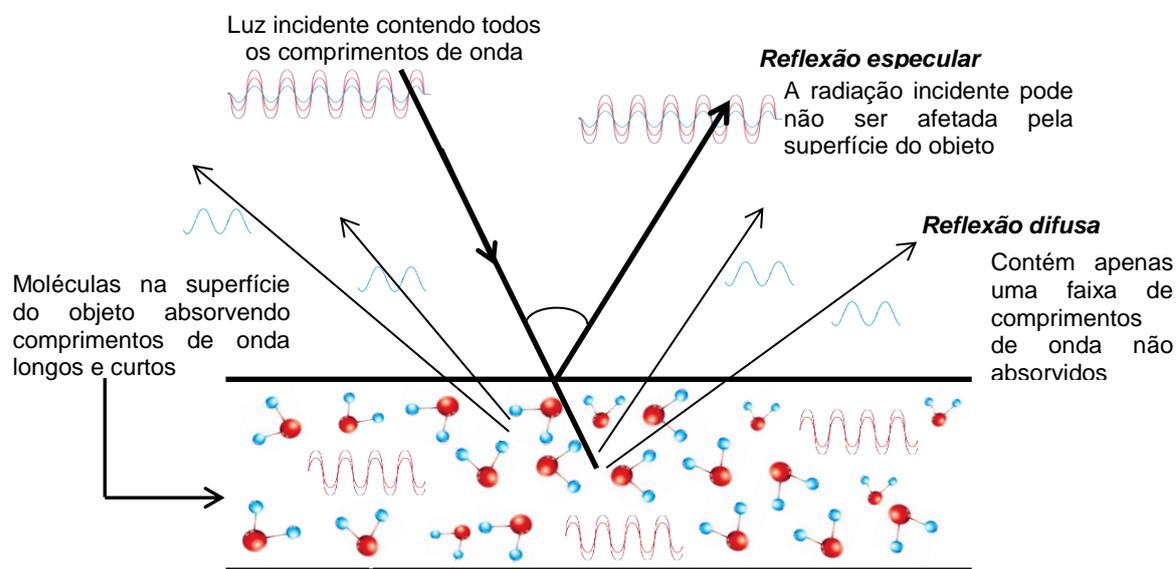


Figura 2.5. Reflexão da luz.
Fonte: Modificada de Fraser (2005).

2.2.1.2 Transmissão da luz

No processo transmissivo, a radiação incidente passa através do objeto e parte dessa radiação pode ou não ser absorvida pelo meio, alterando a composição dos comprimentos de onda da luz (FRASER,

2005). A fração da radiação incidente, com um determinado comprimento de onda, que atravessa um objeto é denominada *transmitância*. Objetos com alto grau de transparência permitem a travessia de toda luz incidente, como mostrado na **Figura 2.6** (LEÃO, 2005).

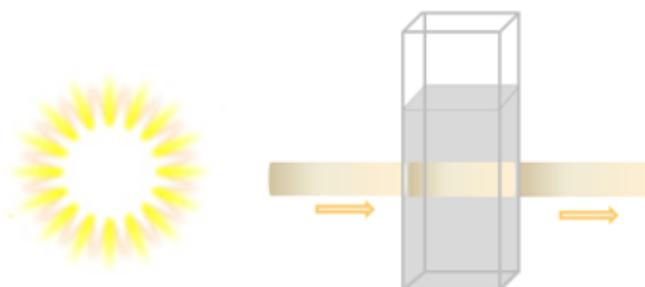


Figura 2.6. Transmissão da luz.

2.3 O sistema visual humano

Para responder a um estímulo, o sistema sensorial humano (visual, auditivo, gustativo, olfativo e táctil) opera integradamente captando sinais que serão posteriormente processadas pelo cérebro (ALEXANDRE e TAVARES, 2007).

Entre os sentidos que respondem pela percepção humana, a visão é o que apresenta maior capacidade de identificar e distinguir os objetos a sua volta. Além disso, seus sinais atingem o cérebro a uma velocidade bem maior que os demais sistemas, fazendo com que as suas informações sejam processadas mais rapidamente (ALEXANDRE e TAVARES, 2007).

A percepção das cores está relacionada a fenômenos fisiopsicológicos e é profundamente influenciada pela natureza da radiação e os processos desencadeados no olho humano (GONZALEZ e WOODS, 2000).

A **Figura 2.7** mostra o olho humano a partir de um corte vertical. Para que a imagem de um determinado objeto seja formada, a radiação luminosa incidente sobre o mesmo deve penetrar no olho a partir de uma

abertura frontal, a íris, atravessar o cristalino e seguir em direção à retina (BATISTA, 2004).

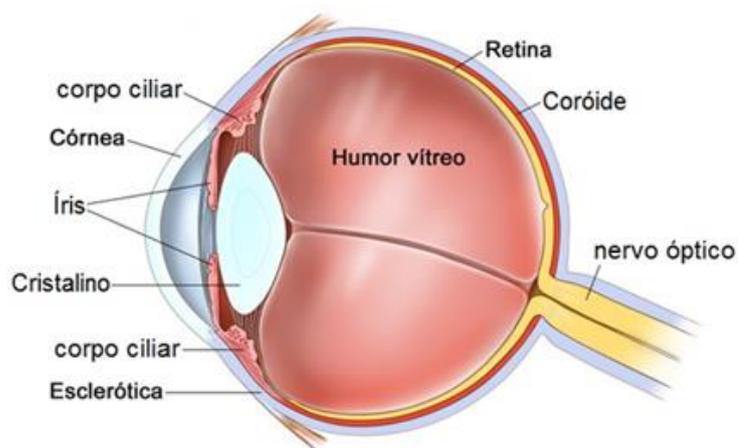


Figura 2.7. Representação do olho humano.
Fonte: Adaptada de <http://www.multiclin.com.br/samba/>
Acessado em: abril de 2013

A íris controla a quantidade de luz que penetra no olho. Quando uma cena é focalizada pelo olho, a imagem correspondente é projetada sobre a retina, onde estão distribuídos dois tipos de receptores fotossensíveis: os *cones* e os *bastonetes* (**Fig. 2.8**). Estes receptores têm a função de converter energia luminosa em sinais elétricos que serão interpretados pelo cérebro, fornecendo características como cor, forma e distância (MARQUES FILHO e VIEIRA NETO, 1999; BATISTA, 2004; ALEXANDRE e TAVARES, 2007).

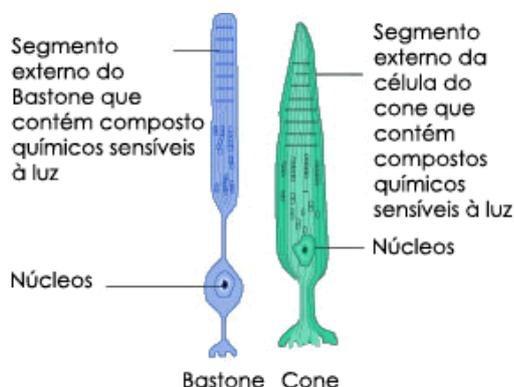


Figura 2.8. Representação de um bastonete e de um cone.
Fonte: <http://saude.hsw.uol.com.br/visao2.htm>
Acessado em: abril de 2013

Os bastonetes, cerca de 75 a 150 milhões em cada olho, são sensíveis apenas a baixos níveis de iluminação, não são capazes de distinguir as cores e por isso são chamados de *visão de baixa luminosidade* (BATISTA, 2004).

Os cones, embora presentes em menor quantidade, 6 a 7 milhões em cada olho, são sensíveis as cores e possibilitam alta resolução de detalhes, proporcionando a *visão de alta luminosidade*. Estes sensores são sensíveis a basicamente três faixas de comprimentos de onda predominantes: vermelho, verde e azul, as cores primárias do espectro, o que lhe atribuiu às denominações cones vermelho, verde e azul. As diferentes cores percebidas pelo olho humano são obtidas a partir da combinação dessas três cores primárias (BATISTA, 2004).

2.4 Modelos de cores

Os modelos de cores têm como objetivo principal a especificação das cores em uma forma padrão. Neste sentido, eles são direcionados de acordo com sua aplicação, seja em dispositivos de impressão, captura (*scanners* e máquinas digitais) ou visualização (monitores e projetores) de imagens (GONZALEZ e WOODS, 2000).

Em geral, um modelo de cor pode ser representado por um sistema de coordenadas tridimensionais e um subespaço dentro deste sistema onde cada ponto das coordenadas espaciais corresponde a uma determinada cor (GONZALEZ e WOODS, 2000).

Os modelos mais comumente utilizados são o RGB (R – *Red*, G – *Green* e B – *Blue*), o CMY (C – *Cyan*, M – *Magenta* e Y – *Yellow*) e o HSI (H – *Hue*, S – *Saturation* e I – *Intensity*). Nas próximas seções serão apresentados apenas os modelos RGB e HSI, os quais foram empregados neste trabalho.

2.4.1 Modelo RGB de cores

O modelo RGB baseia-se no mecanismo de formação das cores no olho humano, onde combinações, em diferentes níveis, das radiações luminosas nos comprimentos de onda vermelho, verde e azul fornecem radiações de diferentes cores (BATISTA, 2004) (**Fig. 2.9**). Neste mecanismo estão envolvidos fenômenos de absorção e reflexão da luz, uma vez que as cores percebidas pelos olhos humanos num objeto estão relacionadas à natureza da luz refletida pelo mesmo (GONZALEZ e WOODS, 2000).

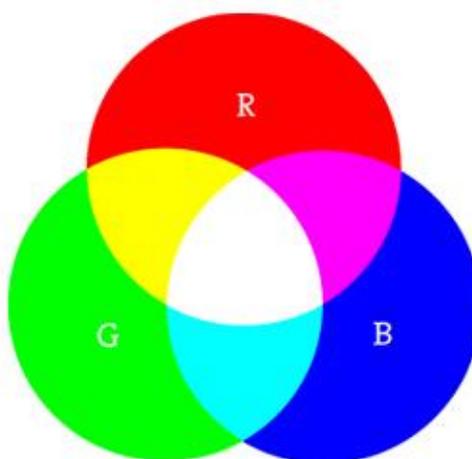


Figura 2.9. Cores primárias aditivas e as cores resultantes de suas combinações.

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

Acesso em: abril de 2013

Corpos capazes de refletir a luz são percebidos na cor branca quando ocorre uma combinação equilibrada de todos os comprimentos de onda visíveis do espectro. Para corpos com reflectância limitada, apenas alguns tons de cores são exibidos. Por exemplo, um objeto de cor amarela reflete radiações nas faixas do verde e vermelho do espectro e absorve radiação azul, que é então chamada *cor oposta* ao amarelo. Se não ocorre reflexão de luz, a cor preta é percebida (GONZALEZ e WOODS, 2000).

O modelo RGB pode ser representado por um sistema de coordenadas cartesianas, a partir do qual, é gerado um cubo sólido conhecido como cubo de cores, apresentado na **Figura 2.10**.

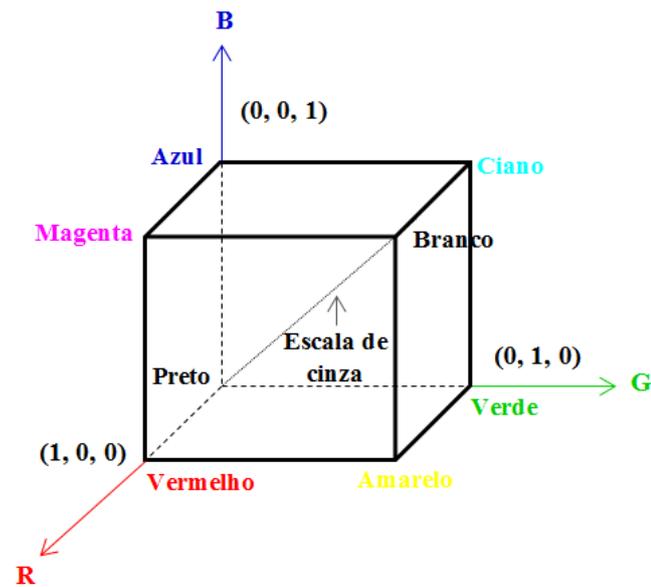


Figura 2.10. Cubo de cores RGB.
Fonte: Adaptada de Gonzalez e Woods (2000).

Como pode ser observado, a combinação de duas cores primárias, em igual proporção, resulta na formação das cores secundárias, tais como: magenta (Vermelho + Azul), ciano (Azul + Verde) e amarelo (Vermelho + Verde). A cor preta, localizada na origem do sistema, indica que não há contribuição das cores primárias enquanto que a cor branca revela a combinação equilibrada destas cores. No modelo RGB cada pixel pode assumir valores de intensidade variando de 0 a 255 para cada componente vermelho, verde e azul.

O nível de cinza é indicado pela linha diagonal estendida da cor preta até a cor branca. Segundo SOLOMON e BRECKON (2011, p.12) “a imagem em escala de cinza é essencialmente uma soma ponderada dos canais de cores vermelho, verde e azul”. Matematicamente, a conversão de uma imagem RGB em escala de cinza é obtida a partir da transformação:

$$I_{\text{escala de cinza}}(n,m) = \alpha I_{\text{cor}}(n,m,r) + \beta I_{\text{cor}}(n,m,g) + \gamma I_{\text{cor}}(n,m,b) \quad (2.1)$$

Onde:

- (n, m) corresponde a localização de um pixel individual na imagem em escala de cinza;
- (n, m, c) corresponde a localização de um pixel (n, m) na imagem colorida em relação a cada componente de cor c (**r**ed, **g**reen, **b**lue) na imagem.

Os coeficientes de ponderação (α, β, γ) são atribuídos considerando-se a resposta de percepção do olho humano para as componentes de cores vermelho, verde e azul com um peso padronizado garantindo a uniformidade $(\alpha = 0.2989, \beta = 0.5870, \gamma = 0.1140)$ (SOLOMON e BRECKON, 2011).

2.4.2 Modelo HSI de cores

No modelo HSI, as cores são descritas a partir de três atributos fundamentais: matiz, saturação e intensidade luminosa. Matiz (ou tonalidade) de uma cor corresponde ao comprimento de onda dominante em uma mistura de radiações luminosas, ou seja, a cor realmente percebida pelo observador. A saturação expressa a pureza relativa da cor, em relação à quantidade de luz branca adicionada. O grau de saturação é inversamente proporcional à quantidade de luz branca adicionada. Com isso, pode-se dizer que as cores puras do espectro são completamente saturadas (GONZALEZ e WOOD, 2000; BATISTA, 2004).

Seguindo os princípios da relação existente entre a cor e a natureza da luz, pode-se considerar que a luz acromática (sem cores) será caracterizada apenas com base nos seus valores de intensidade luminosa, também chamada brilho ou apenas luminosidade. No caso da luz cromática, a caracterização da cor é realizada a partir da junção dos três atributos (GONZALEZ e WOOD, 2000).

O modelo HSI pode ser representado a partir do sólido mostrado na **Figura 2.11a**. Se cortado horizontalmente, esse sólido fornece uma figura geométrica semelhante à apresentada na **Figura 2.11b**, nas quais estão dispostas as cores primárias do espectro e as demais cores

decorrentes de suas combinações. A linha perpendicular que passa pelo centro do sólido na **Figura 2.11a** é utilizada para medir os valores de intensidade luminosa das combinações de cores, que podem ser claras ou escuras, conforme a proximidade do corte em relação ao vértice branco ou preto, respectivamente (MARQUES FILHO e VIEIRA NETO, 1999).

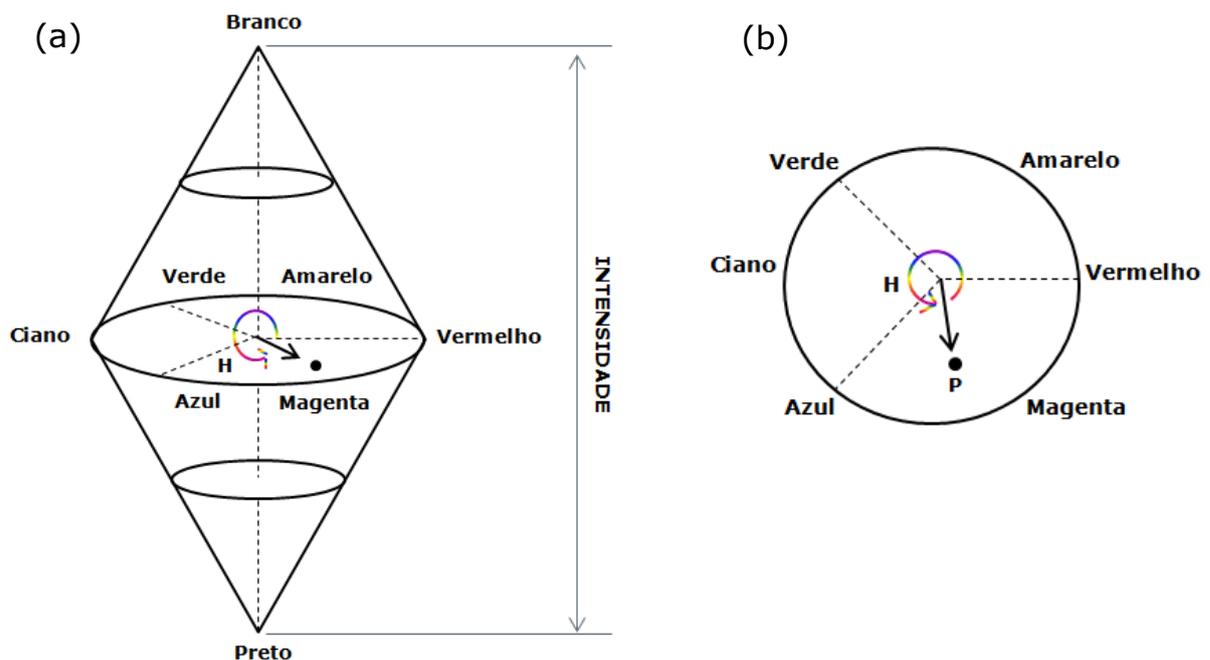


Figura 2.11. Modelo HSI de cores. (a) Sólido de cores, (b) Triângulo de cores.
Fonte: Modificada de Marques Filho e Vieira Neto (1999).

Os atributos matiz e saturação podem ser representados a partir do espectro de cores da **Figura 2.12**. Matiz pode ser entendida como o posicionamento de uma cor no espectro. Considerando-se um movimento circular em volta desse espectro e tendo-se o ângulo 0° como ponto de partida, à medida que ocorre um distanciamento do 0° , seja no sentido horário ou anti-horário, diferentes ângulos e tonalidades são obtidas, ou seja, diferentes matizes. Em contrapartida, a saturação é definida pela distância da região central às bordas da **Figura 2.12** (BATISTA, 2004).

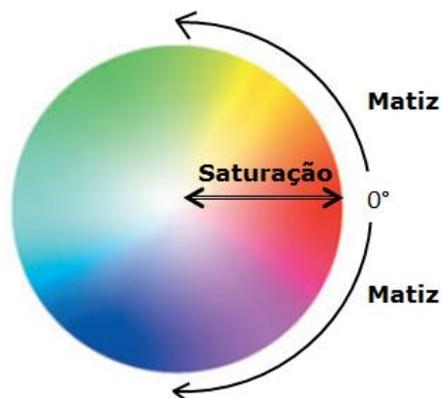


Figura 2.12. Representação dos atributos da cor.

2.5 Das imagens analógicas às imagens digitais

Uma imagem pode ser considerada a representação analógica (ou eletrônica) de uma cena real tridimensional. O processo de conversão de uma cena real em uma imagem analógica, processo chamado transdução optoeletrônica, envolve a redução de dimensionalidade da cena e normalmente é realizado por câmeras CCD (*Charge Coupled Device* - Dispositivo Acoplado por Carga). Estes dispositivos são configurados para armazenar a energia luminosa incidente sobre um determinado objeto e o sinal elétrico conseqüentemente produzido. Este sinal elétrico é condicionado por circuitos eletrônicos especializados gerando como saída um Sinal Composto de Vídeo (SCV) analógico e monocromático, em outras palavras, uma imagem monocromática (MARQUES FILHO e VIEIRA NETO, 1999).

Uma imagem monocromática (**Fig. 2.13**) é descrita matematicamente como sendo uma função bidimensional $f(x, y)$ da intensidade luminosa, que expressa o produto da interação entre a iluminância - quantidade de luz incidente sobre o objeto - e a reflectância - quantidade de luz refletida pelo objeto, que está relacionada às propriedades da superfície do mesmo. O valor da função $f(x, y)$, em qualquer ponto de coordenadas espaciais (x, y) , é proporcional ao brilho (ou nível de cinza) naquele ponto (MARQUES FILHO e VIEIRA NETO, 1999). Matematicamente:

$$f(x,y) = i(x,y) \cdot r(x,y) \quad (2.2)$$

Onde:

- $f(x, y)$ é uma função bidimensional contínua que representa a imagem;
- $i(x, y)$ representa a iluminância, sendo $0 < i(x, y) < \infty$
- $r(x, y)$ representa a reflectância, sendo $0 < r(x, y) < 1$



Figura 2.13. Imagem monocromática. Os níveis de cinza correspondem à intensidade luminosa da imagem f em cada coordenada espacial (x,y) que compõem a imagem.

No caso das imagens coloridas no sistema RGB, existem informações em intervalos ou bandas distintas de frequência, sendo necessária uma função $f(x, y)$ para cada banda, que por sua vez, refere-se às cores primárias aditivas: vermelho (R – *Red*), verde (G – *Green*) e azul (B – *Blue*). Com isso, para gerar uma imagem colorida, as câmeras CCDs devem apresentar, em sua configuração, um conjunto de prismas e filtros de cor que fazem a decomposição da imagem colorida em suas componentes RGB (MARQUES FILHO e VIEIRA NETO, 1999).

2.5.1 Imagens digitais

As imagens analógicas capturadas por dispositivos eletrônicos, monocromáticas ou coloridas, não podem ser processadas por computadores, a não ser em termos de matrizes de números digitais. O

processo de conversão das imagens analógicas em imagens digitais consiste na criação de uma matriz M por N pontos, onde cada ponto é denominado *pixel* ou *elemento de imagem* (QUEIROZ e GOMES, 2001, MARQUES FILHO e VIEIRA NETO, 1999). A **Figura 2.14** expressa a representação matricial, ou digital, de uma imagem analógica monocromática. Como pode ser visto, uma imagem digital pode ser definida como uma matriz, onde no caso das imagens monocromáticas, cada ponto representa o nível de cinza da imagem naquele ponto. Uma imagem de alta resolução apresenta altos valores de M e N , ou seja, quanto maior o número de pixels, maior é a qualidade da imagem (MARQUES FILHO e VIEIRA NETO, 1999).

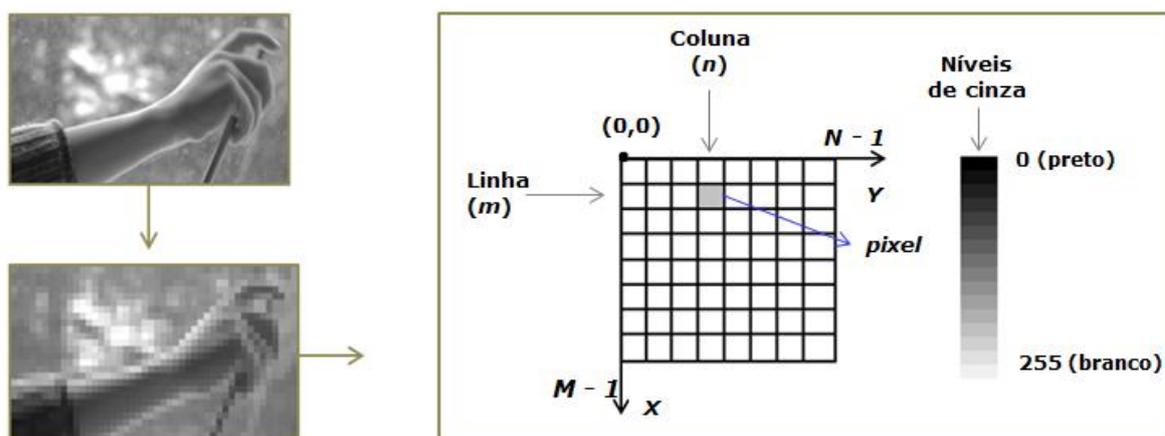


Figura 2.14. Imagem monocromática e sua representação em imagem digital.
Fonte: Adaptada de Queiroz e Gomes (2001).

Nas imagens digitais coloridas no sistema RGB, cada pixel representa um vetor onde as componentes correspondem às intensidades (ou bandas) das cores vermelho, verde e azul da imagem (BATISTA, 2004). Matematicamente:

$$f(x,y) = f_R(x,y) + f_G(x,y) + f_B(x,y) \quad (2.3)$$

onde, $f_R(x, y)$, $f_G(x, y)$ e $f_B(x, y)$ representam, respectivamente, as intensidades luminosas das componentes vermelha, verde e azul da

imagem, no ponto (x, y) . A **Figura 2.15** apresenta as três bandas de uma imagem, separadamente, e a imagem colorida resultante da composição dessas três bandas.

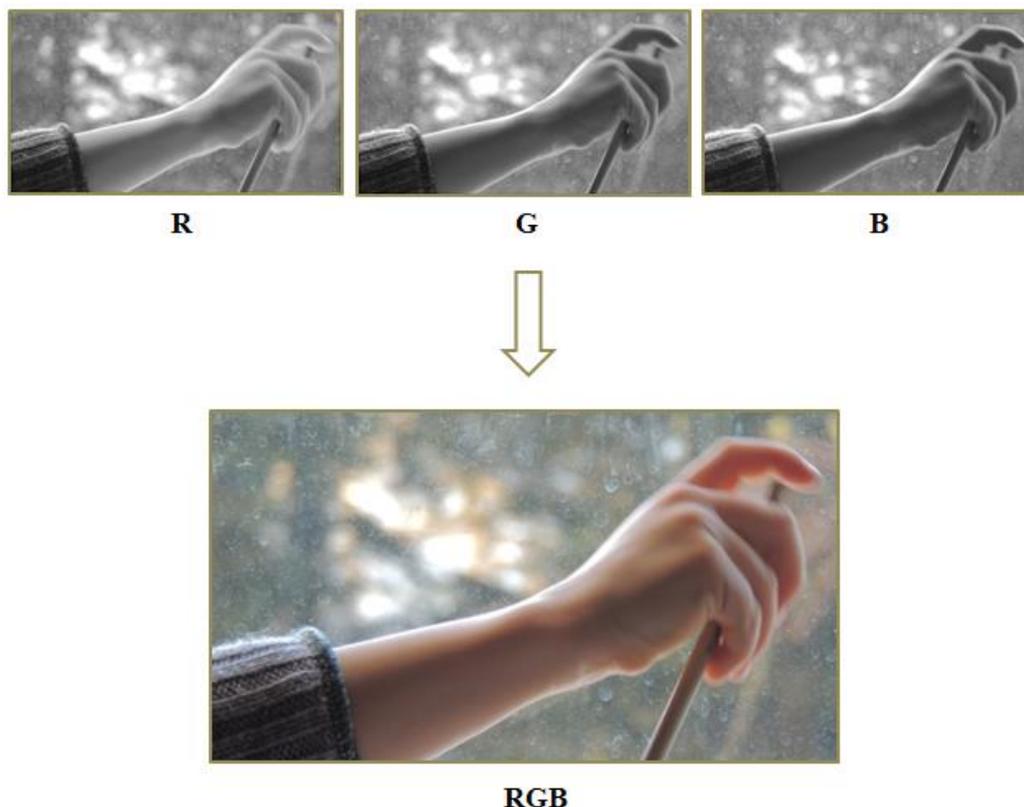


Figura 2.15. As três bandas de uma imagem, separadamente, e a imagem colorida resultante da composição dessas três bandas.

2.5.1.1 Pixel

Em uma imagem digitalizada, cada pixel apresenta 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 (SOLOMON e BRECKON, 2011; MARQUES FILHO e VIEIRA NETO, 1999).

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 x 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.

2.5.1.2 Histogramas

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 apresenta 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, SOLOMON e BRECKON, 2011, QUEIROZ e GOMES, 2001).

Segundo Solomon e Breckon (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 2.16** é apresentada uma imagem colorida e os histogramas resultantes de sua decomposição nas componentes de cores cinza, vermelho, verde e azul.

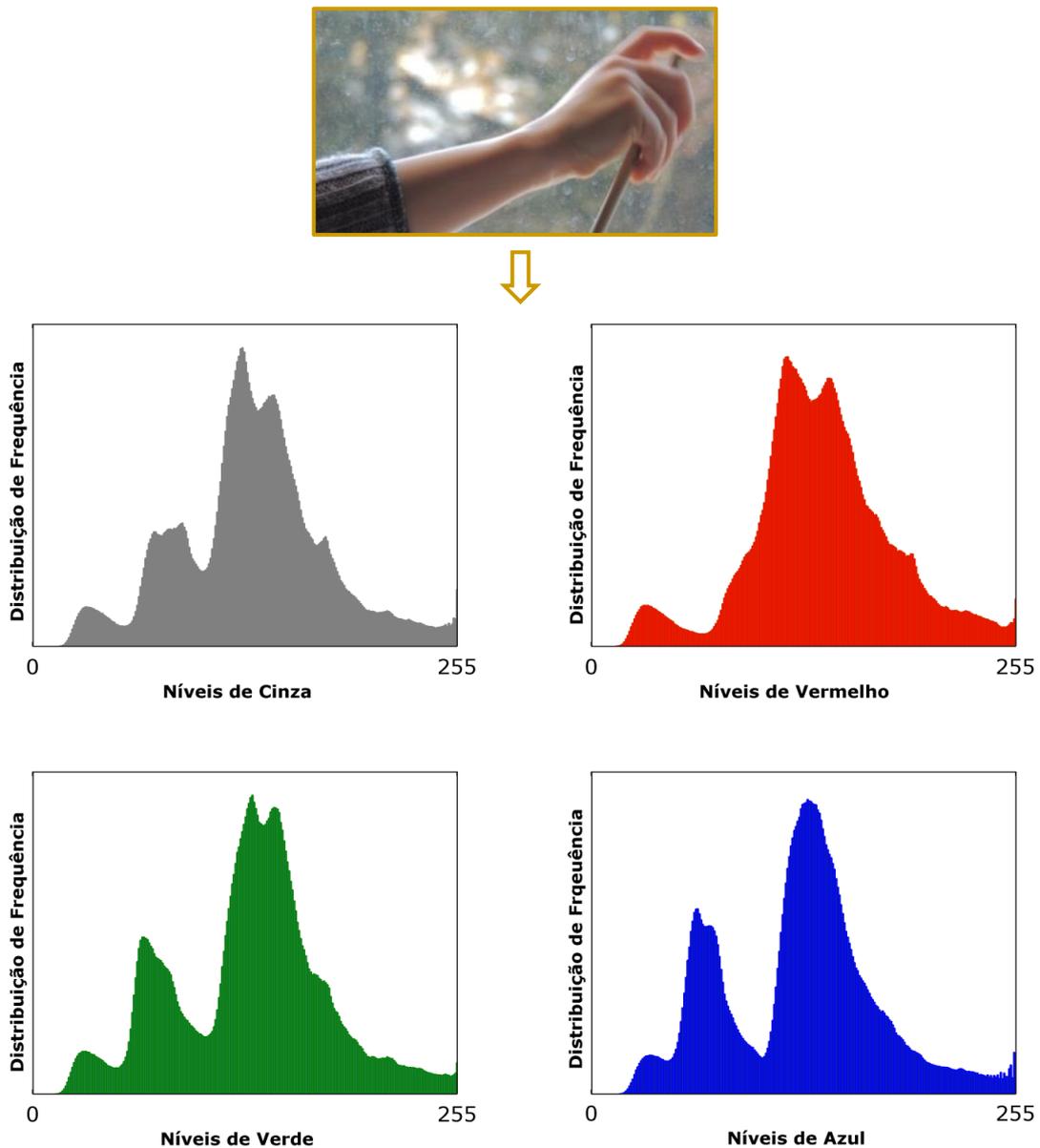


Figura 2.16. Histogramas referentes a distribuição de frequência de todos os possíveis valores de um pixel nos níveis de cinza, vermelho, verde e azul de uma imagem.

Além de representar os níveis de cores para os tons de cinza e as componentes RGB, como mostrado na **Figura 2.16**, os histogramas correspondentes aos atributos da cor (componentes HSI) também podem ser obtidos.

2.6 Ferramentas quimiométricas

A Quimiometria pode ser entendida como um conjunto de procedimentos matemáticos e estatísticos destinados ao planejamento de experimentos e a obtenção do máximo de informação de dados provenientes de análises químicas (OTTO, 1999).

A utilização de ferramentas quimiométricas frente aos dados de natureza multivariada foi impulsionada, sobretudo, pelo desenvolvimento de instrumentos analíticos modernos capazes de analisar mais de um parâmetro por amostra e com isso fornecer quantidades consideráveis de informação, em reduzidos intervalos de tempo (CHAU, LIANG, GAO e SHAO, 2004). Soma-se a isso, a disponibilidade de computadores e programas computacionais que facilitam a aquisição, manipulação e interpretação dos dados (OTTO, 1999; VICINI, 2005).

Na Química Analítica, as análises realizadas com propósito de controle de qualidade são conduzidas através do monitoramento de algum parâmetro da amostra, como por exemplo: índice de acidez, condutância, pH, cor, entre outros. Basicamente, muitas dessas análises consistem em classificar a amostra em categorias, de acordo com seus padrões de conformidade. Nesse contexto, o emprego de técnicas de reconhecimento de padrões é imprescindível. Essas técnicas são comumente divididas em não supervisionadas e supervisionadas (BEEBE, 1998). Uma abordagem teórica de ambas é apresentada nas próximas seções.

2.6.1 Técnicas de reconhecimento de padrões não supervisionadas

As técnicas de reconhecimento de padrões não supervisionadas são utilizadas com intuito de encontrar similaridades e diferenças em conjuntos de amostras com características variadas. Para isso, utilizam-se apenas as informações provenientes de alguma técnica analítica (BEEBE, 1998).

Na literatura, duas técnicas de reconhecimento de padrões não supervisionados vêm se destacando no contexto da Química Analítica: análise de componentes principais (PCA: *Principal Component Analysis*) e

a análise de agrupamentos hierárquicos (HCA: *Hierarchical Cluster Analysis*). Neste trabalho, apenas a PCA foi avaliada.

2.6.1.1 Análise de Componentes Principais (PCA)

Após a obtenção dos dados, os mesmos são organizados em uma matriz de dimensão $n \times m$, onde as linhas (n) representam as amostras e as colunas (m) representam as variáveis correspondentes aos valores dos sinais instrumentais. Na PCA, um conjunto de dados com um número elevado de variáveis pode ser observado com clareza a partir de fatores que apresentam os dados em menores dimensões (BEEBE, 1998).

A PCA consiste na descrição da variação do conjunto de dados através de componentes principais (PCs), ou fatores, resultantes da decomposição da matriz de dados em um novo sistema de coordenadas, formado por eixos ortogonais e independentes entre si, obtidos em ordem decrescente de variância (OTTO, 1999; BEEBE, 1998; MOITA NETO e MOITA, 1998). À medida que a variância decresce, diminui o percentual de informação contido nas PCs e, conseqüentemente, menor será o número de PCs necessárias para representar os dados, obtendo-se a redução da dimensionalidade (BEEBE, 1998).

As PCs são construídas a partir de novas coordenadas das amostras, os escores, e de um conjunto de variáveis obtido pela combinação linear das variáveis originais (BEEBE, 1998) (**Fig. 2.17**). Nesse conjunto, apenas as variáveis realmente representativas estão presentes - o que funciona como uma espécie de filtro para as informações redundantes ou de caráter ruidoso - e os coeficientes da combinação linear, os pesos, correspondem a uma medida da contribuição de cada variável original (BEEBE, 1998).

A PCA pode ser utilizada para (MOITA NETO e MOITA, 1998):

- Observar agrupamentos no conjunto de dados;
- Identificar amostras anômalas (*outliers*);
- Verificar a contribuição das variáveis originais no novo conjunto de variáveis obtido.

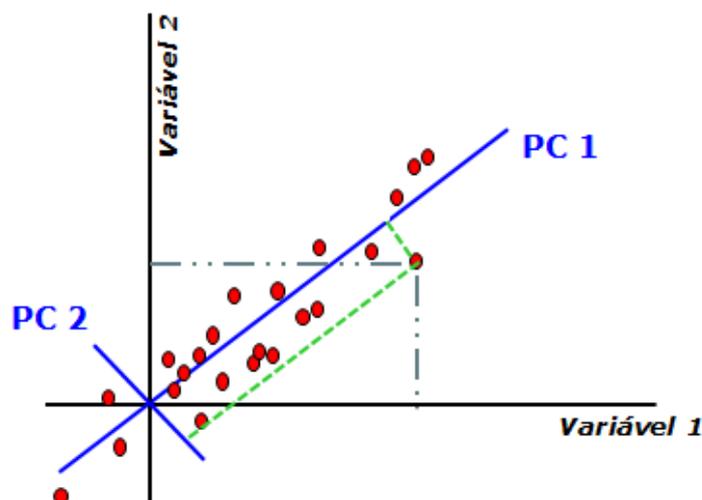


Figura 2.17. Primeira e segunda PCs obtidas a partir da combinação linear das variáveis originais. As linhas tracejadas na cor verde correspondem aos escores e as linhas tracejadas-pontilhadas na cor cinza representam os pesos.

Fonte: Adaptada de BEEBE (1998).

2.6.2 Técnicas de reconhecimento de padrões supervisionadas

Diferentemente das técnicas de reconhecimento de padrões não supervisionadas, as supervisionadas utilizam uma informação adicional sobre os membros (objetos) das classes. Assim, a resposta instrumental é relacionada com os índices de classes e modelos multivariados são obtidos e validados com intuito de classificar amostras de identidades desconhecidas (BEEBE, 1998; CASTRO e PRADO, 2002).

Diversos métodos de reconhecimento de padrões supervisionados têm sido utilizados em problemas analíticos de classificação. Destacam-se: modelagem independente e flexível por analogia de classes (*SIMCA: Soft Independent Modeling of Class Analogy*) análise discriminante linear (LDA: *Linear Discriminant Analysis*) e análise discriminante pelos mínimos quadrados parciais (PLS-DA: *Partial Least Squares-Discriminant Analysis*). Neste trabalho, apenas a LDA e a PLS-DA foram adotadas.

2.6.2.1 Análise Discriminante Linear (LDA)

A LDA baseia-se no cálculo de hiperplanos, ou funções discriminantes (FD) lineares, que descrevem os dados objetivando a máxima variância entre as classes e a mínima variância entre os indivíduos de uma mesma classe (OTTO, 1999). Para isso, as FD lineares são obtidas a partir de combinações lineares das variáveis originais que melhor discriminem as classes ou grupos de objetos (PONTES, 2009).

A **Figura 2.18** apresenta o esquema de obtenção de uma FD linear e como opera a classificação de um elemento.

Os centros de gravidade 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, 2008, p.22).

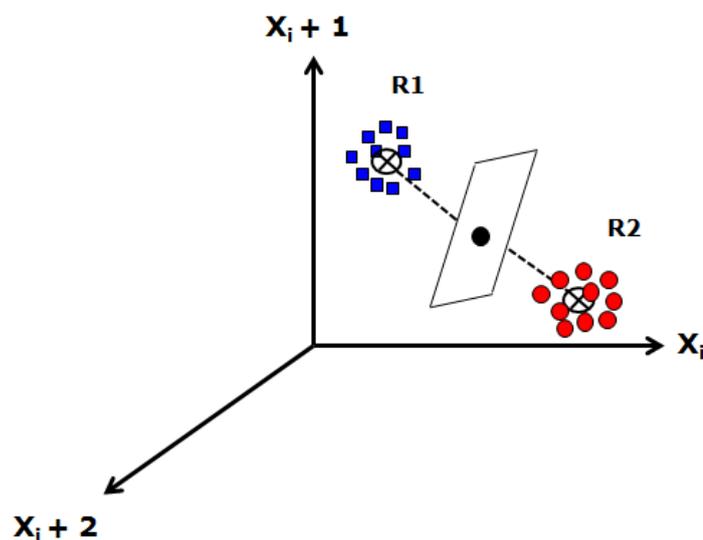


Figura 2.18. Representação esquemática da análise discriminante linear.
Fonte: Adaptada de Gambarra-Neto (2008).

É importante destacar que a LDA é bastante afetada por problemas de colinearidade, sendo frequentemente empregada em dados de

pequenas dimensões. Assim, procedimentos baseados em seleção ou redução de variáveis são constantemente requeridos aos modelos LDA (PONTES, 2009).

No que diz respeito às estratégias de seleção de variáveis, alguns métodos vêm se destacando no âmbito da classificação: Algoritmo das Projeções Sucessivas (SPA) (PONTES et al., 2005), Algoritmo Genético (GA) (LEARDI, 2007) e Stepwise (SW) (CANECA et al., 2006). Neste trabalho, apenas o SPA e SW foram empregados.

2.6.2.2 Algoritmo das Projeções Sucessivas (SPA)

O SPA é uma técnica de seleção de variáveis baseada numa série de procedimentos matemáticos onde as N variáveis existentes no conjunto de dados são projetadas sucessivamente, de forma a obter cadeias com número reduzido de variáveis onde cada elemento (variável) é selecionado de modo a exibir mínima colinearidade com o elemento anterior. A escolha da melhor cadeia é realizada com base na função de custo G , que corresponde ao risco médio de uma classificação incorreta realizada pelas variáveis selecionadas no modelo LDA (PONTES, 2009). Esta função é calculada usualmente com base no conjunto de validação (como mostrado na **Equação 2.4**). Contudo, em um recente trabalho publicado por Pontes et al (2012), o cálculo da função de custo foi realizado no domínio das amostras de treinamento. Tal procedimento é normalmente requerido quando o número de amostras empregado no estudo é, de certo modo, reduzido.

$$G = \frac{1}{K_v} \sum_{k=1}^{K_v} g_k \quad (2.4)$$

Na **Equação 2.4**, K_v é o número de amostras do conjunto de validação e g_k representa o risco de uma classificação incorreta do objeto \mathbf{x}_k da k -ésima amostra de validação, sendo definido como:

$$g_k = \frac{r^2(x_k, \mu_{IK})}{\min_{Ij \neq IK} r^2(x_k, \mu_{IK})} \quad (2.5)$$

Na **Equação 2.5**, 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 previsão são utilizados, o cálculo da função de custo, J_{cost} , é realizado como mostrado na **Equação 2.6**:

$$J_{cost} = \frac{1}{N_{train} - L - C} \sum_{n=1}^{N_{train}} g_n \quad (2.6)$$

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

2.6.2.3 Stepwise (SW)

O Algoritmo Stepwise é outro método de seleção de variáveis que vem sendo também aplicados em problemas de classificação (SILVA et al., 2012). Nele, a capacidade de discriminação de cada variável é investigada individualmente, de forma que todas as variáveis são excluídas ou adicionadas ao modelo de acordo com o método de validação cruzada *leave-one-out* (PONTES, 2009).

O fator de discriminação (D_i) de uma variável n_i baseia-se nas medidas de dispersão dentro das classes e entre as classes (como

mostrado na **Equação 2.7**) e é utilizado como parâmetro de seleção das variáveis (PONTES, 2009).

$$D_i = \frac{SB_i}{SW_i} \quad (2.7)$$

O valor de SW_i é calculado a partir da **Equação 2.8a**, onde s_{ij} é a dispersão de n_i na classe j , expressa na **Equação 2.8b**:

$$(a) \quad SW_i = \sum_{j=1}^c s_{ij} \quad (b) \quad s_{ij} = \sum [x_i^k - m_{ij}]^2 \quad (2.8)$$

Onde x_i^k é o valor n_i na k -ésima amostra e m_{ij} é o valor médio de n_i na classe j :

$$m_{ij} = \frac{1}{n_j} \sum_{k \in I_j} x_i^k \quad (2.9)$$

A dispersão entre as classes, SB_i , é definida como:

$$SB_i = \sum_{j=1}^c n_j [m_{ij} - m_i]^2 \quad (2.10)$$

Dessa forma, a variável n_i com o valor de D_i mais elevado é selecionada e o número de erros obtidos pela validação cruzada é registrado. Após terem sido calculados os valores de D_i para todas as variáveis, o grupo que apresentar o menor número de erros de validação cruzada, sem problemas de colinearidade, é então selecionado (PONTES, 2009).

2.6.2.4 Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA)

O PLS-DA baseia-se no método de 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 (PEERBHAY et al., 2013; ALMEIDA et al., 2013; PONTES et al., 2011). 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 (SILVA et al., 2012) (**Fig. 2.19**).

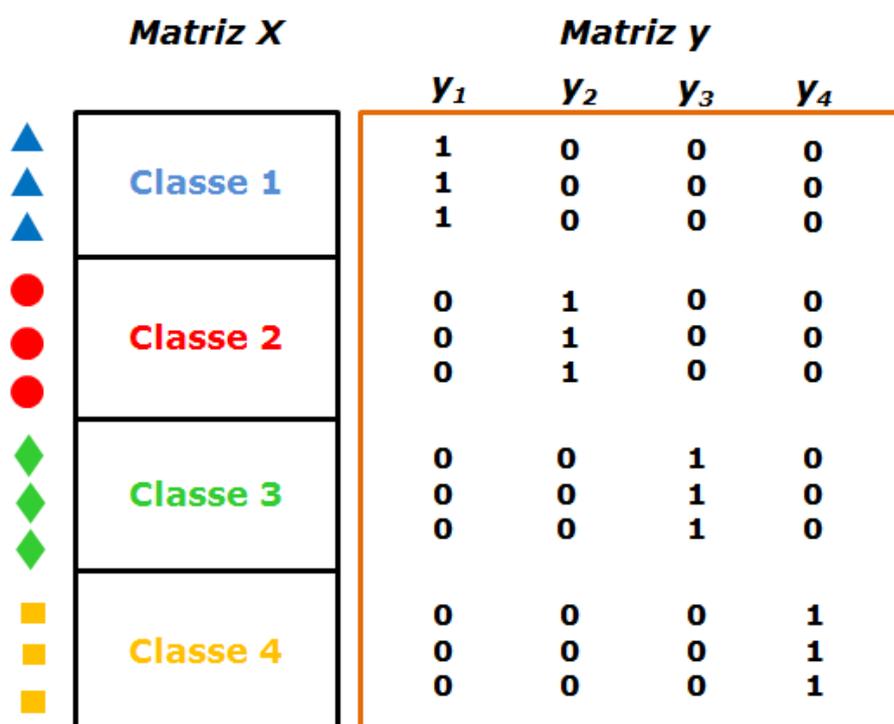
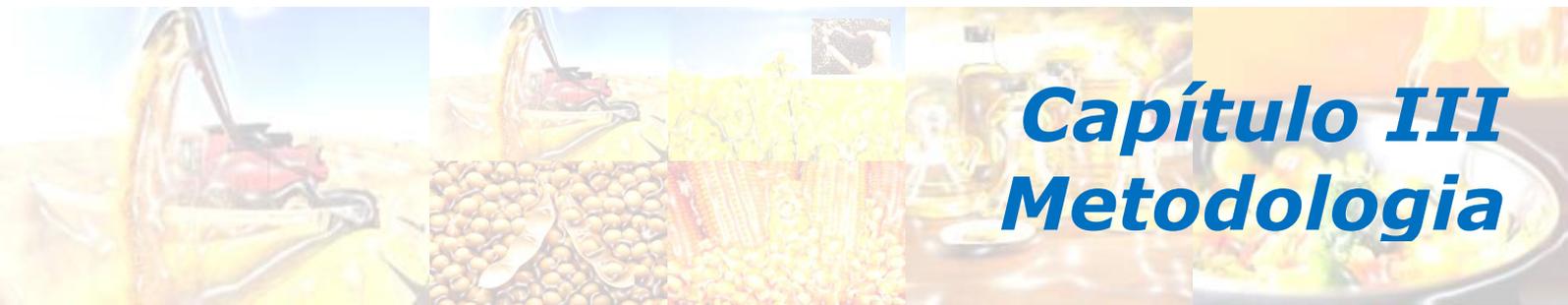


Figura 2.19. Representação do modelo preditivo PLS-DA.

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).

A PLS-DA tem sido aplicada com sucesso em alguns problemas de classificação. Pontes et al. (2011) utilizaram PLS-DA e LDA, com seleção

de variáveis, para detectar adulteração em blendas de diesel/biodiesel e Silva et al. (2012) para detectar adulteração em álcool etílico hidratado.



Capítulo III

Metodología

3 METODOLOGIA

3.1 Amostras

Cento e trinta e duas amostras de óleos vegetais comestíveis dos tipos canola, milho, girassol e soja (**seção 2.1.2**) foram analisadas. Dessas amostras, 60 foram adquiridas nos supermercados da grande João Pessoa estando dentro do prazo de validade e, para garantir a variabilidade dos dados, amostras de diferentes lotes e fabricantes foram selecionadas. As amostras remanescentes (72) foram cedidas pela Universidade Federal Rural de Pernambuco e estavam com o prazo de validade expirado por um período de um mês a dois anos. As amostras não foram submetidas a nenhum tipo de processamento.

3.2 Sistema para aquisição das imagens

A **Figura 3.1** apresenta o sistema empregado para aquisição das imagens. Como pode ser observado, é um sistema relativamente simples e exige um número reduzido de componentes: uma caixa, onde é realizada a captura das imagens; uma *webcam*, uma cubeta de fluxo, uma bomba peristáltica, uma lâmpada fluorescente e um notebook.

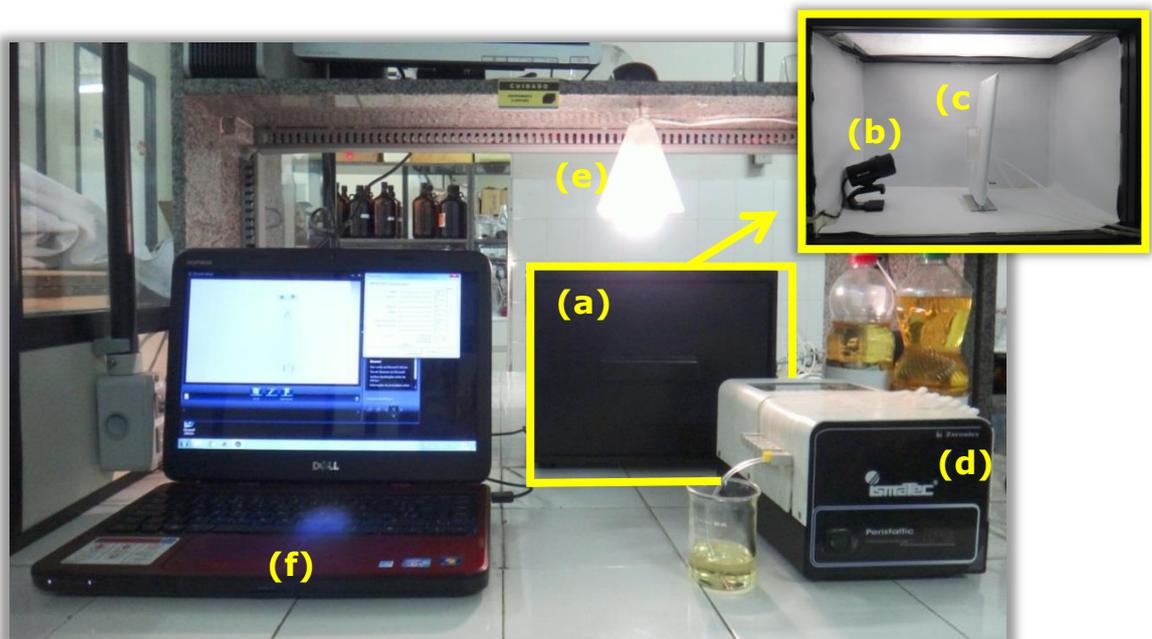


Figura 3.1. Sistema utilizado na aquisição das imagens das amostras de óleos vegetais comestíveis. (a) caixa; (b) webcam; (c) cubeta de fluxo; (d) bomba peristáltica (e) lâmpada fluorescente e (f) notebook.

3.3 Compartimento

Na **Figura 3.2** é apresentado o interior do compartimento empregado para aquisição das imagens. Uma caixa com dimensões de 30 cm × 22 cm × 23 cm foi construída em madeira, com o intuito de isolar as amostras com relação à luz externa e assim facilitar o controle da luz incidente sobre as mesmas, garantindo a qualidade e principalmente, a uniformidade das imagens registradas. Para essa finalidade, a caixa foi revestida internamente com folha de papel ofício de cor branca, seguindo os princípios ópticos da reflexão da luz, para que todas as cores presentes na imagem fossem devidamente detectadas pela *webcam*. No interior do compartimento foram colocadas uma cubeta de fluxo de quartzo de 1 cm de caminho óptico e uma *webcam* Microsoft com resolução HD (1280 × 720), modelo LifeCam Cinema, dispostas estrategicamente para aquisição das imagens.



Figura 3.2. Imagem interna do compartimento.

Para encontrar a melhor disposição dos componentes internos, de modo a se obter imagens uniformes quanto à quantidade de luz incidente ao longo da cubeta, vários testes foram realizados com diferentes posições da cubeta e da *webcam*. A melhor configuração foi escolhida com base na qualidade e uniformidade das imagens registradas.

3.4 Iluminação do sistema

A iluminação do compartimento foi realizada por meio de uma lâmpada fluorescente em formato espiral com 6W de potência e temperatura de cor de 4.000K (cor branca). Com o propósito de uniformizar a iluminação, a lâmpada foi colocada fora do compartimento a uma distância de 18 cm das amostras. A fim de evitar o efeito estroboscópico em lâmpadas fluorescentes, uma folha de papel ofício na cor branca foi utilizada (**Fig. 3.2**). Com isso, além de servir como tampa da parte superior do compartimento, a folha de papel ofício ajudou a dispersar a luz no interior do compartimento.

3.5 Bomba peristáltica

Uma bomba peristáltica Ismatec[®], modelo IPC com 8 canais e 12 roletes, localizada externamente ao compartimento, foi utilizada apenas para propulsão das amostras de óleos vegetais por intermédio de tubos de bombeamento de silicone. Com isso, reduziu-se o tempo de análise e eliminou-se a necessidade de intervenção do analista evitando também qualquer modificação nas condições internas do compartimento.

3.6 Notebook

O notebook foi utilizado com o intuito de atribuir portabilidade ao sistema. Além de servir para conexão da webcam e controle de sua configuração, o notebook permitiu comandar a captura das imagens e está profundamente relacionado a todo o procedimento quimiométrico.

3.7 Aquisição das imagens

A *webcam*, conectada ao microcomputador e posicionada a uma distância de 10 cm da cubeta, após o interrupção do fluxo, realizava o registro de dez imagens sequenciais para cada amostra. As imagens registradas continham 24 bits (16.7 milhões de cores) e 2880 pixels ×

1620 pixels de resolução espacial, sendo armazenadas em formato JPEG (jpg). Na **Figura 3.3** são apresentadas as imagens obtidas para cada tipo de óleo vegetal analisado e, na **Figura 3.4**, são mostradas as imagens de óleos de canola não expirado e expirado, respectivamente. A captura das imagens foi realizada pelo software da própria *webcam* e sua configuração padrão foi mantida.

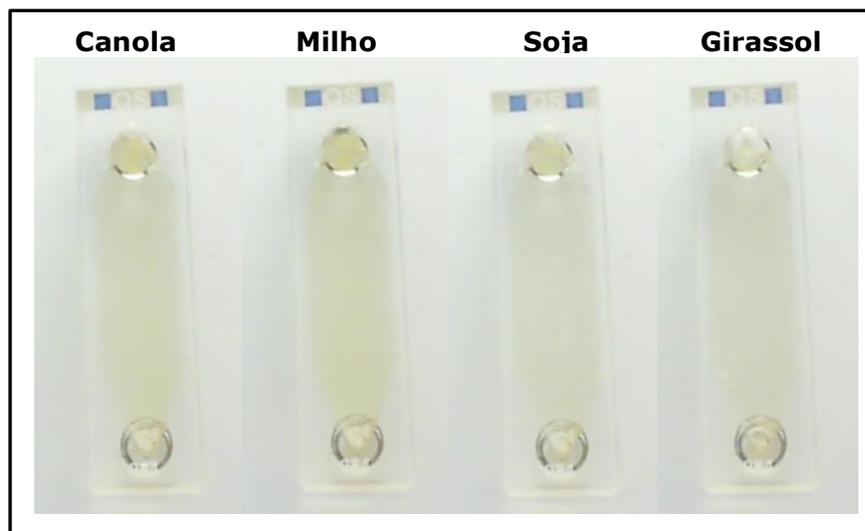


Figura 3.3. Imagens dos óleos vegetais obtidas pela webcam no interior do sistema desenvolvido.



Figura 3.4. Imagens de óleos de canola obtidas para uma amostra não expirada e uma expirada há 24 meses.

Após a captura das dez imagens por amostra, cada imagem foi decomposta em histogramas correspondentes a uma região específica da

imagem. Em seguida, calculou-se a média destes histogramas que, por sua vez, foi utilizada como resposta instrumental referente àquela amostra (**Fig. 3.5**).

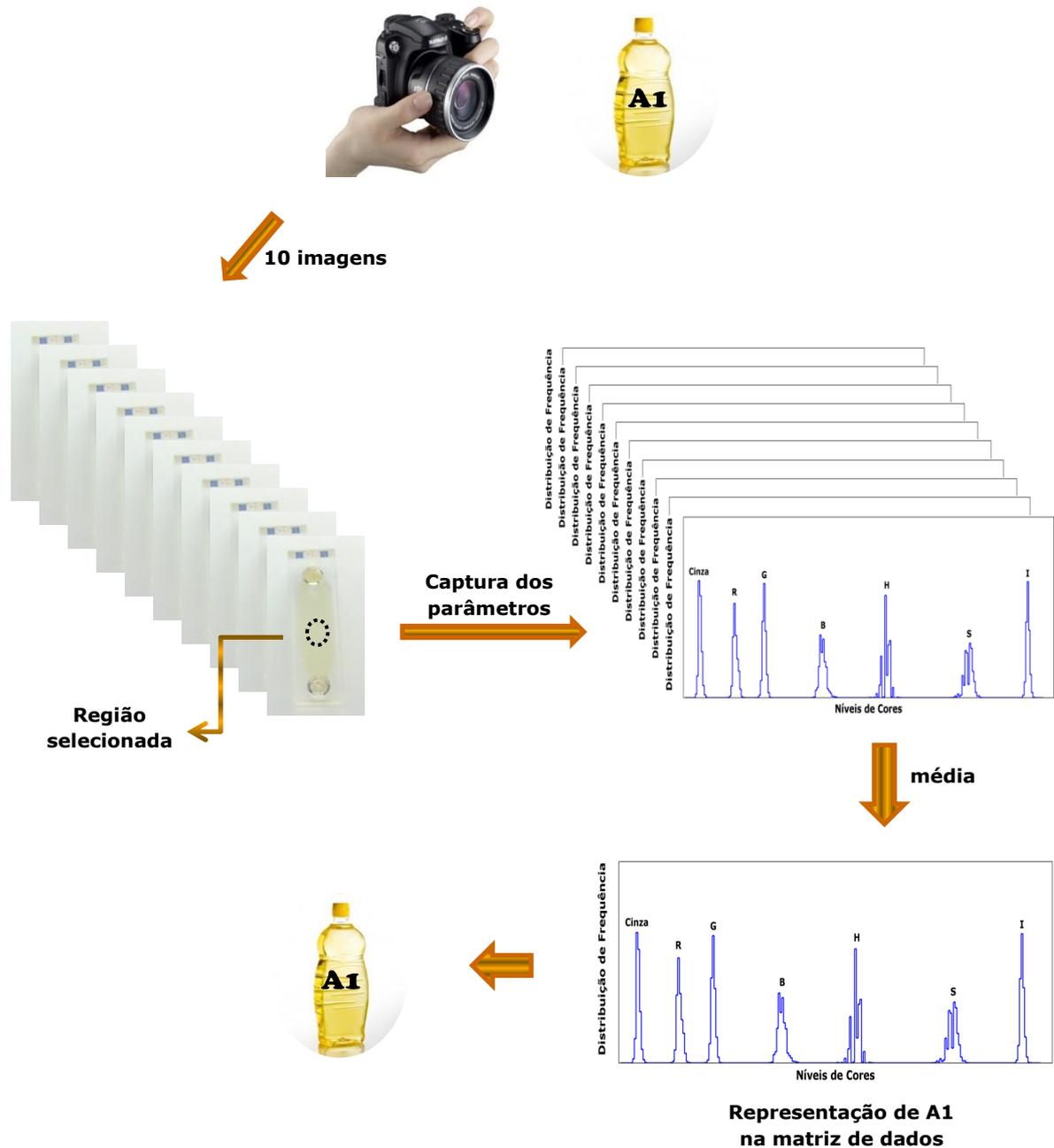


Figura 3.5. Representação de uma amostra na matriz de dados.

A matriz de dados era então formada pelas amostras, dispostas nas linhas, e pelas variáveis correspondentes aos níveis de cores obtidos para

cada componente de cor, dispostas nas colunas. Para obtenção desses histogramas, três modelos para a cor de um pixel foram utilizados: o padrão RGB (R-Red, G- Green e B-Blue), tons de cinza e o sistema HSI (H-Hue, S-Saturation e I- Intensity). A **Figura 3.6** apresenta a composição da matriz de dados.

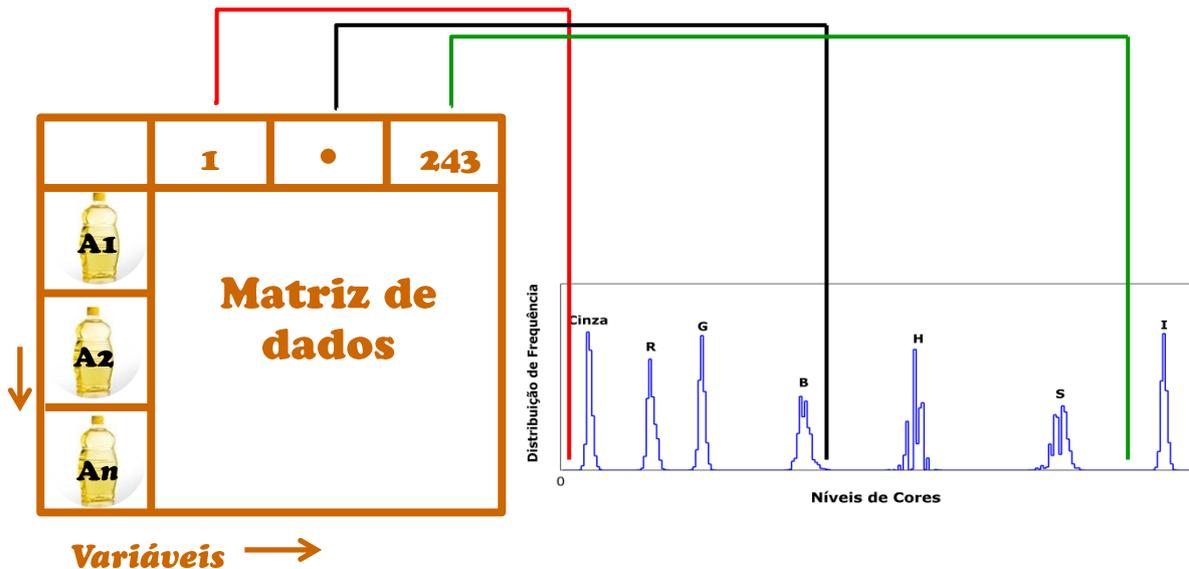


Figura 3.6. Representação da matriz de dados usada no tratamento dos dados.

3.8 Análise Quimiométrica

As informações contidas nos histogramas foram utilizadas para a classificação dos óleos vegetais comestíveis e a capacidade discriminatória da técnica de imagens digitais, aliada às técnicas de reconhecimento de padrões, foi avaliada segundo duas abordagens:

1. *Classificação quanto ao tipo de oleaginosa*
2. *Classificação quanto ao estado de conservação*

No primeiro estudo de caso, os procedimentos foram aplicados ao grupo de amostras com prazo de validade não expirado para classificação dos óleos vegetais nos tipos: *canola, milho, girassol ou soja*.

No segundo estudo de caso, os procedimentos foram aplicados dentro de cada categoria de óleo vegetal em estudo - canola, milho, girassol e soja - para classificação quanto ao estado de conservação: *prazo de validade expirado (E) e não expirado (NE)*.

Em ambos os casos, os conjuntos de dados foram divididos em subconjuntos de treinamento (70%) e previsão (30%) usando o algoritmo clássico Kennard-Stone (KS) (KENNARD e STONE, 1969). As **Tabelas 3.1** e **3.2** apresentam o número de amostras em cada classe nos dois problemas de classificação.

Tabela 3.1. Número de amostras de treinamento e previsão em cada classe, para o primeiro estudo de caso.

Classes	Subconjuntos	
	Treinamento	Previsão
Canola	10	5
Milho	10	5
Girassol	10	5
Soja	10	5

Tabela 3.2. Número de amostras de treinamento e previsão em cada classe, para o segundo estudo de caso.

Classes		Subconjuntos	
		Treinamento	Previsão
Canola	Expirada	9	5
	Não Expirada	10	5
Milho	Expirada	14	6
	Não Expirada	10	5
Girassol	Expirada	12	6
	Não Expirada	10	5
Soja	Expirada	14	6
	Não Expirada	10	5

Estudos preliminares foram realizados com o intuito de avaliar a capacidade de discriminação das componentes de cores utilizadas. Para

isso, uma análise exploratória empregando a PCA, bem como a construção dos modelos SPA-LDA, SW-LDA e PLS-DA foram avaliados com base no uso de componentes de cores individuais e combinadas.

Nesse trabalho, a formulação SPA apresentada por Pontes et al. (2005) foi empregada e, a partir daí, as amostras de treinamento foram usadas como critério de escolha do melhor subconjunto de variáveis.

No caso do algoritmo Stepwise, seis valores de coeficientes de correlação múltipla foram avaliados (0,1, 0,3, 0,7, 0,9, 0,95 e 0,99) a fim de decidir quais variáveis deveriam ser descartadas. O melhor valor foi selecionado com base nos erros de classificação no subconjunto de treinamento. Para os modelos PLS-DA, o limiar empregado foi 0,5.

As PCA e PLS-DA foram realizadas utilizando o programa The Unscrambler® X.1 (CAMO S/A). Os algoritmos KS, Stepwise e SPA foram executados no programa Matlab® 7.10.



Capítulo IV
Resultados e Discussão

4 RESULTADOS E DISCUSSÃO

4.1 Histogramas

A **Figura 4.1** apresenta os histogramas das 132 amostras de óleos vegetais analisados. Esses histogramas, obtidos para as componentes de cores tons de cinza, R, G, B, H, S e I, respectivamente, representam as distribuições de frequência dos níveis (ou tons) de cores presentes na imagem e a quantidade de pixels que correspondem a esses níveis. Para cada componente de cor, 256 níveis podem ser assumidos variando no intervalo de 0 a 255. Considerando-se as sete componentes, 1792 níveis de cores foram obtidos.

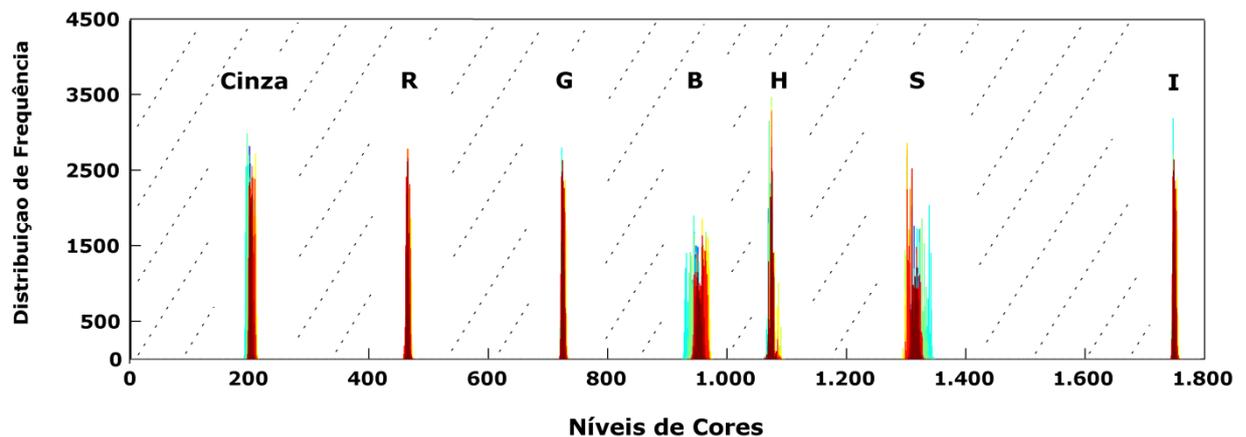


Figura 4.1. Histogramas das 132 amostras de óleos vegetais comestíveis obtidos para as componentes de cores Tons de cinza, R: *Red*; G: *Green*; B: *Blue*; H: *Hue*; S: *Saturation* e I: *Intensity*.

Como pode ser observado, existem grandes espaços onde a resposta é igual à zero, indicando que nem todos os 256 níveis de cores permitidos para um pixel estão presentes na imagem. Estas variáveis foram removidas e os histogramas resultantes, com 243 níveis de cores, são mostrados na **Figura 4.2**.

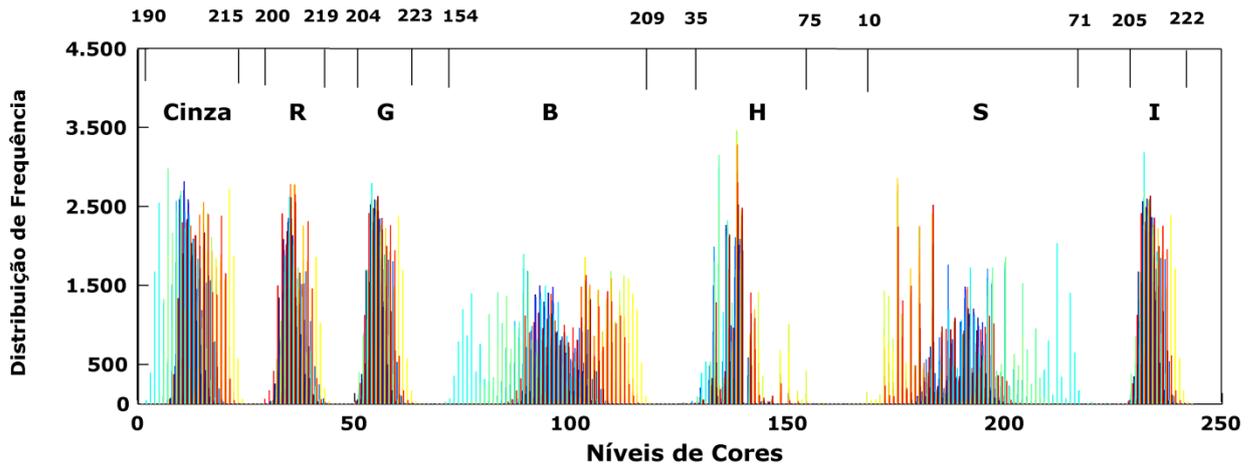


Figura 4.2. Histogramas das 132 amostras de óleos vegetais comestíveis após a remoção dos zeros. Tais histogramas foram usados no tratamento quimiométrico. Tons de Cinza, R: Red; G: Green; B: Blue; H: Hue; S: Saturation e I: Intensity. Os valores apresentados na parte superior da figura correspondem aos níveis de cores presentes nos histogramas originais.

4.2 Classificação com respeito ao tipo

Para a classificação com respeito ao tipo, 60 amostras de óleos vegetais com prazo de validade não expirado foram utilizadas, sendo 15 de cada classe: Canola, Milho, Girassol e Soja.

Na **Figura 4.3** são apresentados os histogramas das amostras de óleos vegetais em estudo, obtidos para as componentes tons de cinza, R, G, B, H, S e I.

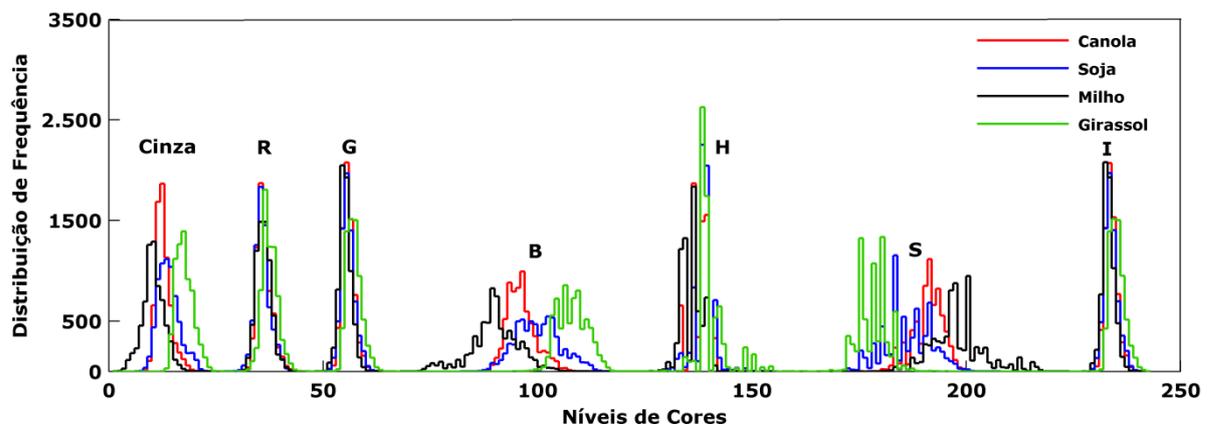


Figura 4.3. Histogramas médios das quatro classes de óleos em estudo, obtidos para as componentes Tons de cinza, R, G, B, H, S e I, respectivamente.

Como pode ser observado, em todos os casos as componentes B e S destacam-se por apresentar maior variabilidade dos dados, o que está relacionado à maior quantidade de níveis de cores preenchidos nessas componentes, embora esses níveis estejam presentes em menor frequência.

A coloração amarelada dos óleos vegetais empregados neste estudo pode ser atribuída aos carotenóides, pigmentos capazes de absorver luz na faixa UV e visível do espectro (MORETTO E FETT, 1998). Para que a cor amarela seja percebida pelo olho humano, é necessário que, de todos os comprimentos de ondas incidentes sobre a oleaginosa, apenas os comprimentos correspondentes à cor azul sejam absorvidos enquanto os demais sofrem reflexão. Os processos de absorção e reflexão da luz e as sensações provocadas no olho humano dependem das características do meio onde está ocorrendo a incidência luminosa, considerando-se principalmente sua composição química (MCMURRY, 2011).

Dessa forma, as diferenças de perfil apresentadas pelas oleaginosas no histograma da componente B podem ser relacionadas às diferentes concentrações de carotenoides em cada tipo de óleo. Além disso, deve-se considerar os demais constituintes presentes em cada óleo e suas capacidades de alterar a intensidade da cor percebida pelo olho humano.

Explicações para o comportamento apresentado na componente S podem estar relacionadas tanto a composição química quanto aos processos de refino aos quais estas amostras foram submetidas. A saturação expressa à pureza da cor em relação à proporção de cor branca adicionada. Uma vez que as amostras têm aproximadamente a mesma tonalidade (amarelo ~ 565 - 590 nm), as variações na cor, e assim seu grau de pureza, são indicados pela sua saturação.

4.2.1 Análise de Componentes Principais (PCA)

Nas **Figuras 4.4a-b** são apresentados os gráficos dos escores obtidos para PC2 x PC1 para a combinação de componentes tons de cinza

+ RGB + HSI e o gráfico dos pesos das variáveis, respectivamente. Na **Figura 4.4a** três agrupamentos principais podem ser visualizados: milho, girassol e canola. Contudo, uma substancial sobreposição das amostras de óleo de soja é observada.

O gráfico dos pesos (**Fig. 4.4b**), usado para verificar a contribuição de cada variável, revela que as componentes Tons de cinza, B e S parecem apresentar maior poder de discriminação. A partir destes resultados, foram obtidos os gráficos dos escores de PC2 x PC1 para as essas componentes de cores (**Fig. 4.5a-c**).

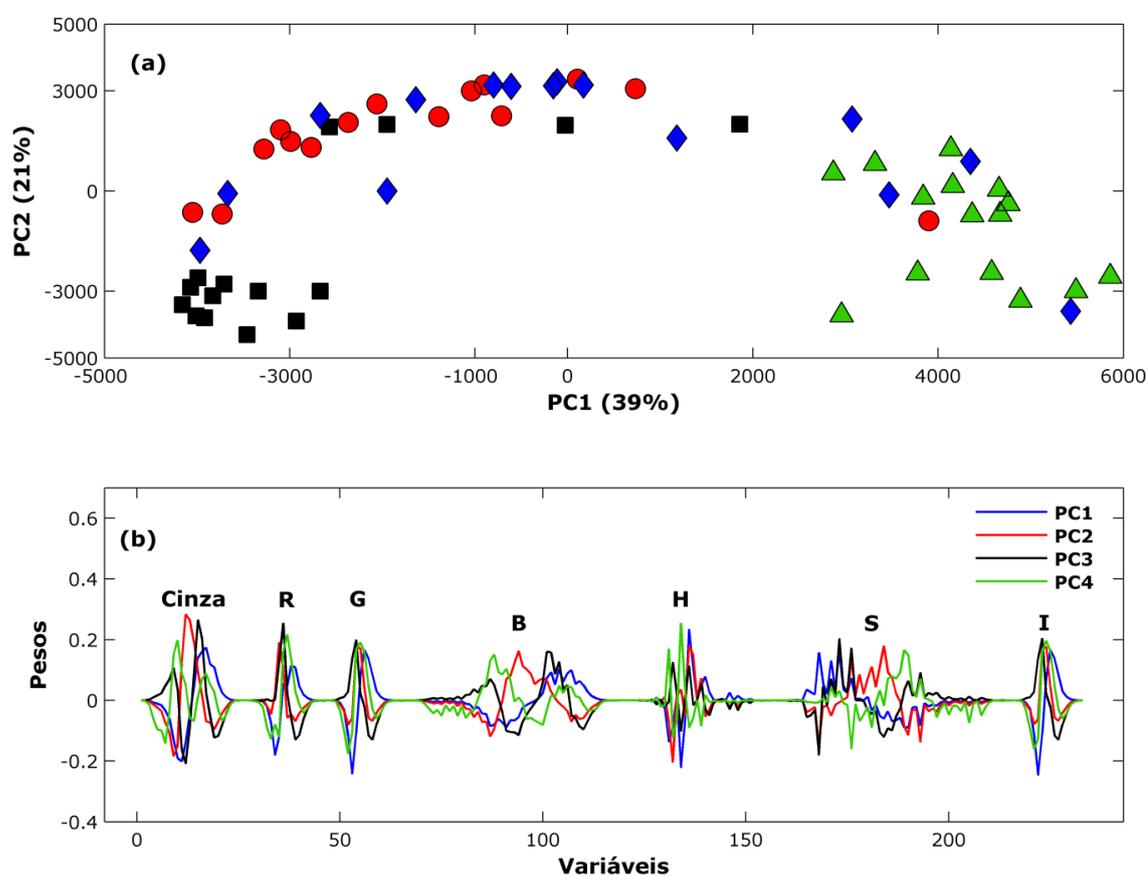


Figura 4.4. (a) Gráficos dos escores de PC1 x PC2, obtidos para as componentes de cores tons de cinza + RGB + HSI e (b) gráfico dos pesos das variáveis em cada componente de cor. (●: canola, ■: milho, ▲: girassol e ◆: soja).

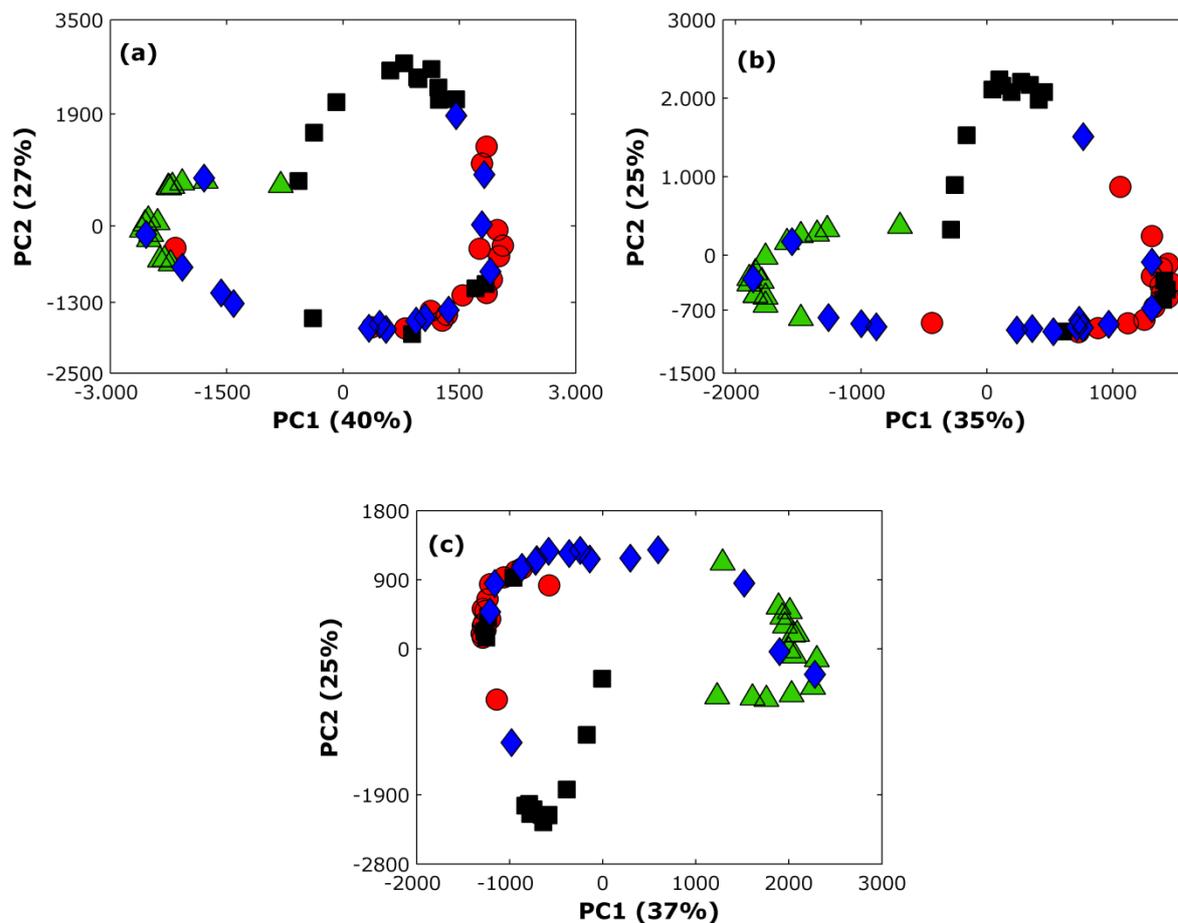


Figura 4.5. Gráficos dos escores de PC1 x PC2, obtidos para as componentes de cores (a) tons de cinza, (b) B e (c) S, para os quatro tipos de óleos vegetais em estudo. (•: canola, ■: milho, ▲: girassol e ◆: soja).

Como pode ser observado, um comportamento similar ao da **Figura 4.4a** é obtido, com três agrupamentos principais detectados (milho, girassol e canola) e sobreposição das amostras de soja.

A aplicação da PCA ao conjunto de dados forneceu resultados promissores quanto à discriminação das classes em estudo. Tais resultados sugerem uma investigação com base no emprego de técnicas de seleção de variáveis e métodos de reconhecimento de padrões supervisionados com o intuito de melhorar a discriminação entre as diferentes amostras, bem como construir modelos capazes de realizar previsões em amostras futuras.

4.2.2 Escolha da faixa de trabalho

A partir dos resultados fornecidos pela PCA, um estudo foi realizado com o intuito de avaliar a capacidade discriminatória de todas as componentes de cores utilizadas. Para isso, foram construídos modelos de classificação LDA/SPA, LDA/SW e PLS-DA usando as componentes de cores individuais e combinadas.

A **Tabela 4.1** apresenta a taxa de classificação correta obtida pelos modelos de classificação aplicados aos conjuntos de treinamento e previsão (em diferentes canais).

Tabela 4.1. Taxa de classificação correta obtida com os modelos de classificação LDA/SPA, LDA/SW e PLS-DA para o conjunto de previsão.

Componentes de cores	Taxa de Classificação Correta (%)					
	LDA/SPA		LDA/SW		PLS-DA	
Componentes de cores	Treinamento	Previsão	Treinamento	Previsão	Treinamento	Previsão
Cinza + RGB + HSI	80	90	83	85	88	90
Cinza	60	55	88	85	70	70
RGB	95	85	100	90	88	90
HSI	88	85	75	90	80	90
Cinza + RGB	95	90	95	90	85	90
Cinza + HSI	73	85	100	85	80	95
RGB + HSI	95	85	100	90	85	90
Red	73	85	78	85	58	70
Green	60	80	65	55	55	60
Blue	80	90	95	90	75	95
Hue	85	80	88	85	75	90
Saturação	95	95	95	90	80	95
Intensidade	63	75	65	65	58	65

De acordo com a **Tabela 4.1**, um número considerável de componentes de cores apresentou bom desempenho de classificação das amostras de óleos, para diferentes modelos gerados. Entretanto, os melhores resultados para os conjuntos de treinamento e previsão, em

todos os modelos, simultaneamente, foram obtidos a partir da componente S. Este resultado corrobora com o perfil apresentado pelos histogramas (**Fig. 4.3**) e com as informações fornecidas por este canal, sendo esta a componente selecionada para a geração dos modelos de classificação, os quais serão apresentados nas próximas seções.

4.2.3 Modelos LDA

A **Figura 4.6a** apresenta função de custo que representa o menor risco de uma classificação incorreta do conjunto treinamento *versus* o número de variáveis selecionadas pelo algoritmo SPA. Como pode ser visto, o mínimo da curva ocorre com 19 variáveis selecionadas. Tais variáveis são apresentadas na **Figura 4.6b**. Essas variáveis foram selecionadas com base na ocorrência de um determinado nível de cor nos diferentes tipos de óleo.

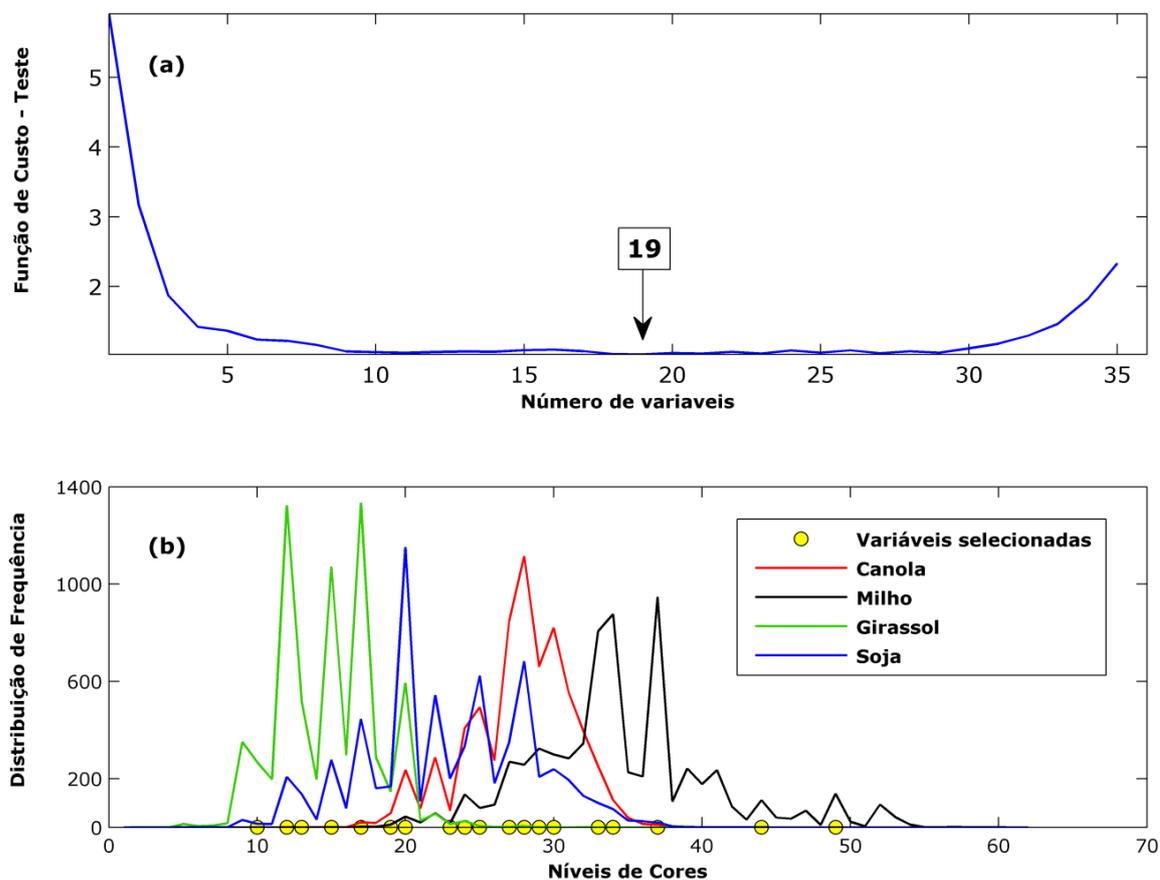


Figura 4.6. (a) Gráfico da função de custo do conjunto de treinamento *versus* o número de variáveis selecionadas pelo algoritmo SPA, (b) variáveis selecionadas pelo SPA.

Quando aplicado às amostras dos conjuntos de treinamento e previsão, o modelo LDA construído com as 19 variáveis selecionadas pelo SPA alcançou uma taxa de classificação correta de 95%, correspondente a um erro de classificação.

Para efeito de comparação, o modelo LDA/SW foi construído utilizando os mesmos subconjuntos de amostras que o modelo LDA/SPA. Um número um pouco maior de variáveis foi selecionado pelo SW: um total de 27. Tais variáveis são destacadas na **Figura 4.7**.

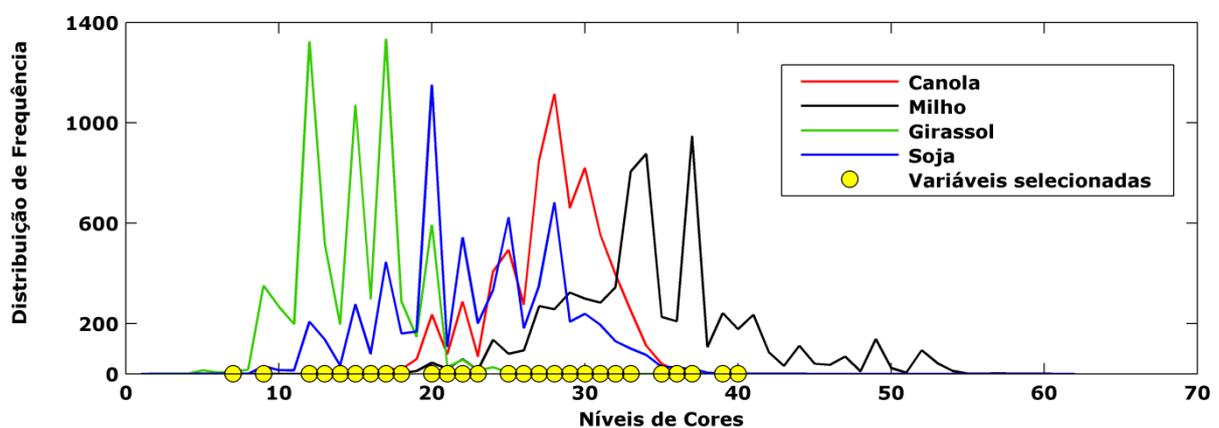


Figura 4.7. Variáveis selecionadas pela formulação SW empregadas na construção do modelo SW-LDA.

Como pode ser visto (**Fig. 4.7**), as variáveis selecionadas pelo SW são bem parecidas com aquelas selecionadas pelo SPA. O modelo LDA construído com essas variáveis classificou corretamente 95% e 90% das amostras de treinamento e previsão (dois erros de classificação em cada conjunto), respectivamente.

As **Figuras 4.8a-b** apresentam os gráficos dos escores obtidos pela LDA com prévia seleção de variáveis pelo SPA (**Fig. 4.8a**) e SW (**Fig. 4.8b**). É possível observar no gráfico dos escores obtido pelo LDA/SPA uma discriminação das classes canola, soja e milho ao longo da FD1. As amostras de girassol, por sua vez, encontram-se bem discriminadas das demais classes com valores de escores mais elevados em FD2 e menores em FD1. Um comportamento parecido, porém com valores de escores

invertidos, pode ser encontrado no gráfico dos escores obtidos pelo LDA/SW (Fig. 4.8b).

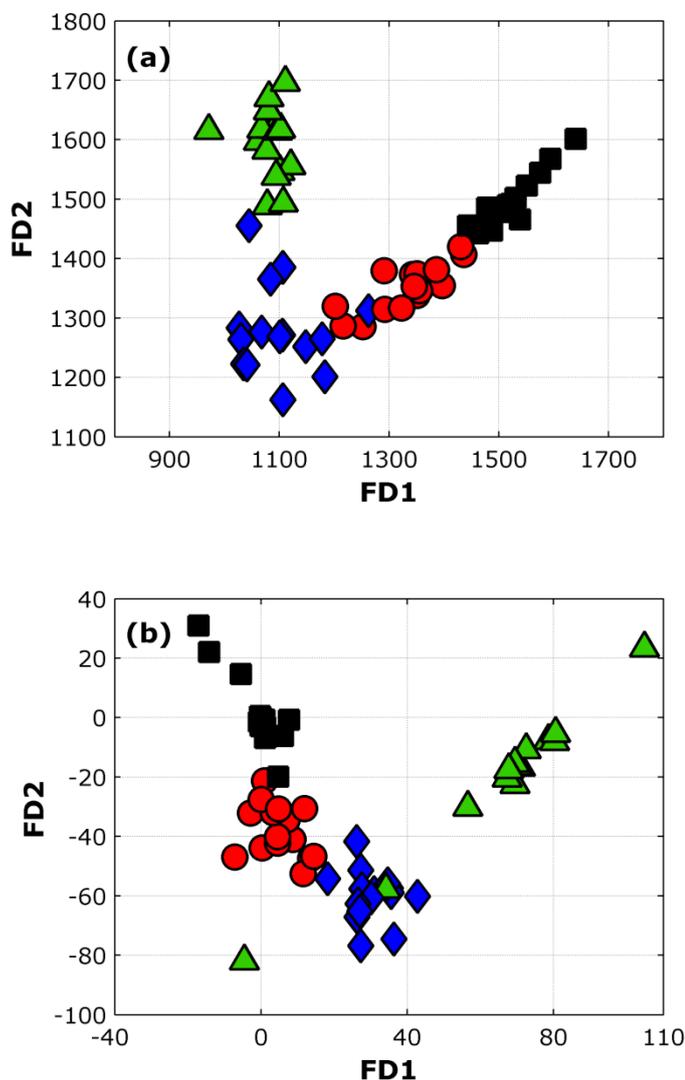


Figura 4.8. Gráficos dos escores das funções discriminantes FD2 x FD1 obtidos pelo (a) SPA e (b) SW para as 60 amostras de óleos vegetais não expiradas. (•: canola, ■: milho, ▲: girassol e ◆: soja).

Quando comparados aos gráficos dos escores obtidos pela PCA para a mesma componente S (Fig. 4.5c), os gráficos dos escores obtidos pela LDA (Fig. 4.8a-b) mostram uma discriminação mais efetiva das classes de óleos vegetais utilizando as variáveis selecionadas tanto pelo SPA, como pelo SW.

4.2.4 Classificação PLS-DA

Para o conjunto de previsão, os resultados obtidos pelo modelo PLS-DA foram similares àqueles encontrados pelo LDA/SPA e superiores aos obtidos pelo LDA/SW. Contudo, uma taxa de classificação correta de apenas 80% foi alcançada para o conjunto de treinamento.

Os resultados detalhados para os modelos LDA/SPA, LDA/SW e PLS-DA podem ser encontrados na **Tabela 4.2**. Os acertos na classificação são obtidos quando os valores encontrados na *Classe Prevista* são iguais aos valores exibidos na *Classe Verdadeira*. Em contrapartida, os erros de classificação correspondem aos valores diferentes encontrados nas *Classes Prevista e Verdadeira*.

Tabela 4.2. Resultados da classificação obtida com os modelos LDA/SPA, LDA/SW e PLS-DA para os conjuntos de previsão e treinamento. (1): amostras de canola, (2): amostras de milho, (3): amostras de girassol e (4): amostras de soja. O número de variáveis ou fatores empregados em cada modelo é indicado entre parênteses. N indica o número de amostras de previsão utilizadas neste estudo.

Conjunto de Treinamento															
Classe Verdadeira		LDA/SPA (19)				LDA/SW (27)*					PLS-DA (4)				
		Classe Prevista				Classe Prevista					Classe Prevista				
		TCC				TCC					TCC				
N	1	2	3	4	95%	1	2	3	4	95%	1	2	3	4	80%
1	10	10	-	-	-	10	-	-	-	-	10	-	-	-	-
2	10	1	9	-	-	1	9	-	-	-	3	7	-	-	-
3	10	-	-	10	-	-	-	10	-	-	-	-	9	1	-
4	10	-	-	1	9	1	-	-	9	-	1	1	2	6	-

Conjunto de Previsão															
Classe Verdadeira		LDA/SPA (19)				LDA/SW (27)*					PLS-DA (4)				
		Classe Prevista				Classe Prevista					Classe Prevista				
		TCC				TCC					TCC				
N	1	2	3	4	95%	1	2	3	4	90%	1	2	3	4	95%
1	5	5	-	-	-	5	-	-	-	-	5	-	-	-	-
2	5	-	5	-	-	-	5	-	-	-	-	5	-	-	-
3	5	-	-	5	-	-	-	3	2	-	-	-	5	-	-
4	5	1	-	-	4	-	-	-	5	-	1	-	-	4	-

*Valor de limiar igual a 0.99

Como pode ser observado, os modelos LDA/SPA e PLS-DA apresentaram desempenho similar para o conjunto de previsão, com uma TCC de 95%, enquanto o LDA/SW apresentou um desempenho inferior de 90%. Em termos de TCC para o conjunto completo de amostras (amostras de treinamento e previsão), os modelos LDA/SPA, LDA/SW e PLS-DA apresentaram 95%, 93% e 82%, respectivamente.

4.3 Classificação com respeito ao estado de conservação

4.3.1 Histogramas

O estado de conservação dos óleos vegetais comestíveis é comprometido por uma série de fatores, entre eles: exposição à luminosidade, altas temperaturas e ação do oxigênio. A velocidade de degradação de amostras de óleos vegetais submetidas às mesmas condições de armazenamento é influenciada principalmente pela sua estrutura molecular. A presença de duplas ligações na cadeia de grupo alquila direciona a estabilidade diante do processo degradativo de rancidez autoxidativa, fazendo com que oleaginosas diferentes sofram processos degradativos em diferentes proporções e intervalos de tempo. Dessa forma, é conveniente realizar a classificação dos óleos com respeito ao estado de conservação dentro de cada categoria de oleaginosa.

A **Figura 4.9a-d** apresenta os histogramas médios das amostras expiradas e não expiradas de cada tipo de óleo vegetal (canola, milho, girassol e soja).

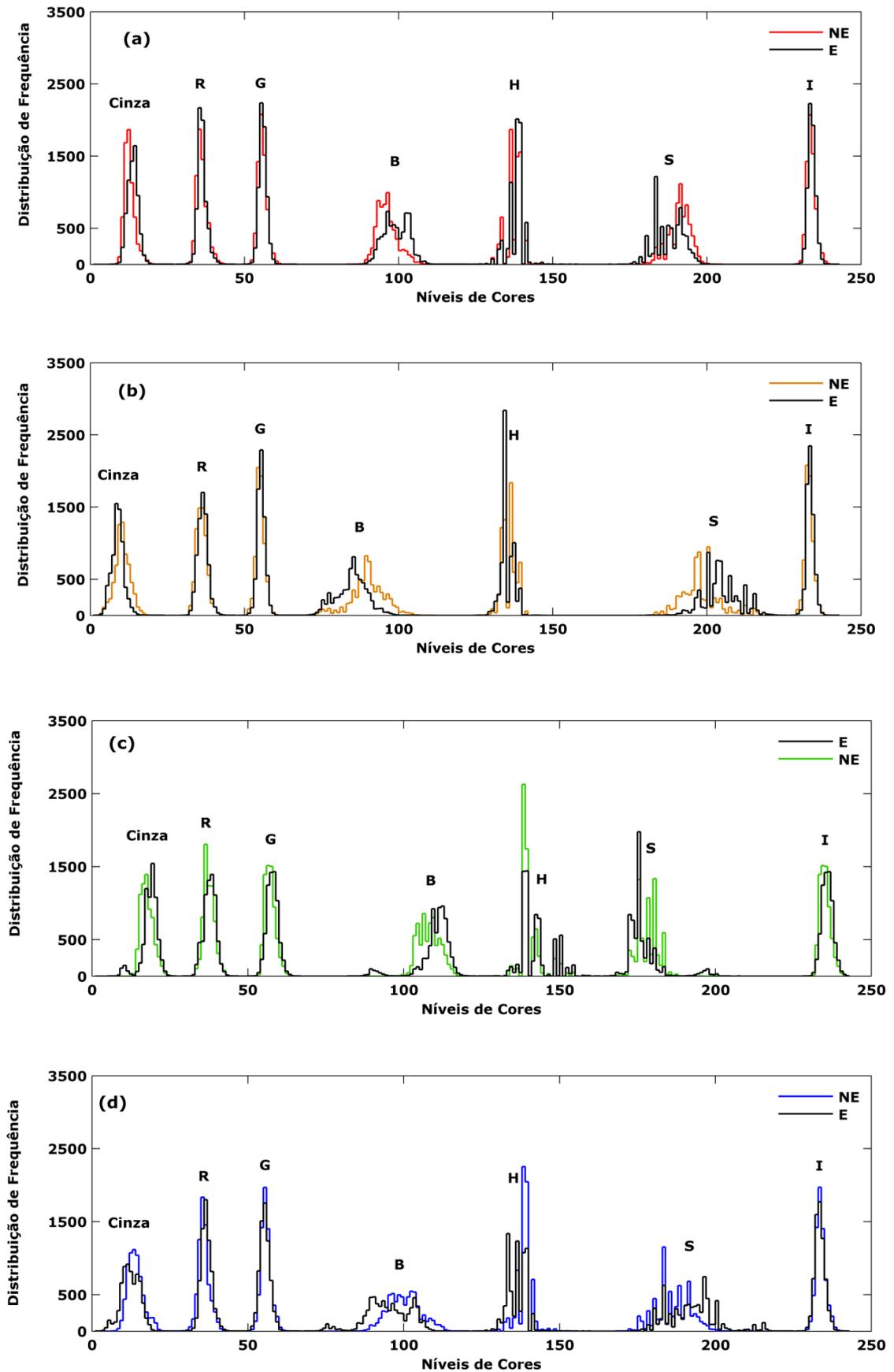


Figura 4.9. Histogramas médios das amostras de óleos expiradas e não expiradas de (a) canola, (b) milho, (c) girassol e (d) soja.

Como pode ser observado, existe uma substancial sobreposição dos histogramas das classes expiradas (E) e não expiradas (NE). Isso ocorre principalmente nas componentes R, G e I. Em contrapartida, a maior diferença em termos de perfil dos histogramas é encontrada nas componentes B, H e S.

4.3.2 Análise de Componentes Principais (PCA)

Nas **Figuras 4.10a-d** são apresentados os gráficos dos escores de todas as componentes de cores obtidos para as amostras expiradas e não expiradas de cada oleaginosa.

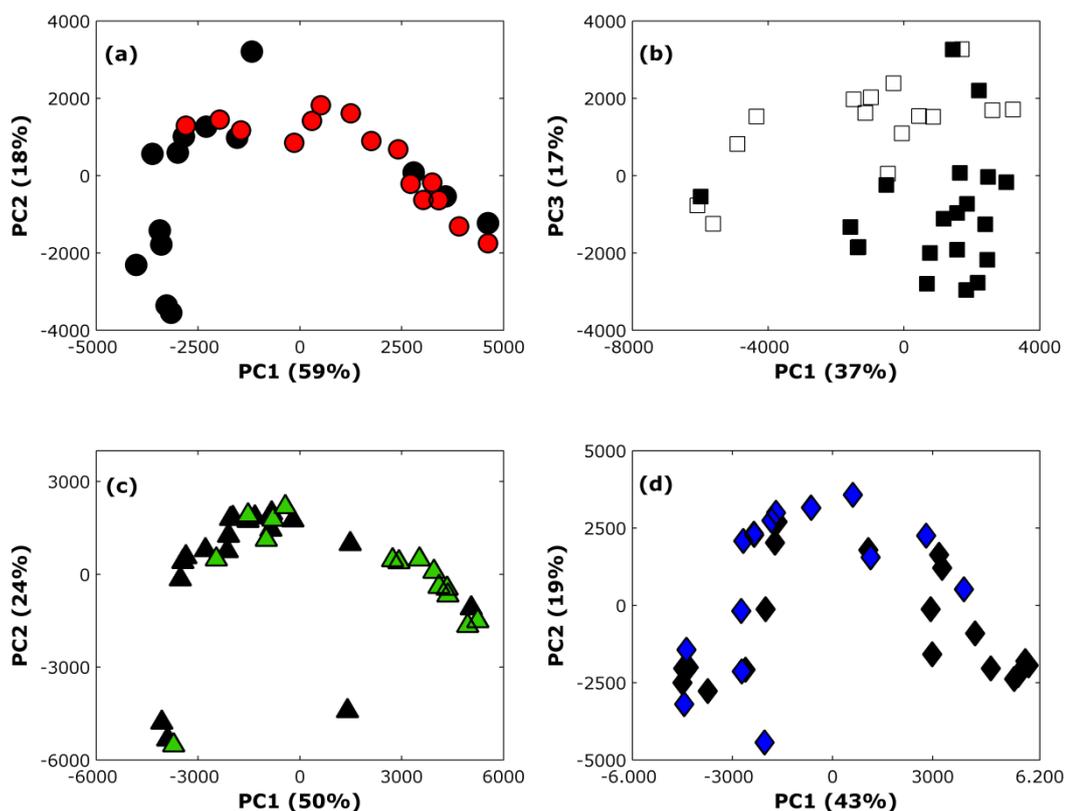


Figura 4.10. Gráficos dos escores obtidos a partir de todas as componentes de cores para as amostras expiradas e não expiradas de (a) canola, (b) milho, (c) girassol e (d) soja. Em todos os casos, os símbolos preenchidos na cor preta correspondem às amostras expiradas.

Como pode ser observado, a melhor discriminação é encontrada na oleaginosa de milho (**Fig. 4.10b**), onde as amostras não expiradas

apresentam-se, em sua maioria, com valores de escores mais positivos ao longo de PC3. Para as oleaginosas de canola e girassol, uma tendência de separação entre as amostras expiradas e não expiradas pode ser observada ao longo de PC1.

4.3.3 Escolha da faixa de trabalho

Um estudo similar ao da **seção 4.2.2** foi realizado para cada tipo de oleaginosa com o propósito de encontrar padrões entre as amostras expiradas e não expiradas. Nas **Tabelas 4.3 a 4.6** são apresentados os melhores resultados de classificação (de forma detalhada) obtidos para cada oleaginosa, em termos da TCC para os conjuntos de treinamento e previsão.

Tabela 4.3. Resultados da classificação obtida com os modelos LDA/SPA, LDA/SW e PLS-DA para os conjuntos de treinamento e previsão das amostras de milho. (E): amostras expiradas, (NE): amostras não expiradas. O número de variáveis ou fatores empregados em cada modelo é indicado entre parênteses. N indica o número de amostras de previsão utilizadas neste estudo.

MILHO										
Conjunto de Treinamento										
Classe Verdadeira	N	LDA/SPA (6) - HSI			LDA/SW (8)* - HSI			PLS-DA (2) - RGB		
		Classe Prevista		TCC	Classe Prevista		TCC	Classe Prevista		TCC
		E	NE		E	NE		E	NE	
E	14	14	-	96%	14	-	100%	12	2	92%
NE	10	1	9		-	10		-	10	
Conjunto de Previsão										
Classe Verdadeira	N	SPA-LDA (6) - HSI			LDA/SW (8)* - HSI			PLS-DA (2) - RGB		
		Classe Prevista		TCC	Classe Prevista		TCC	Classe Prevista		TCC
		E	NE		E	NE		E	NE	
E	6	6	-	100%	6	-	100%	6	-	100%
NE	5	-	5		-	5		-	5	

*valor de limiar igual a 0.7

De acordo com a **Tabela 4.3**, os três métodos avaliados apresentaram desempenhos de previsão semelhantes para as amostras de

milho utilizando os canais HSI (para o LDA/SW e LDA/SPA) e RGB (para o PLS-DA). Levando em conta o número de erros para o conjunto completo de amostras (treinamento e previsão), o modelo LDA/SW forneceu o melhor resultado.

Os resultados para as amostras de soja são mostrados na **Tabela 4.4**.

Tabela 4.4. Resultados da classificação obtida com os modelos LDA/SPA, LDA/SW e PLS-DA para os conjuntos de treinamento e previsão das amostras de soja. (E): amostras expiradas, (NE): amostras não expiradas. O número de variáveis ou fatores empregados em cada modelo é indicado entre parênteses. N indica o número de amostras de previsão utilizadas neste estudo.

SOJA										
Conjunto de Treinamento										
Classe Verdadeira	N	LDA/SPA (13) - Cinza + HSI			LDA/SW (23)* - Cinza + HSI			PLS-DA (2) - RGB		
		Classe Prevista		TCC	Classe Prevista		TCC	Classe Prevista		TCC
		E	NE	100%	E	NE	100%	E	NE	88%
E	14	14	-		14	-		12	2	
NE	10	-	10		-	10		1	9	
Conjunto de Previsão										
Classe Verdadeira	N	LDA-SPA (13) - Cinza + HSI			LDA/SW (23)* - Cinza + HSI			PLS-DA (2) - RGB		
		Classe Prevista		TCC	Classe Prevista		TCC	Classe Prevista		TCC
		E	NE	82%	E	NE	73%	E	NE	82%
E	6	4	2		3	3		4	2	
NE	5	-	5		-	5		-	5	

*valor de limiar igual a 0.99

No caso das amostras de soja, o modelo LDA/SPA gerado com base na combinação de componentes Tons de Cinza + HSI, forneceu os melhores resultados com uma TCC de 100% para o conjunto de treinamento e 82% para o conjunto de previsão. O modelo PLS-DA com melhores resultados foi obtido com base nas componentes RGB, que embora tenha atingido o mesmo percentual de acerto que o LDA/SPA para o conjunto de previsão, mostrou desempenho inferior na classificação das amostras de treinamento. Para as amostras de soja, os melhores resultados em termos de TCC para o conjunto completo de amostras foi

obtido pelo LDA/SPA, correspondente a 94% (2 erros de classificação). As TCC para o conjunto completo de amostras para os modelos LDA/SW e PLS-DA foram 91% (3 erros) e 86% (5 erros), respectivamente.

Para as amostras de canola (**Tabela 4.5**), o modelo LDA/SPA construído com as componentes HSI apresentou melhor desempenho de classificação do conjunto completo de amostras, com 93% de TCC para o LDA/SPA contra 90% para LDA/SW e 79% para PLS-DA, ambos construídos com as componentes RGB. As TCC mencionadas acima correspondem a 2, 3 e 6 erros totais de classificação, respectivamente.

Tabela 4.5. Resultados da classificação obtida com os modelos LDA/SPA, LDA/SW e PLS-DA para os conjuntos de treinamento e previsão das amostras de canola. (E): amostras expiradas, (NE): amostras não expiradas. O número de variáveis ou fatores empregados em cada modelo é indicado entre parênteses. N indica o número de amostras de previsão utilizadas neste estudo.

CANOLA										
Conjunto de Treinamento										
Classe Verdadeira	N	LDA/SPA (16) - HSI			LDA/SW (13)* - RGB			PLS-DA (1) - RGB		
		Classe Prevista		TCC	Classe Prevista		TCC	Classe Prevista		TCC
		E	NE	100%	E	NE	90%	E	NE	79%
E	9	9	-		7	2		6	3	
NE	10		10		-	10		1	9	
Conjunto de Previsão										
Classe Verdadeira	N	LDA/SPA (16) - HSI			LDA/SW (13)* - RGB			PLS-DA (1) - RGB		
		Classe Prevista		TCC	Classe Prevista		TCC	Classe Prevista		TCC
		E	NE	80%	E	NE	90%	E	NE	80%
E	5	5	-		4	1		4	1	
NE	5	2	3		-	5		1	4	

*valor de limiar igual a 0.95

As amostras de girassol (**Tabela 4.6**) foram efetivamente classificadas pelo modelo LDA/SPA construído para a combinação de cores Tons de Cinza + HSI. Tal modelo atingiu uma TCC para o conjunto completo de amostras de 97%, correspondente a apenas 1 erro de classificação encontrado no conjunto de previsão. Diferentemente dos casos analisados anteriormente, a componente S apareceu individualmente realizando uma das melhores classificações. Para os

modelos LDA/SW e PLS-DA foram atingidas TCC de 94% e 88%, respectivamente, correspondentes a 2 e 4 erros de totais de classificação.

Tabela 4.6 Resultados da classificação obtida com os modelos LDA/SPA, LDA/SW e PLS-DA para os conjuntos de treinamento e previsão das amostras de girassol. (E): amostras expiradas, (NE): amostras não expiradas. O número de variáveis ou fatores empregados em cada modelo é indicado entre parênteses. N indica o número de amostras de previsão utilizadas neste estudo.

GIRASSOL										
Conjunto de Treinamento										
Classe Verdadeira	N	LDA/SPA (17) - Cinza + HSI			LDA/SW (16)* - S			PLS-DA (3) - HSI		
		Classe Prevista		TCC	Classe Prevista		TCC	Classe Prevista		TCC
		E	NE	100%	E	NE	91%	E	NE	82%
E	12	12	-		11	1		10	2	
NE	10	-	10		1	9		2	8	
Conjunto de Previsão										
Classe Verdadeira	N	LDA/SPA (17) - Cinza + HSI			LDA/SW (16)* - S			PLS-DA (3) - HSI		
		Classe Prevista		TCC	Classe Prevista		TCC	Classe Prevista		TCC
		E	NE	91%	E	NE	100%	E	NE	100%
E	6	5	1		6	-		6	-	
NE	5	-	5		-	5		-	5	

*valor de limiar igual a 0.99

Observa-se que para todas as oleaginosas, diferentes resultados de classificação foram obtidos a partir de diferentes componentes de cores. Os resultados dos melhores modelos em termos de TCC para o conjunto completo de amostras, considerando-se o número de erros totais, são resumidos na **Tabela 4.7**.

Tabela 4.7. TCC obtida pelo melhor modelo de classificação (em cada oleaginosa) aplicado ao conjunto completo de amostras.

Oleaginosa	Modelo	Componentes de cores	TCC
Canola	LDA/SPA	HSI	93%
Milho	LDA/SW	HSI	100%
Girassol	LDA/SPA	Cinza + HSI	97%
Soja	LDA/SPA	Cinza + HSI	94%

Na **Figura 4.11a-d** são apresentados os gráficos dos escores obtidos pelos melhores modelos LDA aplicados ao conjunto completo de amostras.

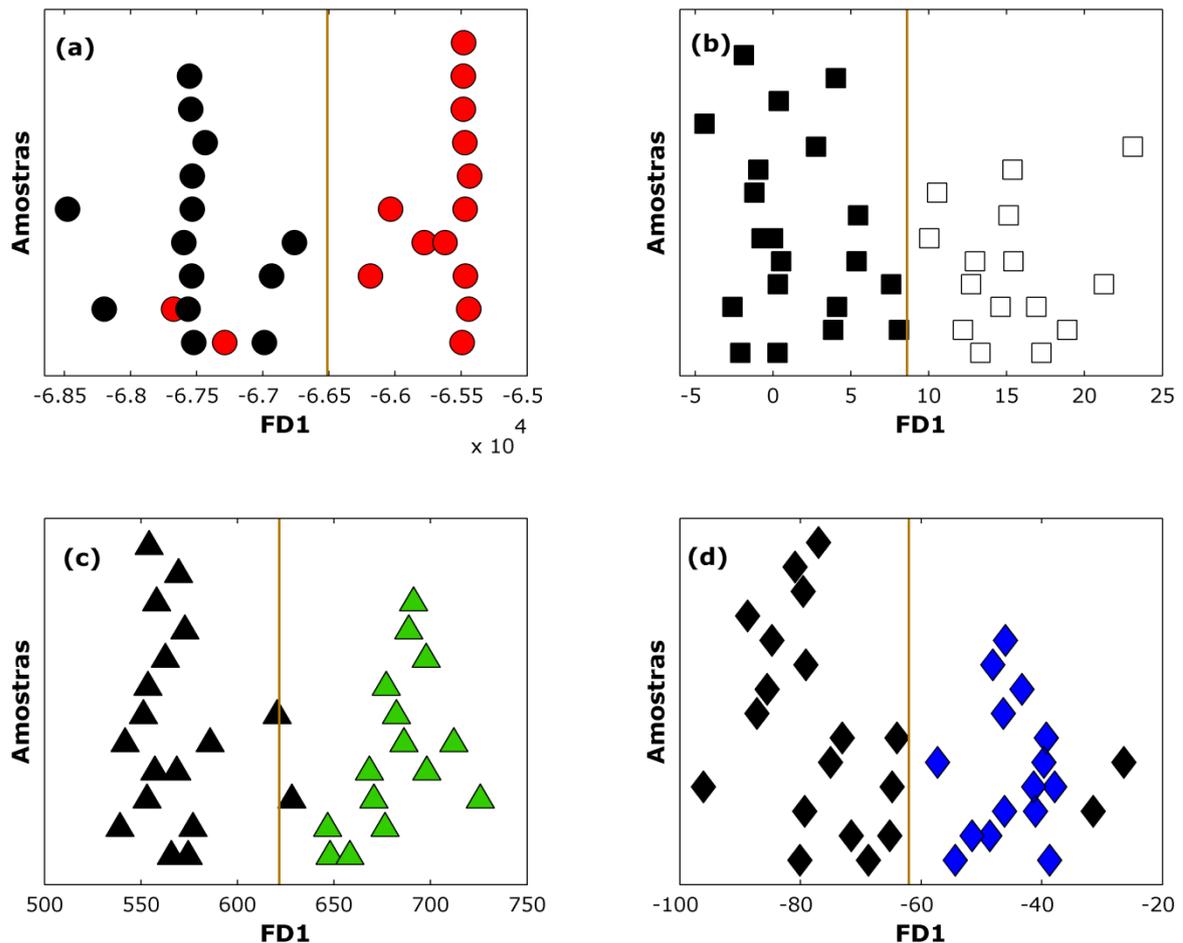


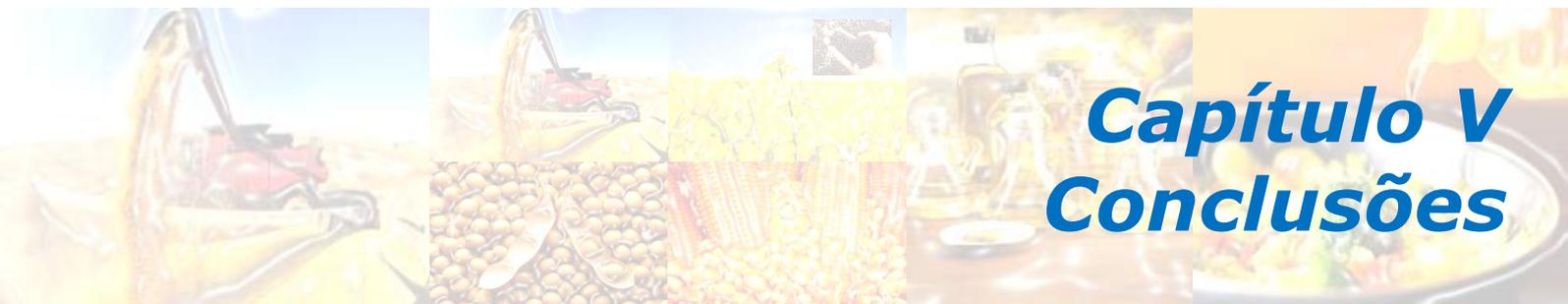
Figura 4.11. Gráficos dos escores da função discriminante 1 (FD1) para as amostras expiradas e não expiradas de: (a) canola, nas componentes HSI, (b) milho, nas componentes HSI, (c) girassol, na combinação de componentes tons de cinza + HSI e (d) soja, na combinação de componentes tons de cinza + HSI. Em todos os casos, as amostras expiradas são representadas pelos símbolos preenchidos na cor preta.

Como pode ser visto na **Figura 4.11**, todas as amostras não expiradas apresentaram valores de escores mais positivos, quando comparados às amostras expiradas.

Na **Figura 4.11a** as amostras de canola não expiradas classificadas como expiradas apresentam-se dentro do prazo de validade estabelecido pelo fabricante, restando ainda um mês para expiração da validade de

uma amostra e 4 meses para a outra. Entretanto, pode-se observar que os processos degradativos já foram iniciados e, como resultado, a alteração na cor do óleo pôde ser percebida. Este fato pode está relacionado à composição do óleo de canola, que apresenta altos teores de ácido linolênico sendo, portanto, mais suscetível à deterioração pelo processo de rancidez autoxidativa (MORETTO e FETT, 1998).

Além disso, é possível observar que, na maioria dos casos (**Fig. 4.11c-d**), os erros ocorrem nas amostras expiradas sendo classificadas como não expiradas. Embora as amostras de girassol classificadas erroneamente (**Fig. 4.11c**) tenham se expirado há aproximadamente 18 meses e as amostras de soja (**Fig. 4.11b**) há 24 meses, possivelmente, nesses casos, o processo de degradação sofrido por tais amostras não alterou substancialmente sua cor a ponto dessa variação ser detectada pelo método proposto.



Capítulo V

Conclusões

5 CONCLUSÕES

Neste trabalho foi apresentada uma metodologia baseada na utilização de imagens digitais e técnicas de reconhecimento de padrões para classificação de óleos vegetais comestíveis com respeito ao tipo e ao estado de conservação.

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 LDA (com base nas técnicas de seleção de variáveis SPA e SW) e, para efeito de comparação, PLS-DA.

Na primeira abordagem (classificação com respeito ao tipo), o modelos LDA/SPA e PLS-DA apresentaram desempenho similar para o conjunto de previsão, com uma TCC de 95%, enquanto o LDA/SW apresentou um desempenho inferior de 90%. Em termos de TCC para o conjunto completo de amostras, os modelos LDA/SPA, LDA/SW e PLS-DA apresentaram 95%, 93% e 82%, respectivamente. Em todos os casos, os melhores resultados de classificação foram alcançados a partir da componente S.

Para a segunda abordagem (classificação com respeito ao estado de conservação), a classificação foi realizada para cada tipo de oleaginosa e o modelo LDA/SPA apresentou melhor desempenho para as amostras de óleos de canola, girassol e soja, com uma TCC de 93%, 97% e 94%, respectivamente, enquanto que o modelo LDA/SW conseguiu classificar corretamente 100% das amostras de milho.

A metodologia proposta apresentou, como principal vantagem, a não utilização de reagentes e a rapidez com que as análises foram realizadas. Além disso, não foi necessária a realização de procedimentos laboriosos para caracterização química das amostras de óleos vegetais.

5.1 Perspectivas

Embora bons resultados de classificação tenham sido obtidos, as imagens digitais não fornecem informação acerca dos processos químicos que ocorrem no sistema em investigação. Entretanto, os resultados obtidos para a classificação de óleos com respeito ao tipo (**seção 4.2**) revelam que as imagens podem fornecer indícios dos fenômenos que ocorrem em um dado material e seus reflexos na mudança de sua coloração. Dessa forma, apresentam-se as seguintes propostas para trabalhos futuros:

- Utilizar os métodos clássicos (índice de acidez, peróxido e iodo) usualmente empregados para controle da qualidade dos óleos, com o propósito de comparar os resultados assim obtidos com os resultados fornecidos pelas imagens digitais;
- Realizar análises espectroscópicas para acompanhar transformações moleculares e relacioná-las às mudanças de coloração;
- Avaliar a viabilidade da utilização de imagens através da comparação dos resultados obtidos pelos métodos clássicos e espectroscópicos.



Capítulo VI
Referências Bibliográficas

REFERÊNCIAS BIBLIOGRÁFICAS

ABOISSA. Composição Química da Semente de Girassol. Disponível em: <http://www.aboissa.com.br/produtos/view/313/oleo_de_girassol>. Acessado em: 6 de maio de 2013.

ALEXANDRE, D.S.; TAVARES, J.M.R.S. Factores da percepção visual humana na visualização de dados, CMNE/CILAMCE, © APMTAC, Portugal 2007.

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*, **109: 170, 2013**.

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*, **515: 3, 2004**.

ANVISA. Resolução RDC nº 270, de 22 de setembro de 2005. Disponível em: <<http://s.anvisa.gov.br/wps/s/r/eig>>. Acessado em: 27 de agosto de 2013.

ANVISA. Resolução nº 482, de 23 de setembro de 1999. Disponível em: <http://www.anvisa.gov.br/legis/resol/482_99.htm>. Acessado em: 6 de maio de 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>>. Acessado em 01 de agosto de 2013.

BEEBE, K.R.; PELL, R.J.; SEASHOLTZ, M.B. ***Chemometrics: a practical guide***, New York: John Wiley and Sons, Inc., 1998.

BORTOLETO, G.G.; PATACA, L.C.M.; BUENO, M.I.M.S. A new application of X-ray scattering using principal component analysis – classification of vegetable oils, *Analytica Chimica Acta*, **539: 283, 2005**.

BRAMLEY, P.M.; ELMADFA, I.; KAFATOS, A.; KELLY, F.J.; MANIOS, Y.; ROXBOROUGH, H.E.; SCHUCH, W.; SHEEHY, P.J.A.; WAGNER, K-H. Review Vitamin E, *Journal of the Science of Food and Agriculture*, **80: 913, 2000**.

BRANDÃO, P.A.; COSTA, F.G.P.; BARROS, L.R.; NASCIMENTO, G.A.J. Ácidos graxos e colesterol na alimentação humana, *Agropecuária Técnica*, **26: 5, 2005**.

BROWN, T.L.; LEMAY Jr, H.E.; BURSTEN, B.E.; BURDGE, J.R. ***Química: a ciência central***. Tradução de Robson Mendes Matos. 9ª Ed. São Paulo: Pearson Prentice Hall, 2005. 972p. Título original: Chemistry – The central Science.

CAMPESTRE. Óleo de milho. Disponível em: <<http://www.campestre.com.br/oleo-de-milho.shtml>>. Acessado em: 6 de maio de 2013.

CAMPESTRE. Óleo de girassol. Disponível em: <<http://www.campestre.com.br/oleo-de-girassol.shtml>>. Acessado em: 1 de agosto de 2013.

CAMPESTRE. Óleo de soja. Disponível em: <<http://www.campestre.com.br/oleo-de-soja.shtml>>. Acessado em: 6 de maio de 2013.

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*, **70: 344, 2006.**

CASTRO, A.A.M.; PRADO, P.P.L. Algoritmos para reconhecimento de padrões, *Revista Ciências Exatas*, Taubaté, **v.5-8: 129, 1999 - 2002.**

CATALDO, A.; PIUZZI, E.; CANNAZZA, G.; DE BENEDETTO, E. Classification and adulteration control of vegetable oils based on microwave reflectometry analysis, *Journal of Food Engineering*, **112: 338, 2012.**

CHAU, F-T.; LIANG, Y-Z.; GAO, J.; SHAO, X-G. ***Chemometrics From Basics to Wavelet Transform***, vol. 164, New Jersey: John Wiley & Sons, Inc., Publication, 2004.

CONAB. Girassol. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/11_08_05_12_24_50_girassoljulho2011..pdf>. Acessado em: 24 de agosto de 2013.

CONCHA-HERRERA, V.; LERMA-GARCÍA, M.J., HERRERO-MARTÍNEZ, J.M.; SIMÓ-ALFONSO, E.F. Classification of vegetable oils according to their botanical origin using amino acid profiles established by High Performance Liquid Chromatography with UV-vis detection: A first approach, *Food Chemistry*. **120: 1149, 2010.**

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 screen teas, *Analytical Methods*, **4: 2648, 2012.**

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

EMBRAPA. Demanda por energia e alimento deve garantir bons negócios para a soja. Disponível em: <<http://www.embrapa.br/imprensa/noticias/2007/julho/5asemana/noticia.2007-08-01.0199136630/?searchterm=produ%C3%A7%C3%A3o%20mundial%20de%20soja#>>. Acessado em 1 de julho de 2013.

FAO. Prospects by Major Sector. Disponível em: <<http://www.fao.org/docrep/004/y3557e/y3557e08.htm>>. Acessado em 01 de agosto de 2013.

FERRARI, R. A.; OLIVEIRA, V. S.; SCABIO, A. Biodiesel de soja – taxa de conversão em ésteres etílicos, caracterização físico-química e consumo em gerador de energia, *Química Nova*, **28 (1): 19, 2005.**

FRASER, B.; MURPHY, C.; BUNTING, F. ***Real World Color Management***. 2nd Edition. United States of America: Peachpit Press, 2005.

GAMBARRA-NETO, F.F. *Classificação de óleos vegetais utilizando voltametria de onda quadrada e métodos quimiométricos*. João Pessoa, Programa de Pós-Graduação em Química, UFPB, 2008. Dissertação de Mestrado, 98p.

GAMBARRA-NETO, F.F.; MARINO, G.; ARAÚJO, M.C.U.; GALVÃO, R.K.H.; PONTES, M.J.C.; MEDEIROS, E.P.; LIMA, R.S. Classification of edible vegetable oils using square wave voltammetry with multivariate data analysis, *Talanta*, **77: 1660, 2009.**

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*, **31: 1485, 2008.**

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.

GUNSTONE, F.D. ***Vegetable Oils. In: Bailey's Industrial Oil and Fat Products.*** 6th Edition. Six Volume. Canada: John Wiley & Sons, Inc., 6, 213-267, 2005.

JAVIDNIA, K.; PARISH, M.; KARIMI, S.; HEMMATEENEJAD, B. Discrimination of edible oils and fats by combination of multivariate pattern recognition and FT-IR spectroscopy: A comparative study between different modeling methods, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **104: 175, 2013.**

KAMM, W.; DIONISI, F.; HISCHENHUBER, C.; ENGEL, K-H. Authenticity assessment of fats and oils, *Food Reviews International*, **17(3): 249, 2001.**

KENNARD, R. W.; STONE, L. A., *Computer-aided design of experiments*, Technometrics. **11: 137, 1969.**

LEÃO, A.C. *Gerenciamento de cores para imagens digitais*. Belo Horizonte, Programa de Pós-Graduação em Artes Visuais, UFMG, 2005. Dissertação de mestrado, 135p.

LEARDI, R. Genetic algorithms in chemistry, *Journal of Chromatography A*, **1158: 226, 2007**.

LERMA-GARCÍA, M.J.; RAMIS-RAMOS, G.; HERRERO-MARTÍNEZ, J.M.; SIMÓ-ALFONSO, E.F. Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy, *Food Chemistry*, **118: 78, 2010**.

LIZHI, H.; TOYODA, K.; IHARA, I. Discrimination of olive oil adulterated with vegetable oils using dielectric spectroscopy, *Journal of Food Engineering*, **96: 167, 2010**.

LUNA, A.S.; SILVA, A.P.; FERRÉ, J.; BOQUÉ, R. Classification of edible oils and modeling of their physico-chemical properties by chemometric methods using mid-IR spectroscopy, *Spectrochimica Acta Part A*, **100: 109, 2013**.

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

MCMURRY, J. ***Química Orgânica***, combo; Tradução All Tasks; revisão técnica Robson Mendes Matos. São Paulo: Cengage Learning, 2011.

MELCHIADES, F.G.; BOSCHI, A.O. Cores e Tonalidades em Revestimentos Cerâmicos, *Cerâmica Industrial*, **4: 1, 1999**.

MENDOZA, F.; DEJMEK, P.; AGUILERA, J.M. Color and image texture analysis in classification of comercial potato chips, *Food Research International*, **40: 1146, 2007.**

MENDOZA, F.; VALOUS, N.A.; ALLEN, P.;KENNY, T.A., WARD, P.; SUN, D.W. Analysis and classification of commercial ham slice images using directional, *Meat Science*, **81: 313, 2009.**

MOITA NETO, J.M.; MOITA, G.C. Uma introdução à análise exploratória de dados multivariados, *Química Nova*, **21 (4): 467, 1998.**

MORETTO, E.; FETT, R. ***Tecnologia de Óleos e Gorduras Vegetais na Indústria de Alimentos.*** São Paulo: Varela Editora e Livraria, 1998.

O'BRIEN, R.D. ***Fat and Oils: Formulating and Processing for Applications.*** 3rd Edition. Boca Raton: Taylor and Francis Group, 2009.

OTTO, M. ***Chemometrics: statistics and computer application in analytical chemistry***, New York: Wiley-VCH, 1999.

PEERBHAY, K.Y.; MUTANGA, O.; ISMAIL, R. Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu-Natal, South Africa, *ISPRS Journal of Photogrammetry and Remote Sensing*, **79: 19, 2013.**

PONTES, M.J.C. *Algoritmo das Projeções Sucessivas para Seleção de Variáveis Espectrais em Problemas de Classificação.* João Pessoa, Programa de Pós-Graduação em Química, UFPB, 2009. Tese de Doutorado, 123p.

PONTES, M.J.C.; GALVÃO, R.K.H.; ARAÚJO, M.C.U.; MOREIRA, P.N.T.; NETO, O.D.P.; JOSÉ, G.E.; SALDANHA, T.C.B. The successive projections algorithm for spectral variable selection in classification problems, *Chemometrics and Intelligent Laboratory Systems*, **78: 11, 2005.**

PONTES, M.J.C.; GOMES, A.A.; GALVÃO, R.K.H.; ARAÚJO, M.C.U. Internal and external validation in SPA-LDA: A comparative study involving diesel/biodiesel blends, *NIR news*, **23 (5): 6, 2012.**

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*, **85: 2159, 2011.**

POULLI, K.I.; MOUSDIS, G.A.; GEORGIU, C.A. Classification of edible and lampante virgin olive oil based on synchronous fluorescence and total luminescence spectroscopy, *Analytica Chimica Acta*, **542: 151, 2005.**

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.**

RAMALHO, V.C.; JORGE, N. Antioxidantes utilizados em óleos, gorduras e alimentos gordurosos, *Química Nova*, **29 (4): 755, 2006.**

REDA, S.Y.; CARNEIRO, P. I. B. Óleos e gorduras: Aplicações e implicações, *Revista Analytica*, **27:60, 2007.**

SHAHIDI, F.; ZHONG, Y. ***Lipid Oxidation: Measurement Methods. In: Bailey's Industrial Oil and Fat Products.*** 6th Edition. Six Volume. Canada: John Wiley & Sons, Inc., 8, 357-385, 2005.

SIKORSKA, E.; GÓRECKI, T.; KHMELINSKII, I.V.; SIKORSKI, M.; KOZIOL, J. Classification of edible oils using synchronous scanning fluorescence spectroscopy, *Food Chemistry*, **89: 217, 2005.**

SILVA JUNIOR, C.; SASSON, S. ***Biologia 1.*** 7^a Ed. São Paulo: Atual, 1991.

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*, **93: 129, 2012.**

SILVA, F.A.M.; BORGES, M.F.M.; FERREIRA, M.A. Métodos para avaliação do grau de oxidação lipídica e da capacidade antioxidante, *Química Nova*, **22 (1): 94, 1999.**

SILVA, P.R.; REGINAL, J.B; MURILLO, M. A importância da qualidade dos produtos para manter a competitividade das organizações, *Revista Científica Eletônica de Administração*, **10: 6, 2006.**

SILVA, T.C.O.; OLIVEIRA, A.E. Análise de imagens para classificação de amostras de cervejas através de análise de componentes principais (ACP). Anais eletrônicos do III Seminário de Pós-Graduação da UFG, Goiânia, 2006, 3, n.p.

SINELLI, N.; CERRETANI, L.; DI EGIDIO, V.; BENDINI, A.; CASIRAGHI, E. Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as

a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity, *Food Research International*, **43: 369, 2010.**

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.

TINOCO, S.M.B.; SICHIERI, R.; MOURA, A.S.; SANTOS, F.S.; DO CARMO, M.G.T.; Importância dos ácidos graxos essenciais e os efeitos dos ácidos graxos trans do leite materno para o desenvolvimento fetal e neonatal, *Cad. Saúde Pública*, **23(3): 525, 2007.**

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*, **16: 417, 2012.**

VICINI, L. *Análise multivariada da teoria à prática.* Santa Maria, Especialização, UFSM, 2005. Monografia, 215p.

YANG, H.; IRUDAYARAJ, J.; PARADKAR, M.M. Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy, *Food Chemistry*, **93: 25, 2005.**

ZHENG, C.; SUN, D.W.; ZHENG, L. Recent developments and applications of image features for food quality evaluation and inspection – a review, *Trends in Food Science and Technology*, **17: 642, 2006.**