



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

DISSERTAÇÃO DE MESTRADO

**CLASSIFICAÇÃO DE VINHOS PRODUZIDOS NO VALE DO SÃO FRANCISCO
BASEADA EM IMAGENS DIGITAIS E QUIMIOMETRIA**

Carlos Monteiro de Lima

SAPIENTIA AEDIFICAT

João Pessoa – PB – Brasil

Novembro/2017



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

DISSERTAÇÃO DE MESTRADO

**CLASSIFICAÇÃO DE VINHOS PRODUZIDOS NO VALE DO SÃO FRANCISCO
BASEADA EM IMAGENS DIGITAIS E QUIMIOMETRIA**

Carlos Monteiro de Lima

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

Orientador: Prof. Dr. MÁRIO CESAR UGULINO DE ARAÚJO

Coorientador: Prof. Dr. ADRIANO GOMES DE ARAÚJO

João Pessoa – PB – Brasil

Novembro/2017

Catálogo na publicação
Seção de Catalogação e Classificação

L732c Lima, Carlos Monteiro de.

Classificação de vinhos produzidos no Vale do São Francisco baseada em imagens digitais e quimiometria / Carlos Monteiro de Lima. - João Pessoa, 2017.
113 f. : il.

Orientação: Mário Cesar Ugulino de Araújo.

Coorientação: Adriano Gomes de Araújo.

Dissertação (Mestrado) - UFPB/CCEN.

1. Química. 2. Quimiometria. 3. Imagens digitais. 4. Vinho - Fabricação - Vale do São Francisco. I. Araújo, Mário Cesar Ugulino de. II. Araújo, Adriano Gomes de. III. Título.

UFPB/BC

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.

Aprovada pela banca examinadora:



Prof. Dr. Mário César Ugulino de Araújo
Orientador/Presidente

Dr. Wellington da Silva Lyra
Examinador



Profa. Dra. Kátia Messias Bichinho
Examinadora

Para minha amada Mical e nossas amadas e
desejadas filhas, Ana Carla e Tarsila.

Pelas contribuições para a realização deste trabalho, com todo o carinho, agradeço:

A minha Mãe e o meu Pai, Isaura e Casimiro, pessoas simples e de pouco letramento que me ensinaram os valores da honestidade, do respeito e da valorização das pessoas e a crença no amor, pilares da minha vida.

A minha amada Mical, por todo o esforço realizado no cuidado com nossas filhas enquanto eu me fazia distante e por todo o apoio para que eu pudesse desenvolver meus estudos e por me inspirar a ir mais longe todos os dias. Te amo!

As minhas filhas, Ana Carla e Tarsila. Por vocês eu irei sempre mais longe!

Os Professores Dr. Mário Cesar Ugulino de Araújo e Dr. Adriano Gomes de Araújo, por aceitarem o desafio de orientar a mim, um sujeito fora da trajetória acadêmica convencional, retomando a formação acadêmica catorze anos após concluir a graduação.

O Dr. David. Amigo, sem sua colaboração/orientação e seus incentivos eu não teria ido tão longe.

@s amig@s Arquimedes, Kelly, Matheus e Willy – “Mestrados 2016.1”, pela excelente convivência e pelos incentivos.

O amigo Adenilton, porque suas dicas foram muito valiosas para este trabalho.

O Dr. Giuliano Elias Pereira, *expert* em enologia e viticultura, pela parceria na realização deste trabalho.

Licarion Neto, que me recebeu no LAQA e me prestou importantes ensinamentos.

O LAQA/UFPB, incluindo seus colaboradores, por fornecer toda a estrutura (física e intelectual) para que eu desenvolvesse esta pesquisa: o LAQA vai comigo, onde quer que vá!

O cunhado/irmão João Marcos, pelos incentivos e inspiração e pela revisão deste texto.

O cunhado/irmão Micael, pelo apoio na construção do DMCbid.

As vinícolas Santa Maria, Botticelli, Bianchetti-Tedesco e Garziera, a EMBRAPA-Semiárido e o IF SERTÃO-PE. Suas doações foram fundamentais para o desenvolvimento deste trabalho.

As amigas Silvia e Vanessa, pelas consultorias acerca de cores, design, fotografia e mais.

A Capes, pelo financiamento.

Este trabalho, embora verse sobre enologia, não faz de mim enólogo,
continuo enófilo. Pois, como afirmou Luiz Groff, enólogo é um
sujeito que diante do vinho toma decisões, enquanto o enófilo, diante
de decisões, toma vinho.

RESUMO

Este trabalho teve como objetivo propor uma metodologia simples, rápida e de baixo custo, baseada em imagens digitais e em técnicas de reconhecimento de padrões, para classificação de vinhos produzidos no Vale do São Francisco, região vitivinícola localizada no nordeste brasileiro, contribuindo para o processo de certificação de indicação geográfica destes vinhos. Foram usadas amostras de vinhos tintos, finos e tranquilos, comprados em supermercados de Natal/RN e de João Pessoa/PB ou doadas por vitivinicultores e instituições de pesquisa localizadas na Região do Vale do São Francisco (VSF). Três abordagens de classificação foram consideradas, tomando como referência a origem geográfica (com duas classes: VSF e Mundo), os vitivinicultores (com três viticultores do VSF) e as variedades de uvas (Cabernet Sauvignon, Syrah e Touriga Nacional). Cem amostras foram usadas na abordagem de classificação em função da origem geográfica dos vinhos e os modelos apresentaram Taxa de Classificação Correta (TCC) do conjunto de teste de 80,7% e 93,6%, respectivamente, com os dados RGB e HSI e modelagem com SPA-LDA e de 61,3% e 83,9%, respectivamente, com os dados RGB e HSI e modelagem com PLS-DA. Na abordagem de classificação em função dos vitivinicultores, 70 amostras foram usadas e os valores de TCC dos conjuntos de teste foram, igualmente para os dados RGB e HSI, de 100% e 95,5%, respectivamente, para os modelos construídos com SPA-LDA e PLS-DA. Na abordagem de classificação considerando-se as variedades de uvas presentes na composição dos vinhos varietais, 48 amostras foram usadas e os modelos construídos com SPA-LDA e PLS-DA obtiveram TCC de 100% nos conjuntos teste para os pares Cabernet Sauvignon *versus* Touriga Nacional e Syrah *versus* Touriga Nacional, tanto com os dados RGB quanto com os dados HSI. Na comparação entre as Cabernet Sauvignon *versus* Syrah, os modelos SPA-LDA alcançaram iguais resultados com os dados RGB e HSI, com TCC de 72,7%, para os conjuntos teste, enquanto os modelos obtidos com PLS-DA alcançaram TCC iguais 72,7% e 81,8% nos conjuntos de teste, respectivamente, com os dados RGB e HSI. Os resultados demonstram a viabilidade do uso imagens digitais associadas a ferramentas quimiométricas para classificação de vinhos em função de sua origem geográfica, vitivinicultor e composição varietal, de forma simples, rápida, com baixo consumo de amostras, sem usar qualquer pré-tratamento, reagentes químicos ou diluição das amostras e com baixa geração de resíduos.

Palavras-chave: Vinho. Origem geográfica. Imagens digitais. Quimiometria. Algoritmo das Projeções Sucessivas.

ABSTRACT

This work proposes a simple, fast and low-cost methodology based on digital images and pattern recognition techniques for the classification of wines produced in the São Francisco Valley (VSF), a wine region located in the Brazilian Northeast, contributing to the certification process of the wines geographical indication. Red wines samples, purchased from Natal/RN and João Pessoa/PB supermarkets or donated by winegrowers and research institutions located in the Region of the São Francisco Valley were used. Three classification approaches were considered, taking as reference the geographical origin (with two classes: VSF and Word), winegrowers (with three VSF winegrowers) and grape varieties (Cabernet Sauvignon, Syrah and Touriga Nacional). In the classification approach according to the geographical origin of the wines, a hundred samples were used, and the models presented Correct Classification Rate (CCR) of the test set of 80.7% and 93.6%, respectively, with the RGB and HSI data and modeling with SPA-LDA and 61.3% and 83.9%, respectively, with the RGB and HSI data and PLS-DA modeling. In the classification approach as a function of winegrowers, seventy samples were used and the CCR values of the test sets were equal, with RGB and HSI data, 100% and 95.5%, respectively, for the models constructed with SPA-LDA and PLS-DA. In the classification approach using the varieties of grapes present in the composition of varietal wines, forty-eight samples were used, and the models constructed with SPA-LDA and PLS-DA obtained CCR values equal to 100% in the test sets for the Cabernet Sauvignon versus Touriga Nacional and Syrah versus Touriga Nacional, both with the RGB and HSI data. In the comparison between Cabernet Sauvignon versus Syrah, the SPA-LDA models achieved equal results with the RGB and HSI data, with CCR equal to 72.7%, for the test sets, while the models obtained with PLS-DA achieved equal CCR of 72.7% and 81.8% in the test sets, respectively, with the RGB and HSI data. The results demonstrate a feasibility of using digital images associated to chemometrics for wine classification, in a simple, fast, low sample consumption, without chemical reagents or dilution of the samples and with low generation of residues, according to their origin geographical, winegrower or varietal composition.

Keywords: Wine. Geographical origin. Digital images. Chemometrics. Successive Projections Algorithm.

LISTA DE ILUSTRAÇÕES

Figura 3.1 – Fluxograma geral das vinificações de brancos e de tintos.	24
Figura 3.2 – Estrutura do cátion flavilium.	26
Figura 3.3 – Fórmula estrutural das antocianidinas presentes em uvas e vinhos.	27
Figura 3.4 – Fórmula estrutural geral das antocianinas 3-glicosídeas (a) e 3-glicosídeas aciladas com ácido p-cumárico (b)	28
Figura 3.5 – Mapa vitivinícola mundial.	29
Figura 3.6 – Consumo de vinhos por países, em milhões de hectolitros e em percentagens, no ano de 2016.	31
Figura 3.7 – Mapa vitivinícola do Vale do São Francisco.	32
Figura 3.8 – Fotografia de um parreiral no Vale do São Francisco com quatro fases da videira.	33
Figura 3.9 – Esboço representativo da estrutura qualitativa dos vinhos produzidos no Brasil.	36
Figura 3.10 – Modalidades de sensores para aquisição de imagens bidimensionais: sensor único do tipo fotodiodo (a), arranjo de sensores em linha (b) e arranjo matricial de sensores (c).	38
Figura 3.11 – Esquema, em padrão Bayer (RGB), de filtros de cores para detectores CCD e CMOS.	39
Figura 3.12 – Espectro eletromagnético.	40
Figura 3.13 – Diagrama simplificado de um corte lateral do olho humano.	41
Figura 3.14 – Absorção da luz pelos cones azuis (a) , bastonetes (b) , cones verdes (c) e cones vermelhos (d)	42
Figura 3.15 – Representação gráfica do sistema RGB com indicações para o preto, o branco, as cores primárias e secundárias e a escala de cinza.	43
Figura 3.16 – Representação do espaço de cores HSI baseado no cone reto.	45
Figura 3.17 – Imagem de dez moedas em fundo escuro (a) e seu histograma (b)	46
Figura 3.18 – Representação genérica de dados químicos analíticos em forma de matriz	48
Figura 3.19 – Eixos das componentes principais que definem um novo conjunto de vetores de base para o espaço de medição definido pelas variáveis X, Y e Z.	50
Figura 3.20 – Representação gráfica do método PCA.	51

Figura 3.21 – Esquema de discriminação entre duas classes (A e B), baseando-se nas variáveis originais (var1 e var2) e em linha discriminante.	52
Figura 3.22 – Exemplo da seleção de variáveis com SPA.	54
Figura 3.23 – Modelo de análise discriminante de mínimos quadrados parciais (PLS-/DA) para mais de duas classes.	55
Figura 4.1 – Esboço do DMCdib e suas partes principais: a tampa fixadora do <i>smartphone</i> (a), aparelho <i>smartphone</i> (b), caixa principal (c) e gaveta com suporte para cubeta (d).	59
Figura 4.2 – Imagens de três amostras de vinho registradas pelo DMCdib.	60
Figura 4.3 – Imagem da janela de trabalho do <i>software</i> QUIMAGENS, como utilizado neste trabalho.	61
Figura 5.1 – Histogramas das imagens no sistema RGB das 100 amostras usados na modelagem de classificação de vinhos tintos em função da sua origem geográfica, após retirada das variáveis nulas. (–) VSF e (–) Mundo.	64
Figura 5.2 – Histogramas das imagens no sistema HSI das 100 amostras usados na modelagem de classificação de vinhos tintos em função da sua origem geográfica, após retirada das variáveis nulas. (–) VSF e (–) Mundo.	66
Figura 5.3 – Gráfico de escores PC1 x PC2 de classificação de vinhos em função da sua origem geográfica, baseado nos sistemas de cores RGB (a) e HSI (b). VSF (●) e Mundo (●).	67
Figura 5.4 – Gráficos dos custos de validação (G) em função do número de variáveis incluídas no modelo (a) e das variáveis selecionadas (b) pelo algoritmo SPA-LDA, com dados RGB.	68
Figura 5.5 – Gráficos dos custos de validação (G) em função do número de variáveis incluídas no modelo (a) e das variáveis selecionadas (b) pelo algoritmo SPA-LDA com dados HSI.	68
Figura 5.6 – Gráficos dos escores das Funções Discriminantes (DF) calculadas com o SPA-LDA, usando os dados de cores RGB (a) e HSI (b), para classificação de vinhos tintos em função da origem geográfica: VSF (■) e Mundo (■).	69
Figura 5.7 – Histogramas de cores das 70 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função dos seus produtores, após seleção a priori de variáveis, com dados RGB. Classes 1 (–), 2 (–) e 3 (–).	72

Figura 5.8 – Histogramas de cores das 70 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função dos seus produtores, após seleção <i>a priori</i> de variáveis, com dados HSI. Classes 1 (–), 2 (–) e 3 (–).	73
Figura 5.9 – Gráfico de escores PC1 \times PC2 de classificação de vinhos do Vale do São Francisco em função dos seus produtores, baseado no sistema de cores RGB (a) e HSI (b). Classes 1 (●), 2 (●) e 3 (●).	74
Figura 5.10 – Gráficos dos custos de validação (G) em função do número de variáveis (a) e das variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras (b), usando dados RGB.	75
Figura 5.11 – Gráficos dos custos de validação (G) em função do número de variáveis (a) e das variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras (b), usando dados HSI.	75
Figura 5.12 – DF SPA-LDA, com sistema de cores RGB, para classificação de vinhos do Vale do São Francisco em função dos seus produtores, baseado no sistema de cores RGB (a) e HSI (b). Classes 1 (■), 2 (■) e 3 (■).	76
Figura 5.13 – Histogramas de intensidades de cores RGB das 34 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função das variedades das uvas, Cabernet Sauvignon (–) vs Touriga Nacional (–), com dados RGB (a) e HSI (b), após seleção <i>a priori</i> de variáveis.	79
Figura 5.14 – Gráficos dos escores de PCA baseado nos dados de cores RGB (a) e HSI (b) para a classificação de vinhos do Vale do São Francisco em função das variedades das uvas Cabernet Sauvignon (●) vs Touriga Nacional (●).	80
Figura 5.15 – Gráficos das variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras: RGB (a) e HSI (b).	81
Figura 5.16 – Gráficos das Funções Discriminantes (DF) da classificação de vinhos do Vale do São Francisco em função das variedades das uvas baseado nos sistemas de cores RGB (a) e HSI (b), para as variedades Cabernet Sauvignon (■) e Touriga Nacional (■).	82
Figura 5.17 – Histogramas de cores das 30 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função das variedades das uvas Syrah (–) vs Touriga Nacional (–), baseado nos sistemas de cores RGB (a) e HSI (b), após seleção <i>a priori</i> de variáveis.	83

Figura 5.18 – Gráfico de escores PC1 \times PC2 da classificação de vinhos do Vale do São Francisco em função das variedades das uvas Syrah (●) e Touriga Nacional (●), baseado no sistema de cores RGB (a) e HSI (b).	84
Figura 5.19 – Gráficos das variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras, dos dados RGB (a) e HSI (b).	85
Figura 5.20 – Gráfico das Funções Discriminantes (DF) da classificação de vinhos do Vale do São Francisco em função das variedades das uvas, para as variedades Syrah (■) e Touriga Nacional (■), baseado nos sistemas de cores RGB (a) e HSI (b).	85
Figura 5.21 – Histogramas de cores das 32 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função das variedades das uvas Cabernet Sauvignon (–) vs Syrah (–), baseado no sistema de cores RGB (a) e HSI (b), após seleção <i>a priori</i> de variáveis.	87
Figura 5.22 – Gráfico de escores de PCA (PC1 \times PC2) da classificação de vinhos do Vale do São Francisco em função das variedades das uvas Cabernet Sauvignon (●) vs Syrah (●), baseado no sistema de cores RGB (a) e HSI (b).	88
Figura 5.23 – Variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras dos sistemas RGB (a) e HSI (b).	88
Figura 5.24 – Gráfico das Funções Discriminantes (DF) da classificação de vinhos do Vale do São Francisco em função das variedades das uvas, para as variedades Cabernet Sauvignon (■) e Syrah (■), baseado nos sistemas de cores RGB (a) e HSI (b).	89

LISTA DE TABELAS

Tabela 3.1 – Produção de vinhos, por países, em milhões de hectolitros e em porcentagens, no ano 2016.	30
Tabela 4.1 – Quantidades de amostras, em função das classes levantadas, em cada uma das abordagens de classificação utilizadas neste trabalho.	58
Tabela 4.2 – Quantidades de amostras dos conjuntos de treinamento e de teste.	63
Tabela 5.1 – Matriz de confusão dos modelos de classificação construídos com os algoritmos SPA-LDA e PLS-DA, baseados em dados de imagens digitais nos sistemas de cores RGB e HSI.	70
Tabela 5.2 – Matriz de confusão dos modelos de classificação construídos com os algoritmos SPA-LDA e PLS-DA para a classificação de vinhos do Vale do São Francisco, em função dos vitivinicultores e baseados em imagens digitais, em sistemas de cores RGB e HSI.	77
Tabela 5.3 – Matriz de confusão do modelo de classificação de vinhos do Vale do São Francisco em função das variedades de uvas Cabernet Sauvignon e Touriga Nacional, baseado em imagens digitais em sistemas de cores RGB e HSI e construído com o algoritmo PLS-DA.	82
Tabela 5.4 – Matriz de confusão do modelo de classificação de vinhos do Vale do São Francisco em função das variedades de uvas Syrah e Touriga Nacional, baseado em imagens digitais em sistemas de cores RGB e HSI e construído com o algoritmo PLS-DA.	86
Tabela 5.5 – Matriz de confusão dos modelos de classificação de vinhos do Vale do São Francisco em função das variedades de uvas Cabernet Sauvignon e Syrah, baseados em imagens digitais em sistemas de cores RGB e HSI.	90

LISTA DE ABREVIATURAS E SIGLAS

AAS – Espectrometria de absorção atômica
B – Componente azul do sistema de cores RGB
CCD – dispositivo de carga acoplada
CFA – filtro de cores
cm – centímetro
CMOS – semicondutor de metal-óxido complementar
CS – Cabernet Sauvignon
DF – Função discriminante
DIB – Baseado em imagens digitais
DMCdib – Dispositivo para medições colorimétricas baseado em imagens digitais
DO – Denominação de Origem
EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária
EUA – Estados Unidos da América
FN – Falso negativo
FP – Falso positivo
G – Componente verde do sistema de cores RGB
H – Componente matiz do sistema de cores HSI
HPLC – Cromatografia líquida de alta eficiência
HSI – Sistema de cores baseado na percepção humana (matiz, saturação e intensidade)
I – Componente intensidade do sistema de cores HSI
i.e. – Isto é.
IBRAVIN – Instituto Brasileiro do Vinho
ICP-AES – Espectrometria de emissão atômica com plasma indutivamente acoplado
ICP-MS – Espectrometria de massa com plasma indutivamente acoplado
IF SERTÃO-PE – Instituto Federal do Sertão Pernambucano
INPI – Instituto Nacional da Propriedade Industrial
IP – Indicação do Procedência
KS – Algoritmo Kennard-Stone
LDA – Análise Discriminante Linear
LED – Diodo emissor de luz
mhl – milhões de hectolitros, correspondente a 10^8 litros
OIV – Organização Mundial da Vinha e Vinho

PC – Componente Principal

PCA – Análise de Componentes Principais

PLS-DA - Análise Discriminante pelos Mínimos Quadrados Parciais

R – Componente vermelho do sistema de cores RGB

RGB – Modelo de cores vermelho-verde-azul

S – Componente saturação do sistema de cores HSI

SPA – Algoritmo das Projeções Sucessivas

Sy – Syrah ou Shiraz

TCC – Taxa de Classificação Correta

TN – Touriga Nacional

UFPE – Universidade Federal do Pernambuco

UNEB – Universidade do Estado da Bahia

UNIVASF – Universidade Federal do Vale do São Francisco

UPLC – Cromatografia líquida de ultra alta eficiência

UV-Vis – Espectrofotometria de radiação ultravioleta e visível

V. vinifera – *Vitis vinífera*

VINHOVASF – Instituto do Vinho do Vale do São Francisco

VN – Verdadeiro negativo

VP – Verdadeiro positivo

VSF – Vale do São Francisco

ZPV - Zona de Produção Vitivinícola

SUMÁRIO

1	INTRODUÇÃO	17
2	OBJETIVOS	19
	2.1 Objetivo geral	19
	2.2 Objetivos específicos	19
3	FUNDAMENTAÇÃO TEÓRICA	20
	3.1 Vinhos	20
	3.1.1 <i>Definição e história</i>	20
	3.1.2 <i>Classificação de vinhos</i>	22
	3.1.3 <i>Vinificação</i>	23
	3.1.4 <i>Composição do vinho</i>	26
	3.1.5 <i>Produção de vinhos no Brasil e no mundo</i>	29
	3.1.6 <i>A viticultura no Vale do São Francisco/Brasil</i>	31
	3.2 Indicações Geográficas	34
	3.3 Imagens digitais	37
	3.3.1 <i>Conceito e fundamentos</i>	37
	3.3.2 <i>Sistemas de aquisição de imagens digitais</i>	38
	3.3.3 <i>O sistema visual humano</i>	40
	3.3.4 <i>Sistemas de cores</i>	42
	3.3.4.1 <i><u>Sistema de cores RGB</u></i>	43
	3.3.4.2 <i><u>Sistema de cores HSI</u></i>	44
	3.3.4.3 <i><u>Conversão entre os sistemas RGB e HSI</u></i>	45
	3.3.5 <i>Distribuição de pixels: histogramas</i>	46
	3.4 Técnicas para o controle de autenticidade de origem geográfica de vinhos	47
	3.5 Ferramentas quimiométricas	48
	3.5.1 <i>Técnicas de reconhecimento de padrões</i>	49
4	EXPERIMENTAL	57
	4.1 Amostras	57
	4.2 Aparelho para aquisição das imagens digitais	58
	4.3 Aquisição dos dados baseados em imagens digitais	60
	4.4 Procedimentos quimiométricos e softwares	62
5	RESULTADOS E DISCUSSÃO	64

5.1 Classificação dos vinhos com base na Origem Geográfica	64
5.1.1 <i>Análise exploratória com PCA</i>	66
5.1.2 <i>Análise discriminante com SPA-LDA e PLS-DA</i>	67
5.2 Classificação com base nos vitivinicultores	71
5.2.1 <i>Análise exploratória com PCA</i>	73
5.2.2 <i>Análise discriminante com SPA-LDA e com PLS-DA</i>	74
5.3 Classificação baseada nas variedades de uvas da composição dos vinhos varietais	78
5.3.1 <i>Cabernet Sauvignon versus Touriga Nacional</i>	78
5.3.1.1 <i><u>Análise exploratória com PCA</u></i>	80
5.3.1.2 <i><u>Análise discriminante com SPA-LDA e PLS-DA</u></i>	81
5.3.2 <i>Syrah versus Touriga Nacional</i>	83
5.3.2.1 <i><u>Análise exploratória com PCA</u></i>	83
5.3.2.2 <i><u>Análise discriminante com SPA-LDA e PLS-DA</u></i>	84
5.3.3 <i>Cabernet Sauvignon versus Syrah</i>	86
5.3.3.1 <i><u>Análise exploratória com PCA</u></i>	87
5.3.3.2 <i><u>Análise discriminante com SPA-LDA e PLS-DA</u></i>	88
6. CONCLUSÕES	92
REFERÊNCIAS	94
ANEXO A – Informações indicadas nos rótulos das amostras de vinhos usadas na abordagem de classificação em função da origem geográfica.....	100
ANEXO B – Informações indicadas nos rótulos das amostras de vinhos usadas na abordagem de classificação em função dos vitivinicultores.....	105
ANEXO C – Informações indicadas nos rótulos das amostras de vinhos usadas na abordagem de classificação em função das variedades de uvas.....	109
ANEXO D – Pedido nacional de patente de invenção do Dispositivo de Medições Colorimétricas baseado em imagens digitais	112

1 INTRODUÇÃO

A produção de vinhos no mundo ocorre tradicionalmente em regiões de clima temperado, entre os paralelos 30° e 50° no Hemisfério Norte e entre 30° e 45° no Hemisfério Sul. Durante muito tempo a produção vitivinícola fora destas faixas foi vista como improvável (AMARANTE, 2015; PEREIRA, 2013; PUCKETTE; HAMMACK, 2015).

O Brasil é pioneiro na produção de vinhos finos em regiões de clima tropical, tendo iniciado a produção há cerca de 30 anos, na região do Vale do São Francisco (VSF), localizada entre 8° e 10° de Latitude Sul, às margens do rio São Francisco, entre os municípios de Remanso/BA e Sobradinho/BA, nos estados da Bahia e Pernambuco (ABS-PR, 2012; AMARANTE, 2015; COMITÊ DA BACIA HIDROGRÁFICA DO RIO SÃO FRANCISCO, 2015).

Sendo tão recente a implantação desta região produtora, e mesmo que já se tenha um esforço coletivo, inclusive havendo nas proximidades instituições de ensino e de pesquisa, como a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), a Universidade do Estado da Bahia - UNEB, a Universidade Federal do Pernambuco (UFPE), a Universidade Federal do Vale do São Francisco (UNIVASF) e o Instituto Federal do Sertão Pernambucano (IFSERTÃO-PE), a região ainda é pouco estudada, carecendo de mais estudos qualitativos e quantitativos, tanto sobre o plantio de videiras, adequando as melhores castas às características do solo e do clima, quanto aos processos de produção e, ainda, relativos à qualidade dos vinhos já produzidos ali (PEREIRA, 2013). Nesse *terroir*, é cultivado um grande número de variedades vitivinícolas, como Syrah, Cabernet Sauvignon, Alicante Bouschet, Tannat, Ruby Cabernet, Touriga Nacional, Chenin Blanc, Moscato Canelli e Sauvignon Blanc, que estão se adaptando muito bem às condições do sertão nordestino brasileiro (AMARANTE, 2015; PEREIRA, 2013; TONIETO; PEREIRA, 2011).

As Indicações Geográficas referem-se a produtos ou serviços que tenham uma origem geográfica específica, está relacionada à reputação, às qualidades e às características vinculadas a certo local, comunicam ao mundo que uma certa região se especializou e tem capacidade de produzir um artigo diferenciado e de excelência. Nos termos da Lei nº. 9.279 (BRASIL, 1996), a Indicação de Procedência é o nome geográfico de país, cidade, região ou localidade de seu território, que se tenha tornado conhecido como centro de extração, produção ou fabricação de determinado produto ou de prestação de determinado serviço, quando as qualidades ou características de tais produtos se devam exclusiva ou essencialmente ao meio geográfico, incluídos fatores naturais e humanos, a indicação geográfica será de Denominação de Origem.

Os produtores da região do VSF, com forte apoio da EMBRAPA, estão desenvolvendo esforços para a obtenção do Registro de Indicação Geográfica, o que deixa mais patente a necessidade de realização de pesquisa de qualidade dos vinhos ali produzidos. De certo, com um estudo que comprove a singularidade dos seus vinhos, a região poderá ganhar Registro de Indicação Geográfica, diretamente, como Denominação de Origem (DO).

2 OBJETIVOS

2.1 Objetivo geral

Propor uma metodologia simples, rápida e de baixo custo, baseada em imagens digitais (DIB) e em técnicas de reconhecimento de padrões, para classificação de vinhos produzidos no Vale do São Francisco, região vitivinícola localizada no nordeste brasileiro, contribuindo para o processo de certificação de indicação geográfica destes vinhos.

2.2 Objetivos específicos

- I. Estudar vinhos do Vale do São Francisco e de outras regiões do mundo, utilizando imagens digitais e comparando seus dados de intensidades nos sistemas de cores RGB e HSI;
- II. Construir um instrumento analítico baseado em imagens digitais, de baixo custo e operacionalmente simples, visando transferência de tecnologia para a indústria;
- III. Realizar um estudo exploratório empregando a análise de componentes principais (PCA) para reconhecimento de padrões não supervisionados em dados de imagens digitais que diferenciem de vinhos do Vale do São Francisco de outros produzidos noutras regiões do mundo;
- IV. Desenvolver modelos de classificação de vinhos em função de sua origem geográfica que possam ser usados para confirmar ou negar suas procedências em relação ao Vale do São Francisco, utilizando os fundamentos de análise discriminante linear com seleção de variáveis pelo algoritmo das projeções sucessivas (SPA-LDA) e análise discriminante pelos mínimos quadrados parciais (PLS-DA).

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Vinhos

3.1.1 Definição e História

Vinho é uma bebida alcoólica preparada a partir da fermentação de uvas. Embora seja possível a fermentação de inúmeras outras frutas para a obtenção de bebidas alcóolicas, apenas as uvas podem produzir a bebida reconhecida como vinho (PUCKETTE; HAMMACK, 2016). No Brasil, a produção de vinho é regulamentada pela Lei nº. 7.678, de 8 de novembro de 1988, que define vinho como “a bebida obtida pela fermentação alcoólica do mosto simples de uva sã, fresca e madura” (BRASIL, 1988). A mesma Lei define o mosto simples de uva como o “produto obtido pelo esmagamento ou prensagem da uva sã, fresca e madura, com a presença ou não de suas partes sólidas” (BRASIL, 1988). Na prática, mosto é o próprio suco da uva – que para a produção de vinhos coloridos (tintos e rosados) deve incluir as cascas e as sementes, e o vinho é o produto da fermentação desse mosto.

A uva é o fruto da videira, uma planta trepadeira da família botânica *Vitaceae* ou *Ampelidaceae* que possui mais de 3.000 espécies conhecidas. O único gênero que tem interesse na vitivinicultura é o *Vitis*, que abrange 36 espécies, dentre elas, a *Vitis vinifera* que dá origem a vinhos de qualidade. A *Vitis vinifera* tem mais de 300 variedades ou cultivares que são conhecidas como videiras europeias, mas nem todas são usadas na vinificação. Alguns países, incluindo o Brasil, cultivam e vinificam uvas de outras espécies viníferas, como *Vitis Lambrusca*, *Vitis burquina*, *Vitis aestivalis*, *Vitis riparia*, *Vitis rupestris*, conhecidas como videiras americanas, e híbridos destas com *V. vinifera* que são mais resistentes a pragas e intempéries, mas que produzem vinhos de qualidade inferior. Noutros países, como Chile, Argentina, França e grande parte da Europa, as espécies americanas só podem ser usadas como porta-enxertos para as videiras europeias (*V. vinifera*), não podendo ser plantadas para colheita direta de uvas (AMARANTE, 2015; JOHSON; ROBINSON, 2014).

Há registros da existência da vinha e do vinho que datam de mais de 7.000 anos, sendo mais provável que tenham surgido nas encostas de montanhas do Cáucaso, região localizada entre o Mar Cáspio e o Mar Negro, onde hoje se localizam os territórios da Geórgia, Armênia e Azerbaijão, e depois na Mesopotâmia, onde hoje fica o Iraque (AMARANTE, 2015; GAUTIER, 2013; JOHNSON, 2009; JOHNSON; ROBINSON, 2014). Tais registros consistem de ajuntamentos de sementes hermafroditas de uvas encontradas por arqueólogos em

escavações realizadas na Turquia, na Síria no Líbano e na Jordânia. O fato de as sementes serem hermafroditas comprovaria que o homem já possuía certo domínio da viticultura, com capacidade de selecionar cultivares, enquanto que o fato de estarem acumuladas indica que foram processadas, como numa vinificação (JOHNSON, 2009).

A vitivinicultura chegou ao Egito a partir do III Milênio a.C. Lá foram produzidos vinhos para uso em cerimônias religiosas, geralmente ligado a Osíris - o deus egípcio dos ciclos da vegetação e dos mistérios da vida após a morte (GAUTIER, 2013; TATTERSALL; DESALLE, 2015). Depois, a cultura do vinho chegou à Grécia e, por volta do ano 600 a.C, ao Império Romano que a levaram a toda a costa do Mediterrâneo. Na Grécia, onde o termo civilização se confunde com a reverência ao vinho, Dionísio era a divindade reverenciada, tida como pai e fundador da cultura da vinha e do vinho, enquanto em Roma, Baco é o deus da vinha do vinho e das festividades ou a própria imagem de Dionísio (AMARANTE, 2015; GAUTIER, 2013).

Segundo Amarante (2015), a viticultura chegou no Brasil em 1532 trazida pelos colonizadores portugueses e foi primeiramente instalada na Capitania de São Vicente, onde hoje é o Estado de São Paulo, mas não se obteve grande sucesso de adaptação das videiras. Mais tarde, com a chegada dos imigrantes italianos à Serra Gaúcha, na década de 1870, o cultivo de *Vitis vinifera* chegou ao Rio Grande do Sul, onde encontrou melhores condições de adaptação. Mas foi somente no século XIX que o cultivo de cepas europeias para produção de vinhos finos ganhou perfil comercial (AMARANTE, 2015; PROTAS; CAMARGO; MELLO, 2006). Segundo Amarante (2015), até o final da década de 1960 houve crescimento lento e gradual da vitivinicultura brasileira, sendo que na década seguinte houveram maiores avanços com a criação das cooperativas vinícolas e a chegada de empresas internacionais à Serra Gaúcha. Entre 1960 e 1970, houve também uma expansão para o extremo sul e para o nordeste do país. Na região Nordeste, as atividades vitivinícolas foram iniciadas às margens do Rio São Francisco, sendo que o primeiro vinho fino foi produzido em 1984, no município de Santa Maria da Boa Vista/PE (AMARANTE, 2015).

De acordo com Amarante (2015), desde a antiguidade o vinho era usado como medicamento, indicado como anestésico pré-operatório, antitérmico, laxante, diurético, antisséptico local, a ponto de Louis Pasteur ter a ele se referido como a bebida mais saudável e mais higiênica. Segundo ele, o consumo de vinho sempre esteve associado à boa saúde e longevidade, o que sempre fez atrair mais degustadores.

Inúmeros estudos recentes comprovam os vários efeitos benéficos do vinho à saúde humana, como os efeitos na redução de doenças cardiovasculares quando se consome de 3 a 5 doses (de 150 mL) diariamente (CVEJIĆ; GOJKOVIC-BUKARICA, 2016). Outros estudos

demonstraram que o consumo regular pode reduzir até 60% as chances de doenças cardíacas e infartos do miocárdio, prevenção de demências, inibição de câncer de mama e de próstata, de doenças reumáticas, entre outras (AMARANTE, 2015; CVEJÍC; GOJKOVIC-BUKARICA, 2016). As propriedades farmacológicas, antioxidantes e benéficas do vinho estão relacionadas à presença de diversos compostos fenólicos, particularmente o resveratrol, em sua composição (CVEJÍC; GOJKOVIC-BUKARICA, 2016).

3.1.2 Classificação de vinhos

Existem diversas maneiras de classificar vinhos e, de acordo com Amarante (2015), a forma mais universal classifica os vinhos em vinhos de mesa ou vinhos tranquilos (ou, simplesmente, vinhos), vinhos espumantes (ou, simplesmente, espumantes) e vinhos licorosos. Contudo, existem várias outras classificações baseadas nas espécies das uvas, na cor, na presença de misturas de variedades de uvas, na região de origem, entre outras.

No Brasil, a Lei Nº 7.678/1990 e o Decreto Nº. 8.198/2014 regulam a produção, circulação e comercialização do vinho e derivados da uva e do vinho. Nos termos dessas normas, os vinhos são classificados como finos se forem elaborados a partir de uvas do grupo das europeias da espécie *Vitis vinifera*, enquanto os vinhos de mesa (ou comuns) são aqueles de uvas do grupo das americanas, das espécies americanas (*Vitis lambrusca* e *Vitis bourquina*) e de suas híbridas. Importante destacar que a mesma legislação admite o uso, ainda que opcional, da expressão “de mesa” junto ao termo “fino”, reforçando a ideia que a expressão “de mesa” e o termo “tranquilo” são sinônimos.

De modo geral, quanto a sua cor, os vinhos são classificados em brancos, tintos e rosés. Os tintos são necessariamente produzidos com uvas tintas (vermelhas ou negras) e sua fermentação ocorre na presença das cascas, a fim de haver a extração da matéria corante das uvas. Os vinhos brancos podem ser obtidos de uvas brancas ou tintas, mas a fermentação ocorre na ausência das cascas e, portanto, não extraem delas os pigmentos vermelhos. Por fim, os vinhos rosés (rosados) são obtidos a partir de uvas tintas e com o contato de suas cascas com o mosto por tempo reduzido ou de modo pouco ou nada ortodoxo pode-se misturar branco e tinto, uma prática que não é incomum, especialmente em vinhos de baixo preço (ALEJANDRE, 2007; GUERRA, 2010; MANFRÓI, 2010; PICONERO, 2015).

Quanto às variedades de uvas viníferas usadas na vinificação, se um vinho apresentar, no mínimo, 75% do seu volume proveniente de uma única casta, com o restante volume provenientes de uvas da mesma espécie, poderá indicar o nome dessa variedade em seu rótulo

e será classificado como varietal. Os vinhos que apresentem cortes, isto é, misturas de duas ou mais castas na sua composição e que todas elas apresentem fração em volume inferior a 75% serão chamados genéricos ou *assemblages*. Geralmente, na vinificação dos vinhos *assemblages*, cada casta é vinificada em separado e, depois, são realizados os cortes adequados (BRASIL, 1990; PICONERO, 2015).

Recentemente, uma classificação baseada no clima da região tem se tornado bastante evidente. Assim, quando os vinhos são produzidos nas regiões tradicionais, entre as latitudes 30° e 50°, dos dois hemisférios, onde as estações do ano são bem definidas, permitindo às videiras um longo período vegetativo, são classificados como vinhos temperados ou tradicionais; ao contrário, quando são produzidos nas novas fronteiras vitivinícolas, em regiões que se afastam da região tradicional em direção à linha do Equador (paralelo 0°), onde o clima é predominantemente tropical, com pouca pluviosidade e muita insolação, os vinhos são denominados vinhos tropicais (AMARANTE, 2015; TONIETTO, 2004; TONIETTO; PEREIRA, 2012; TONIETTO, TEIXEIRA, 2004; PEREIRA, VANDERLINDE, LIMA, 2011).

3.1.3 Vinificação

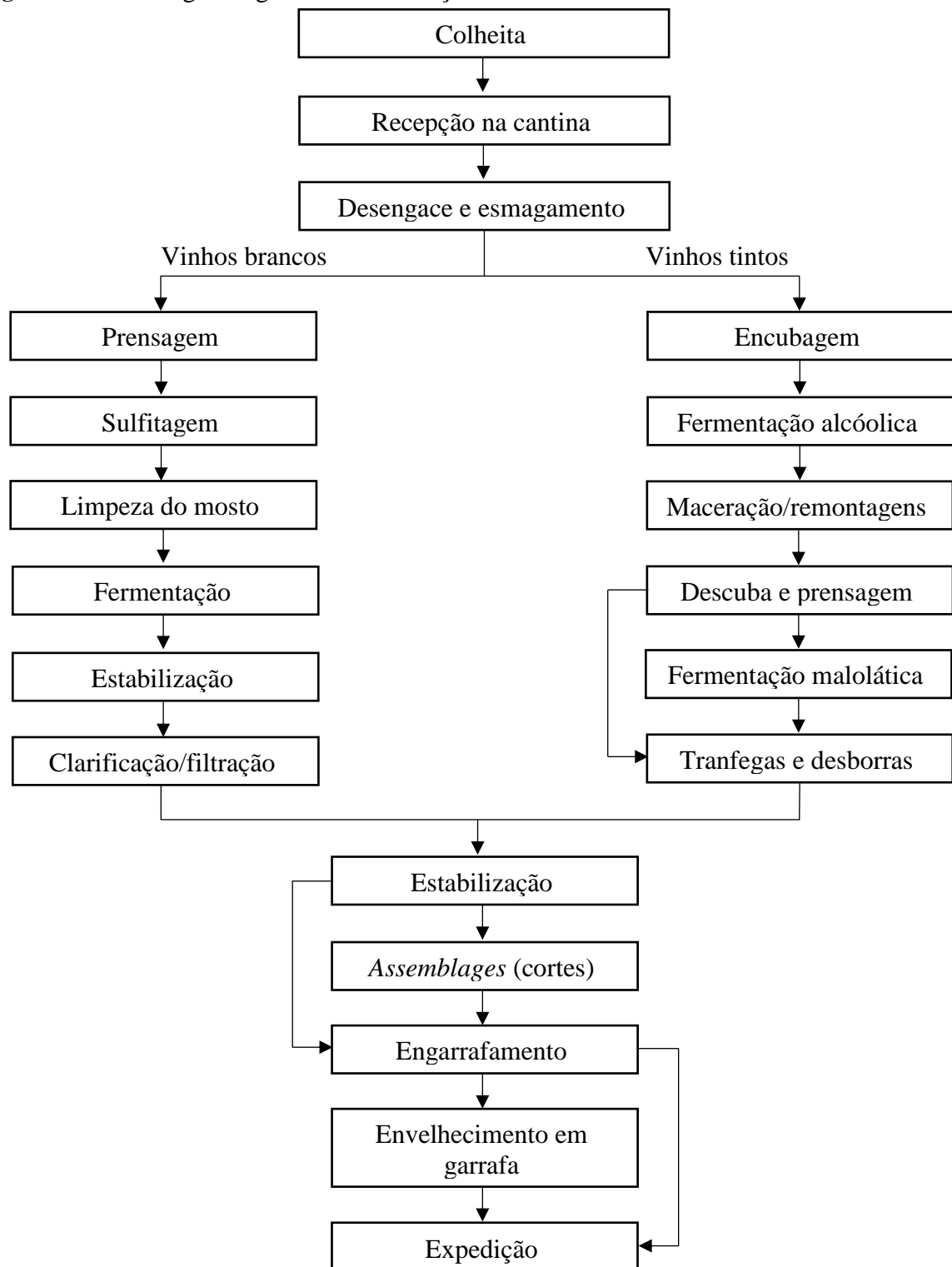
Vinificação é um processo no qual microrganismos (leveduras), mediante fermentação, transformam os açúcares do mosto de uvas selecionadas em vinho, com intervenção humana a fim de garantir que o produto obtido guarde qualidade superior, portanto, trata-se de um processo biotecnológico (GUERRA, 2010). As leveduras encontradas naturalmente nas próprias uvas já são capazes de realizar a fermentação do mosto, no entanto, tem sido frequente o uso de leveduras selecionadas, com o a finalidade de melhorar o desempenho da fermentação.

Na **Figura 3.1** é mostrado o fluxograma geral das vinificações de brancos e de tintos. Em termos gerais, a vinificação de qualquer tipo de vinho contempla as etapas de colheita (vindima), recepção das uvas na cantina, desengace, extração do mosto, fermentação, adição de anidrido sulfuroso (sulfitagem), engarrafamento e expedição. Contudo, há grandes particularidades para cada tipo de vinho, como exemplo, na vinificação de brancos as cascas são retiradas antes da fermentação, enquanto nos tintos a extração da matéria corante é fator determinante na qualidade do vinho e a fermentação ocorre na presença das cascas das uvas (MANFRÓI, 2010; GUERRA, 2010).

A colheita pode ser manual ou mecânica, se for manual, já permite uma seleção dos cachos, evitando-se uvas apodrecidas, folhas e deformidades (ALEXANDRE, 2010; GUERRA, 2010). Na recepção na cantina, há uma seleção complementar tanto dos cachos

como dos grãos, melhorando a uniformidade das uvas. O desengace também pode ser mecânico ou manual, usando-se este na elaboração de vinhos de alto padrão, e consiste na retirada das partes lenhosas que dão sustentação aos cachos, inclusive os pinceis (parte dos engaces que ficam no interior dos grãos), soltando as uvas (ALEJANDRE, 2010; GUERRA, 2010).

Figura 3.1 – Fluxograma geral das vinificações de brancos e de tintos.



Fonte: baseado em Manfrói (2010) e Guerra (2010).

A extração do mosto começa com o esmagamento dos grãos, que pode ser feito ao mesmo tempo do desengace, seguida de prensagem para separar as partes sólidas (nos brancos) ou de maceração (nos tintos) (ALEJANDRE, 2010; GUERRA, 2010). Segundo Guerra (2010, a sulfitagem visa a conservação da bebida, evitando sua oxidação e deve acompanhar todo o processo, iniciando logo após a encubagem (tintos) ou a prensagem (brancos) até o envase, podendo também ser usada para atuar na seleção de microrganismos para a fermentação. A fermentação alcoólica é a etapa em que os açúcares do mosto das uvas, essencialmente glicose e frutose, são transformados em etanol e gás carbônico, por ação de leveduras (principalmente a base de *Saccaromyces cerevisiae*) e enzimas (pectinases), conforme representado pela **Equação 1**. Em alguns vinhos, mais frequentemente nos tintos, se faz uma segunda fermentação chamada malolática, em que o ácido málico é transformado em ácido lático por ação da bactéria láctica (*Leuconostoc oenos*), representada pela **Equação 2** (ALEJANDRE, 2010; GUERRA, 2010).

Fermentação alcoólica:



Fermentação malolática:



Segundo Guerra (2010, pag. 222), a etapa de maceração “consiste na extração seletiva de certos compostos presentes nas partes sólidas da uva, que concorrem para a alta qualidade do vinho”, sendo típica dos vinhos tintos, e geralmente ocorre em paralelo à fermentação alcoólica, podendo iniciar antes ou terminar depois desta. As remontagens são operações que ocorrem durante a maceração e consistem em homogeneizar as fases líquida e sólida, a fim de aumentar a eficiência das extrações, uma vez que o gás carbônico produzido na fermentação tende a suspender a fase sólida. Terminadas a fermentação e a maceração, as fases sólida e líquida são separadas, numa etapa chamada de descuba. A fase sólida é prensada para recuperação de parte do líquido que ainda se encontra misturado nessa fase. As tranfegas consistem nas transferências do vinho entre reservatórios, deixando a matéria sólida decantada (borras), para obter um vinho mais límpido e de melhor qualidade olfativa. Em seguida, seguem-se as etapas de estabilização, em barris de madeira ou tanques de inox, que leva à precipitação de ácidos orgânicos sob a forma de sais insolúveis como o tartarato de potássio ou cálcio, de polifenóis, de compostos nitrogenados, como proteínas, peptídeos e aminoácidos, ou microbiológica. Tanto na produção de tintos como de brancos, o vinificador pode optar por

engarrafar os vinhos da forma como foram vinificados como foram produzidos (varietais) ou realizar mesclas (cortes) de vinhos de uvas diferentes (*assemblages*) para, em seguida, engarrafá-los. Mesmo os varietais podem receber pequenos cortes, mas devem manter 75% do volume da variedade que dá nome ao vinho (BRASIL, 1990; GUERRA, 2010).

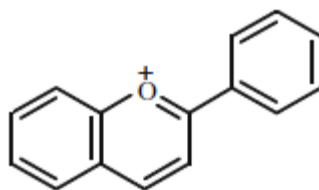
As características cromáticas de um vinho dependem tanto das uvas quanto da vinificação (MONAGAS; BARTOLOMÉ, 2009). A influência das uvas na coloração dos vinhos pode ser facilmente verificada no fato de ser impossível se obter um vinho tinto de uvas brancas. Enquanto a influência das técnicas empregadas na vinificação na coloração dos vinhos pode ser demonstrada, por exemplo, na obtenção de vinho branco a partir de uvas tintas. Assim, na produção de vinhos tintos, a etapa de maceração, que consiste no conjunto de operações visando a extração dos compostos fenólicos responsáveis pela cor, aroma e sabor do vinho, torna-se fundamental na produção desses vinhos, e torna-se responsável pelas características sensoriais do vinho (GUERRA, 2010).

3.1.4 Composição de vinhos

O vinho é uma mistura complexa, de compostos orgânicos, complementada por elementos inorgânicos, tendo água (entre 70 a 90% em volume), álcoois, ácidos orgânicos, polifenóis, polissacarídeos, açúcares, compostos aromáticos, minerais e vitaminas como principais componentes (GUERRA, 2010).

Segundo Ribéreau-Gayon et al. (2006), os compostos fenólicos estão em posição de destaque na composição de vinhos, respondem diretamente pela cor e o sabor dos vinhos tintos e têm propriedades saudáveis, que os credenciam como responsáveis pelo "paradoxo francês", com ação antibacteriana, antioxidante e vitamínica que aparentemente protegem os consumidores das doenças cardiovasculares.

Figura 3.2 – Estrutura do cátion flavilium.



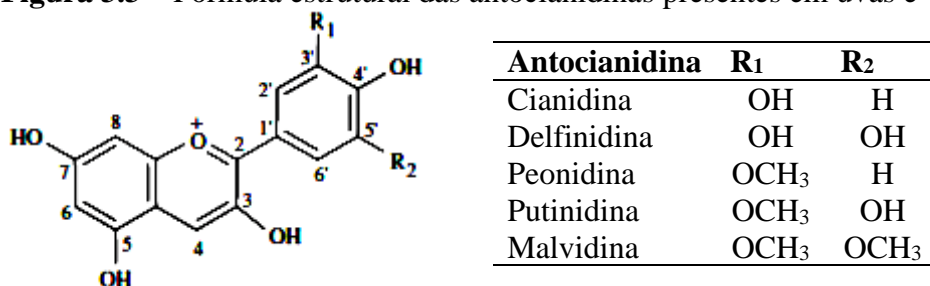
Fonte: Monagas; Bartolomé (2009).

As antocianinas e seus derivados, um grupo de compostos fenólicos, são responsáveis pela coloração dos vinhos tintos e rosados ou rosés e, por isso, têm papel especial na

composição dos vinhos (RIBÉREAU-GAYON et al., 2006). As antocianinas, cuja estrutura geral consiste do cátion flavilium, podem ser classificadas como flavonoides, por apresentarem uma estrutura comum de C6-C3-C6 consistindo em dois anéis fenólicos unidos entre si por um anel de pirano heterocíclico, mostrada na **Figura 3.2** (MONAGAS; BARTOLOMÉ, 2009).

As antocianinas dos vinhos são provenientes das cascas das uvas, embora possam ser advir da polpa de algumas variedades (*teinturier*), como é exemplo da variedade Alicante Bouschet (RIBÉREAU-GAYON et al., 2006). Nas uvas das espécies de *Vitis vinifera* e nos seus vinhos, são encontradas cinco antocianinas, que diferem uma da outra pelo número e pela posição dos grupos hidroxilo e metoxilo (OH e OCH₃) localizados no anel lateral do cátion flavilium e cujas fórmulas estruturais de suas antocianidinas (antocianina aglicona ou não-glicosada) são mostradas na **Figura 3.3** (MONAGAS; BARTOLOMÉ, 2009; RIBÉREAU-GAYON et al., 2006).

Figura 3.3 – Fórmula estrutural das antocianidinas presentes em uvas e vinhos.



Fonte: modificado de Monagas; Bartolomé (2009).

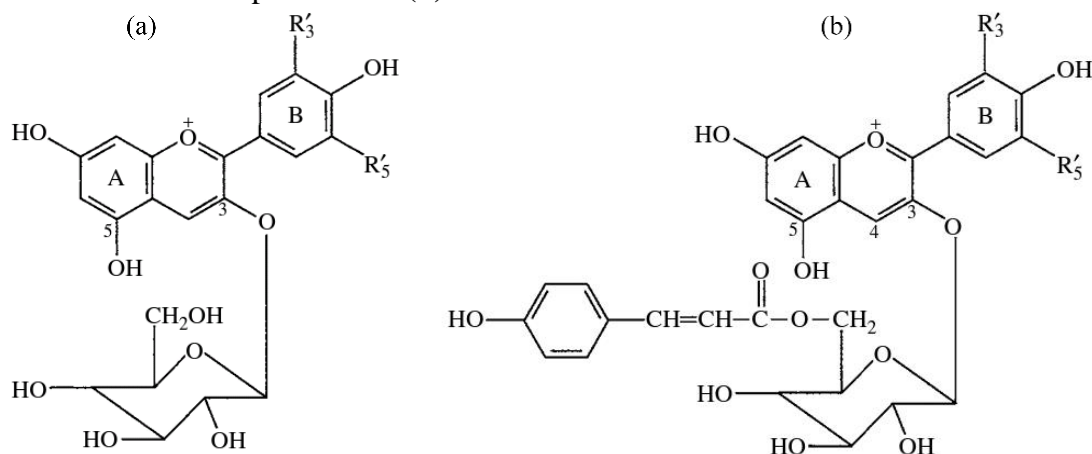
As antocianinas identificadas nas cascas de uvas *Vitis vinifera* e nos seus vinhos, consistem de moléculas de 3-O-monoglicosídeos e os monoglicosídeos 3-O-acilados destas cinco antocianidinas. A acilação ocorre na posição C-6 da molécula de glicose por esterificação com ácidos acético, p-cumárico e caféico (**Figura 3.4**) (MONAGAS; BARTOLOMÉ, 2009; RIBÉREAU-GAYON et al., 2006).

A estrutura molecular e o meio ambiente das antocianinas, além das mudanças do meio (pH, SO₂) provocam alterações na sua coloração: enquanto a substituição do ciclo lateral leva a uma mudança de batocrômico do comprimento de onda de absorção máximo (para violeta), a fixação da glicose e a acilação mudam a cor na direção oposta, ou seja, na direção da laranja (RIBÉREAU-GAYON et al., 2006).

O perfil de antocianinas e suas concentrações nas uvas dependem de vários fatores, incluindo a cultivar, maturidade das uvas, condições climáticas, área plantada e produtividade do vinhedo (MONAGAS; BARTOLOMÉ, 2009). No caso do vinho, o perfil e concentrações

dependerão fortemente das técnicas de vinificação adotadas, por exemplo, o tempo e a temperatura da etapa de maceração, durante a fermentação turbulenta, ou a presença ou ausência da etapa de filtração irão implicar numa maior ou menor extração dos pigmentos (antocianinas) das cascas das uvas. (MONAGAS; BARTOLOMÉ, 2009).

Figura 3.4 – Fórmula estrutural geral das antocianinas 3-glicosídeas (a) e 3-glicosídeas aciladas com ácido p-cumárico (b).



Fonte: modificado de Ribéreau-Gayon et al. (2006).

Segundo Ribéreau-Gayon et al. (2006), a malvidina é a antocianina dominante em todas as variedades de uva, variando de 90% a pouco menos de 50%, respectivamente nas uvas das variedades Grenache e Sangiovese. Ainda segundo este autor, dependendo das variedades de uvas, a faixa de variação vai de 100 mg L^{-1} a 1.500 mg L^{-1} nas uvas, mas diminuem severamente após a fermentação, durante os primeiros anos de envelhecimento (tanto em barris como em garrafa), até atingir um valor mínimo, entre $0\text{-}50 \text{ mg L}^{-1}$, combinando-se e se condensando-se com taninos para formar outras moléculas coloridas mais estáveis.

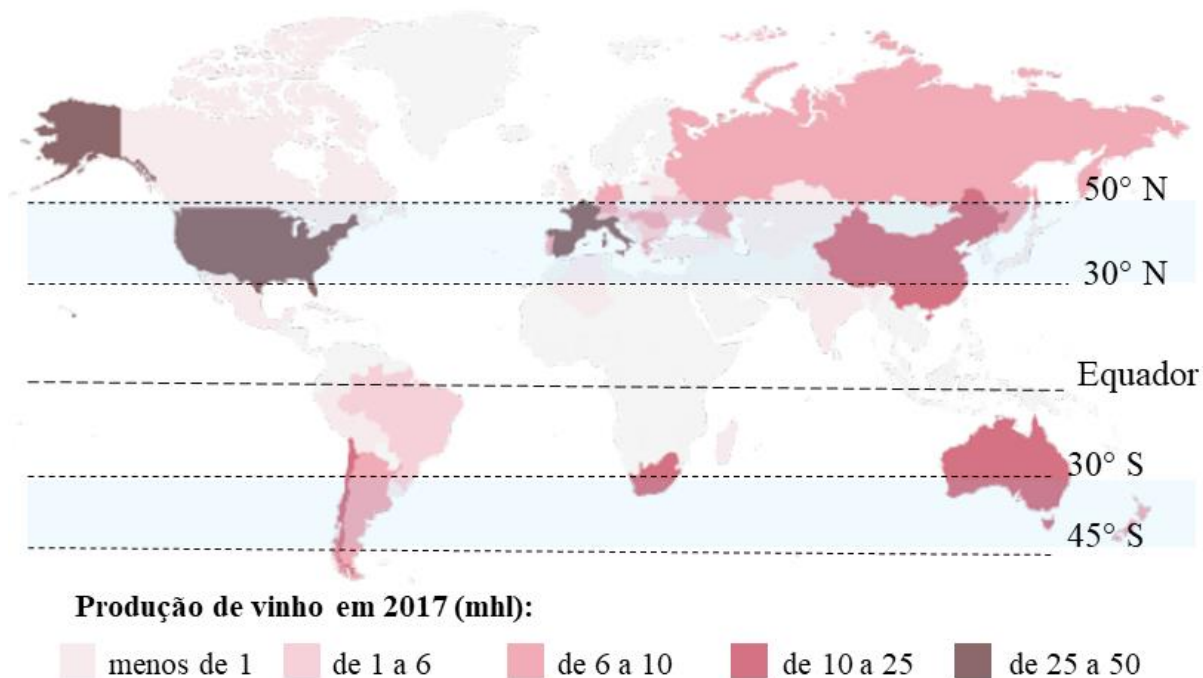
ANDRADE, et al. (2013) realizaram um estudo comparativo do perfil de antocianinas de vinhos produzidos em duas regiões brasileiras (Vale do São Francisco e Rio Grande do Sul) e no Chile, com uso de ferramentas quimiométricas e concluíram haver agrupamentos em função da origem geográfica. O mesmo estudo concluiu que o conteúdo das formas glicosadas (3-glicosídeos) de peonidina, petunidina e malvidina se mostraram mais fortemente responsáveis pela formação dos agrupamentos.

3.1.5 Produção de vinhos no Brasil e no mundo

Atualmente, há vinhas cultivadas nos cinco continentes do globo, de norte a sul, leste a oeste (GAUTIER, 2013; JOHNSON; ROBINSON, 2014). Tradicionalmente, o cultivo de uvas viníferas, especialmente para a produção de vinhos, ocorre em regiões de clima temperado, nas faixas entre as latitudes 30° e 50° no hemisfério Norte e 30° e 45° no hemisfério Sul, representados na **Figura 3.5** pelas faixas transversais coloridas e os países destacados com cores mais intensas são os que têm maior produção de vinhos (AMARANTE, 2015; JOHNSON; ROBINSON, 2014; OIV, 2017; PUCKETTE; HAMMACK, 2016).

Nos últimos anos, os avanços tecnológicos têm levado o cultivo de *V. vinifera* a regiões cada vez mais próximas da linha do Equador, em regiões áridas, dando origem aos vinhos tropicais. O Brasil figura como pioneiro nessa vitivinicultura e tem sua segunda maior região produtora situada na região do Vale do São Francisco, entre 8° e 10° de latitude sul, portanto, bem afastadas das faixas tradicionais de cultivo (AMARANTE, 2015; PEREIRA, 2013).

Figura 3.5 – Mapa da vitivinícola mundial. As faixas horizontais em coloração azul indicam as regiões de vitivinicultura tradicionais.



Fonte: modificado de OIV (2017).

Segundo dados da Organização Mundial da Vinha e Vinho (OIV), a área total de vinhas cultivadas em 2015 somou 7,5 milhões de hectares e rendeu uma produção de 75,8 milhões de toneladas de uvas, sendo que 35,8 milhões de toneladas foram destinadas à produção de vinhos

e o restante foi destinado ao consumo *in natura* ou para a produção de sucos ou de uvas passas (OIV, 2017).

No Brasil, há cultivo de uvas no Rio Grande do Sul, que responde por cerca de 62% da produção nacional, em São Paulo, Pernambuco, Paraná, Santa Catarina, Bahia e Minas Gerais. No entanto, a vinificação de uvas *Vitis vinifera* está restrita aos estados do Rio Grande do Sul, Pernambuco, Bahia e Santa Catarina (AMARANTE, 2015).

Conforme dados estatísticos apresentados na **Tabela 3.1**, no ano de 2016, o mundo produziu 267mhl (milhões de hectolitros) de vinhos. A Itália lidera o ranking mundial de produtores, com 50,9mhl, seguida por França (43,5mhl), Espanha (39,3mhl), Estados Unidos (23,9mhl), Austrália (13,0mhl) e China (11,4mhl). Juntos, estes seis países foram responsáveis por mais de 68% de todo o vinho produzido em 2016. Neste ano, o Brasil apresentou uma queda de 55% na sua produção e somente alcançou a 20ª posição no *ranking* mundial, com uma produção de 1,6mhl de vinhos (OIV, 2017).

Tabela 3.1 – Produção de vinhos, por países, em milhões de hectolitros e em porcentagens, no ano 2016.

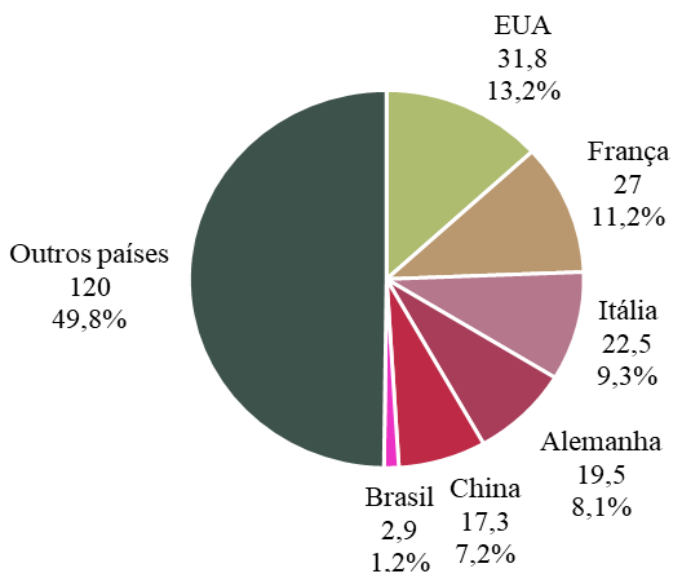
<i>Ranking</i>	<i>País/produtor</i>	<i>Volume (mhl)</i>	<i>Volume (em %)</i>
1º	Itália	50,9	19,1
2º	França	43,5	16,3
3º	Espanha	39,3	14,7
4º	EUA	23,9	9,0
5º	Austrália	13,0	4,9
6º	China	11,4	4,3
7º	África do Sul	10,5	3,9
8º	Chile	10,1	3,8
9º	Argentina	9,4	3,5
10º	Alemanha	9,0	3,4
11º	Portugal	6,0	2,2
12º	Rússia	5,6	2,1
13º	Romênia	3,3	1,2
14º	Nova Zelândia	3,1	1,2
15º	Grécia	2,6	1,0
16º	Sérvia	2,3	0,9
17º	Áustria	2,0	0,7
18º	Hungria	1,9	0,7
19º	Moldávia	1,7	0,6
20º	Brasil	1,6	0,6

Fonte: modificado de OIV (2017).

O consumo mundial registrado em 2016 chegou a 241 mhl (milhões de hectolitros) equivalentes ao total de 32,1 bilhões de garrafas. Conforme se vê na **Figura 3.6**, os Estados Unidos da América (EUA) lideram esse consumo, seguido por França, Itália, Alemanha e

China, e, juntos, estes cinco países são responsáveis por consumir quase metade de toda a produção mundial, somando 118,1 mhl ou 49% de vinho. No mesmo ano, o Brasil ocupou a 15ª colocação no ranking mundial de consumo, com uma fatia de 1,2% deste consumo ou 2,9 mhl consumidos (OIV, 2017).⁷

Figura 3.6 – Consumo de vinhos por países, em milhões de hectolitros e em percentagens, no ano de 2016.



Fonte: Adaptado de OIV (2017).

3.1.6 A viticultura no Vale do São Francisco/Brasil

A região banhada pelo rio São Francisco localizada entre os municípios de Remanso/BA e Paulo Afonso/BA, em ambas margens, é denominada Bacia do Submédio São Francisco, ocupa uma área de 109.827 Km² e engloba 91 municípios dos estados da Bahia e do Pernambuco (COMITÊ DA BACIA HIDROGRÁFICA DO RIO SÃO FRANCISCO, 2015). Na bacia do Submédio São Francisco, uma pequena área, compreendendo os municípios de Casa Nova, localizado no Estado da Bahia, e Lagoa Grande e Santa Maria da Boa Vista, localizados no Estado de Pernambuco, conforme mostrado na **Figura 3.7**, vem se destacando mundialmente na produção de uvas viníferas (*Vitis vinifera* L.) e de vinhos finos, denominada Zona de Produção Vitivinícola (ZPV) Vale do São Francisco (BRASIL, 2006).

Esta Zona de Produção Vitivinícola foi demarcada pelo Ministério da Agricultura, Pecuária e do Abastecimento, por meio da Instrução Normativa No. 01/2006, no ano de 2006 (BRASIL, 2006). Nela, os fatores naturais são muito diferenciados quando comparados com a grande maioria das tradicionais regiões produtoras de uvas (TONIETTO; TEIXEIRA, 2004) e

a ajudam a figurar como a principal região vitivinícola de clima tropical do mundo (GUERRA; ZANUS, 2004).

Figura 3.7 – Mapa vitivinícola do Vale do São Francisco.



Fonte: Modificado de Região Administrativa Integrada de Desenvolvimento do Polo Petrolina e Juazeiro (2014).

A região vitivinícola Vale do São Francisco está situada em zona de clima tropical semiárido, entre 8° e 10° de latitude Sul, em meio a paisagem de caatinga do sertão nordestino, a cerca 330 m de altitude, com temperatura média anual de 26,4 °C, e médias mensais de 28,3 °C e 24,3 °C, respectivamente, para os meses mais quente e mais frio e tem precipitação média

de 559 milímetros por ano, concentrados durante o período de dezembro a abril, tem topografia basicamente plana e solo argilo-silicoso, de baixa fertilidade, possibilitando que a produção de uvas seja escalonada ao longo do ano todo, permitindo, dependendo da cultivar, a colheita duas a três safras anualmente (ABS-PR, 2012; AMARANTE, 2015; PEREIRA, 2013; TONIETO; PEREIRA, 2011; TONIETTO; CARBONNEAU, 2004).

Os parreirais do Vale do São Francisco produzem diversas colheitas alternadamente, sendo que cada planta gera frutos pelo menos duas vezes ao ano, com a possibilidade de possibilidade de se colher uvas em todos os meses e de se encontrar uvas totalmente maduras a alguns metros de uvas totalmente verdes, como se ver na **Figura 3.8**, uma característica única da região, proporcionada pelo ambiente semiárido. No entanto, como em qualquer outra área vitivinícola, as variações do clima têm influência direta no desempenho das videiras (ABS-PR, 2012; IBRAVIN, 2014, PEREIRA, 2013; TONIETO; PEREIRA, 2011).

Segundo Camargo (2004), a possibilidade de programação da colheita e de vinificação ao longo do ano permite a elaboração de grandes volumes de vinho com pequena estrutura industrial. Isso estimulou, na década de 1980, os primeiros plantios comerciais de uvas para vinho no Vale do São Francisco, dando origem aos primeiros vinhos de regiões tropicais produzidos no Brasil.

Figura 3.8 – Fotografia de um parreiral no Vale do São Francisco com quatro fases da videira.



Fonte: Cortesia de Giuliano Elias Pereira.

Atualmente, as cultivares mais produzidas no VSF para a elaboração de vinhos são Syrah, Cabernet Sauvignon, Alicante Bouschet, Tannat, Ruby Cabernet, Touriga Nacional, Chenin Blanc, Moscato Canelli e Sauvignon Blanc (AMARANTE, 2015; PEREIRA, 2013; TONIETO; PEREIRA, 2011). No ano de 2015, a área plantada no Vale do São Francisco somou 400 hectares, distribuídos entre cinco produtores, o que levou a uma produção, naquele ano, de 4 milhões de litros, dos quais 2,8 milhões de litros de espumantes, 1,15 milhões de vinhos tintos e 40 mil litros de vinhos brancos (DARDEAU, 2015; ZANUS, 2016).

3.2 Indicações Geográficas

As indicações geográficas não são invenções modernas, provavelmente, evoluíram da percepção humana acerca de qualidades dos produtos com consequente classificação e, a partir daí, o estabelecimento de preferências por produtos e serviços de determinadas regiões. Segundo Dupim (2015), há referências a indicações geográficas em obras antigas tanto de Homero (do império grego) e Horácio (do império romano), quanto na bíblia antiga. Os egípcios, por volta do ano 2.250 a.C., inventaram algo que equivalia a um sistema de classificação dos seus vinhos, análogo aos *rankings* e denominações modernas, em que as ânforas que guardavam vinho eram rotuladas com o ano do reinado do faraó, o nome da vinha onde foi elaborado e sua localização e até o nome do enólogo e os vinhos podiam ser classificados como genuínos, bons ou muito bons ou nem obterem classificação (GAUTIER, 2013; TATTERSALL; DESALLE, 2015).

A União Europeia, ao regular o uso de indicações geográficas para produtos agrícolas, por meio da EC 510/2006, as reconheceu como uma tentativa de reduzir as adulterações e incrementar valor agregado aos produtos, oferecendo aos consumidores um elo entre produtos e sua região de origem que devem refletir padrões de qualidade ou produtos exclusivos.

No caso de vinhos, diversos países tradicionais na sua produção têm normas de controle de indicações geográficas há muito tempo. Como Portugal, por exemplo, que controla sua produção de Vinhos do Porto desde 1756, quando da criação, pelo Marquês de Pombal, da Companhia Geral da Agricultura das Vinhas do Alto Douro; Espanha e França que têm normas de controle de vinhos da década de 1930; Estados Unidos, que possuem normas de controle da qualidade de seus vinhos desde o ano de 1936 e a Comunidade Europeia, que exerce controle sobre a produção e comercialização de vinhos nos países membros, incluindo a normalização acerca de indicações geográficas, desde o ano de 1970 (REGULATION CE, 1970; ROSA, 2010).

O Brasil estabeleceu, no ano de 1996, por meio da Lei No. 9.279, de 14 de maio de 1999, o uso de indicações geográficas em produtos e serviços, sob duas formas: a Indicação de Procedência e a Denominação de Origem. Enquanto a Indicação de Procedência (IP) faz referência ao nome geográfico de uma região que se tenha tornado conhecida como centro de extração, produção ou fabricação de determinado produto ou de prestação de determinado serviço, a Denominação de Origem (DO) só pode ser concedida quando as qualidades ou características de tais produtos sejam exclusiva ou essencialmente atribuídas ao meio geográfico, combinando-se os fatores naturais e humanos (BRASIL, 1996). Há de se registrar, também, que a competência do Instituto Nacional da Propriedade Industrial (INPI) para conceder registros de Indicações Geográficas não é restrita a produtos e serviços nacionais, sendo também concedida a produtos estrangeiros que tenham circulação no Brasil.

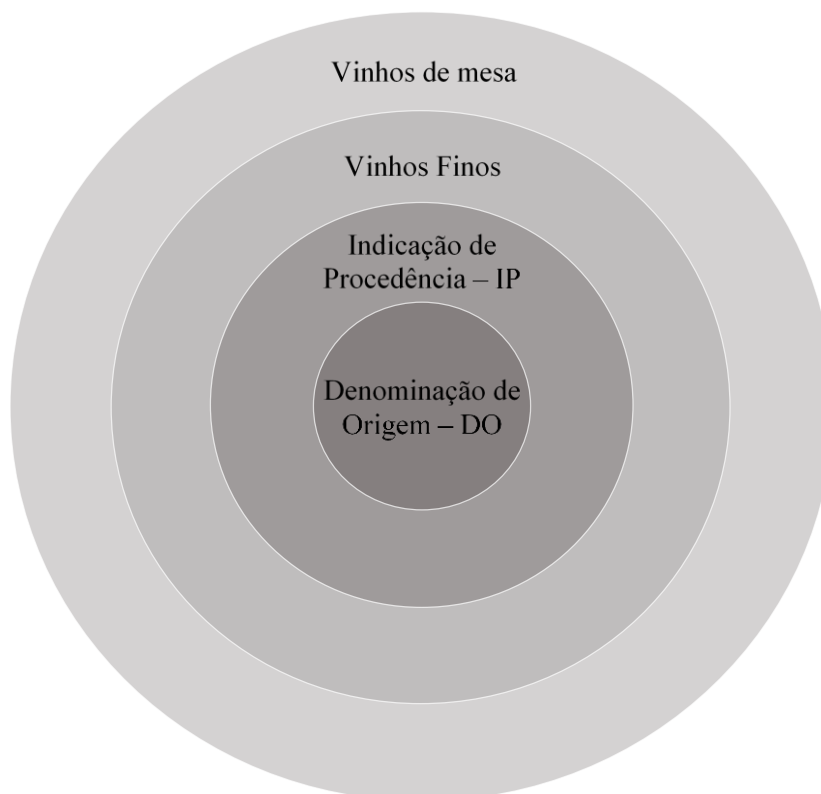
Embora a Denominação de Origem seja mais restritiva, exigindo mais qualificações dos produtos ou serviços, não há necessidade de prévia certificação com Indicação de Procedência, bastando se provar que determinada região se tornou conhecida na produção de produtos ou serviços que se diferenciam de outros equivalentes por ação direta dos fatores naturais e humanos daquela região. Tonietto (1993, p. 11) assevera que a Denominação de Origem “garante, além da origem geográfica precisa do produto, qualidades e características obtidas dos fatores naturais dessa origem e dos fatores humanos, que são regulamentados”.

Atualmente, constam nos registros públicos do Instituto Nacional de Propriedade Industrial, autarquia federal responsável por conceder registros de indicações geográficas, 41 Indicações de Procedência (todas nacionais) e 18 Denominações de Origem – 10 nacionais e 8 estrangeiras (Portugal, França, Itália e Estados Unidos) – concedidas. Seis entre as IP e seis entre as DO concedidas se referem a vinhos ou derivados, no entanto, apenas duas das DO são nacionais.

As Zonas de Produção Vitivinícola (ZPV), demarcadas pelo Ministério da Agricultura, Pecuária e Abastecimento, mesmo não possuindo caráter formal de indicação geográfica, nos termos da Lei de No. 9.279/1996, têm importante papel na identificação da origem dos vinhos brasileiros, uma vez que são, costumeiramente, usadas nos rótulos dos produtos que não são produzidos em regiões com indicações geográficas certificadas. Provavelmente, o uso do nome das ZPV nos rótulos de vinhos brasileiros decorre do grande lapso que o país levou para legislar sobre as indicações geográficas. O normativo mais recente na demarcação de Zonas de Produção Vitivinícolas, o Decreto No. 8.198, de 20 de fevereiro de 2014, dá conta da existência de 25 ZPV demarcadas no Brasil.

Na **Figura 3.9** é mostrada uma representação da estrutura qualitativa dos vinhos brasileiros, adequada à regulamentação atual. Do círculo mais externo para o mais interno, aumentam-se as exigências e, portanto, a qualidade esperada dos vinhos. O círculo mais externo representa os vinhos de mesa, que embora sejam produzidos exclusivamente pela fermentação de mosto de uvas sãs e frescas, admitem o uso de variedades de uvas americanas (*Vitis lambrusca*, *Vitis burquina*) e híbridas destas com variedades europeias (*Vitis vinifera*). O segundo círculo indica os vinhos finos, produzidos exclusivamente a partir de frutos da videira *Vitis vinifera*. O terceiro círculo representa as Indicações de Procedência – IP, que somente são certificadas pelo INPI após análise de processo que demonstre que aquela ZPV se tornou conhecida na produção vitivinícola. O quarto círculo, o mais interno, representa as Denominações de Origem – DO, mais elevada qualificação atribuída aos vinhos brasileiros e que somente é certificada, pelo INPI, após demonstração que os vinhos daquela ZPV apresentam características diferenciadas e que tais características são decorrentes das condições ambientais e dos processos de vinificação adotados na ZPV (BRASIL, 1990; BRASIL, 1996; BRASIL, 2014; TONIETTO, 2003).

Figura 3.9 – Esboço representativo da estrutura qualitativa dos vinhos produzidos no Brasil.



Fonte: modificado de Tonietto (2003).

Os produtores do Vale do São Francisco, representados pelo Instituto do Vinho do Vale do São Francisco (VINHOVASF) e assistidos pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), estão desenvolvendo esforços para a obtenção da certificação de indicação geográfica dos seus vinhos (PEREIRA, 2013).

3.3 Imagens Digitais

3.3.1 Conceitos e fundamentos

Segundo Solomon e Breckon (2013), uma imagem digital consiste numa representação discreta de dados que processam informações espaciais e de intensidade (cor). Para uma imagem bidimensional (2D), uma função $f(x, y)$, em que x e y são coordenadas espaciais e o valor de f , para cada par (x, y) , é chamada de intensidade de cor naquele par (GONZALES; WOODS, 2010). Uma imagem digital também pode ser descrita como uma matriz finita $A(m, n)$, em que m é seu número de linhas e n é o número de colunas. Nesta matriz, cada par (m, n) define uma célula chamada elemento de imagem ou *pixel* (termo derivado do inglês *picture element*) e cada *pixel* terá um valor de intensidade ou nível de cor, sendo a origem da imagem, $(0, 0)$, por definição, o seu extremo superior esquerdo (GONZALES; WOODS, 2010, SOLOMON; BRECKON, 2013).

O processo de construção de uma imagem digital, como na aquisição de uma fotografia ou no escaneamento de uma imagem impressa, implica em atribuir as coordenadas espaciais (m, n) para cada *pixel*, num processo chamado de amostragem, e suas respectivas intensidades (quantização), como ocorre, por exemplo, quando se digitaliza uma imagem impressa num papel fazendo uso de um *scanner* (GONZALES; WOODS, 2010 SOLOMON; BRECKON, 2013).

Segundo Solomon e Breckon (2013), a quantização é dependente dos sensores usados na detecção da cor, assim, a sensibilidade do sensor aos comprimentos de onda que incidem sobre ele é que define os vários níveis de intensidade possíveis. Por exemplo, um sensor monocromático (escala de cinza) com resolução de 8 bits para cada *pixel* irá quantizar 2^8 ou 256 níveis de intensidades, dos quais o primeiro e o último níveis, respectivamente, se referem ao preto e ao branco, enquanto os 254 níveis intermediários são níveis de cinza e, portanto, a escala de cinza desse sensor é quantizada em valores discretos que variam de 0 a 255 (SOLOMON; BRECKON, 2013).

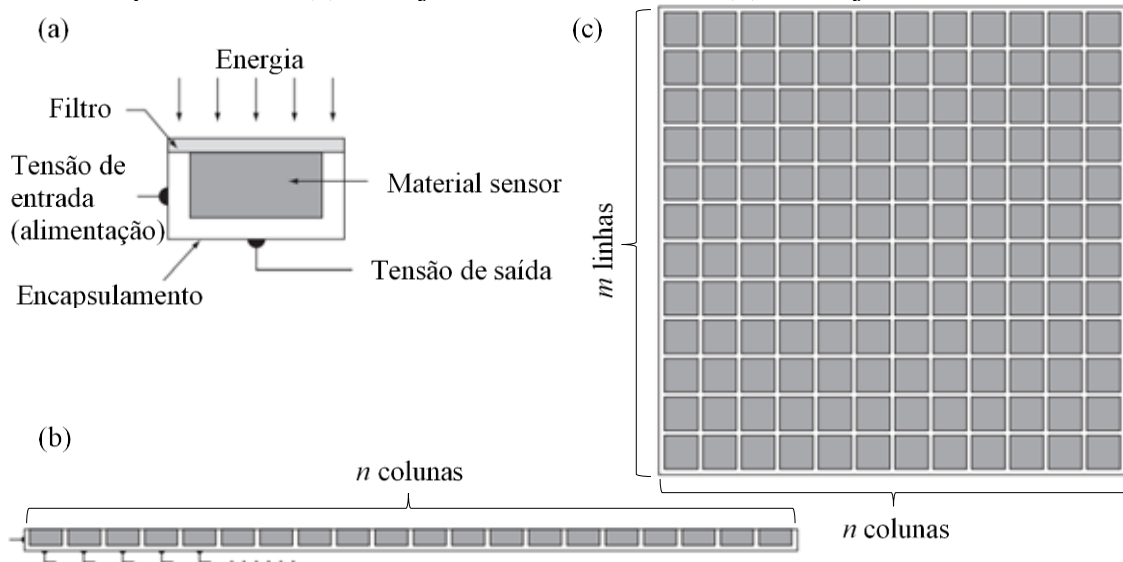
Apesar do uso do termo cor para se referir à amplitude da função $f(x, y)$, o uso de imagens digitais não é restrito às interações com a faixa visível do espectro eletromagnético, cobrindo praticamente todo o espectro, além de fontes acústicas, ultrassônicas e eletrônicas, embora seja comum converter sinais diversos para níveis de cinza a fim de expressá-las numa forma perceptível ao olho humano (GONZALES; WOODS, 2010).

Imagens digitais têm aplicações nas diversas áreas de atividade humana, sendo fortemente usada em medicina, onde se usam imagens de raio-X, tomografias de raios gama e de emissão de pósitrons, imagens de ultrassonografia e de ressonância magnética, mas também tem uso nas áreas da astronomia, de climatologia, aviação (com uso de radares, por exemplo), mais recentemente nas análises químicas, entre outras (GONZALES; WOODS, 2010).

3.3.2 Sistemas de aquisição de imagens digitais

Uma imagem geralmente é obtida pela interação entre a energia irradiada sobre um objeto ou cena, com absorção de parte desta energia e outra parte refletida ou transmitida chegando a um sensor ou conjunto deles (GONZALES; WOODS, 2013).

Figura 3.10 – Modalidades de sensores para aquisição de imagens bidimensionais: sensor único do tipo fotodiodo (a), arranjo de sensores em linha (b) e arranjo matricial de sensores (c).



Fonte: Adaptado de Gonzales e Woods (2013).

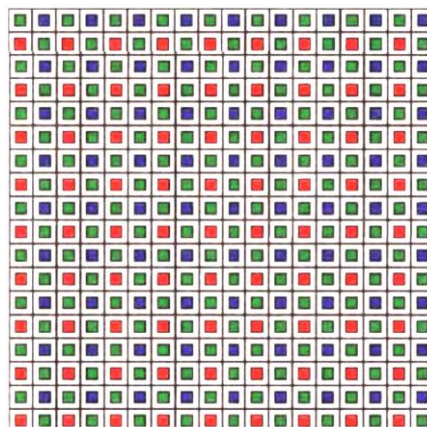
Na **Figura 3.10** são mostradas diferentes modalidades de sensores para se produzir uma imagem bidimensionais, em que se vê um sensor único ou de varredura bidimensional, que deve se movimentar ao longo dos eixos perpendiculares, x e y , para construir a imagem *pixel* por *pixel* (a), um conjunto de sensores em linha ou de varredura de

linha, usado na maioria dos escâneres, deve ter movimento perpendicular a esta linha, para construir imagens linha por linha **(b)** e um conjunto matricial de sensores, com m linhas e n colunas de sensores, muito comum em câmeras fotográficas, que não demandam movimento em relação à cena e constroem toda a imagem de uma só vez **(c)**.

Os sensores de imagens mais comuns são os fotodiodos, baseados no efeito fotoelétrico, construídos com materiais semicondutores e que produzem uma tensão elétrica proporcional à intensidade de luz recebida (GONZALES; WOODS, 2013). São destaques os dispositivos de carga acoplada (CCD, do inglês *charge-coupled device*) e o semicondutor de metal-óxido complementar (CMOS, do inglês *complementary metal-oxide semiconductor*), sensores matriciais que equipam praticamente todas as câmeras fotográficas atuais (FRASER; BANKS, 2013).

Os sensores CCD e CMOS medem a intensidade de energia que incide sobre eles, mas não são capazes de identificar os comprimentos de onda da radiação incidente e, por conseguinte, a cor (FRASER; BANKS, 2013; GONZALES; WOODS, 2013; SOLOMON; BRECKON, 2013). Então, as imagens coloridas são obtidas com auxílio de filtros de cores (CFA, do inglês *color filter array*), como os filtros de Bayer, que são películas que deixam passar determinados comprimentos de onda de certa cor primária e refletem os demais e cobrem as células fotossensíveis. Assim cada *pixel* recebe informação direta de uma única cor e as intensidades das demais cores são obtidas por interpolação (FRASER; BANKS, 2013). Na **Figura 3.11** é mostrado um esquema de arranjo de filtros de cores, em padrão Bayer (RGB), para detectores CCD e CMOS, em que se observa a quantidade dos filtros verdes em dobro em relação aos filtros vermelhos e aos azuis porque o sistema visual humano é mais sensível ao verde do que ao vermelho e ao azul (FRASER; MURPHY; BUNTING, 2005).

Figura 3.11 – Esquema, em padrão Bayer (RGB), de filtros de cores para detectores CCD e CMOS.



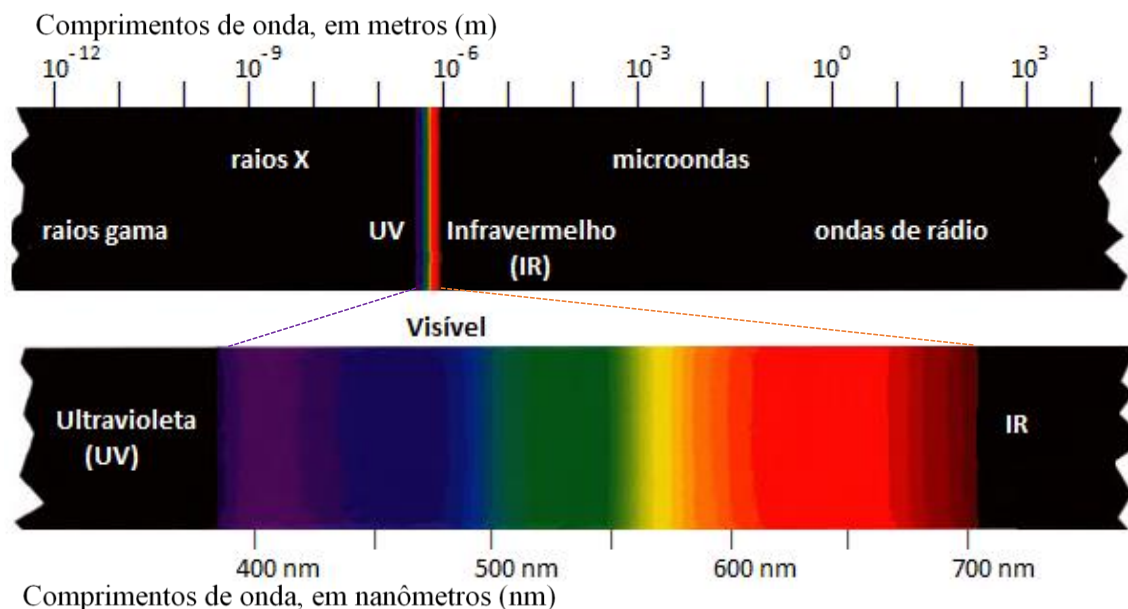
Fonte: Fraser; Murphy; Bunting (2005, p. 376).

3.3.3 O sistema visual humano

O olho humano, uma das mais belas estruturas na natureza, é um órgão praticamente esférico, com diâmetro aproximado de 2,0 cm, revestido por três membranas dentre elas, a mais interna, a *retina* que cobre toda a parte posterior do olho e onde se formam as imagens quando a luz emitida/refletida por um objeto é adequadamente focalizada, pela *córnea* – a lente curva que fica na parte anterior dos olhos – sobre ela (FRASER; MURPHY; BUNTING, 2005, GONZALES; WOODS, 2013).

Os olhos humanos respondem apenas a uma pequena porção do espectro eletromagnético e eles têm diferentes respostas para diferentes partes desta porção, ou seja, os diferentes comprimentos de onda provocam diferentes sensações de cor. A faixa de comprimentos de onda do espectro eletromagnético que está compreendida entre 700 nm e 380 nm, sensível ao olho humano, é chamada de visível e as cores percebidas nesta faixa variam desde os vermelhos, com comprimentos de onda mais próximos de 700 nm, passando pelos laranjas, amarelos, verdes e azuis, até os violetas, com comprimentos de onda próximos de 380 nm (FRASER; MURPHY; BUNTING, 2005). Na **Figura 3.12** é mostrado o espectro eletromagnético completo.

Figura 3.12 – Espectro eletromagnético.



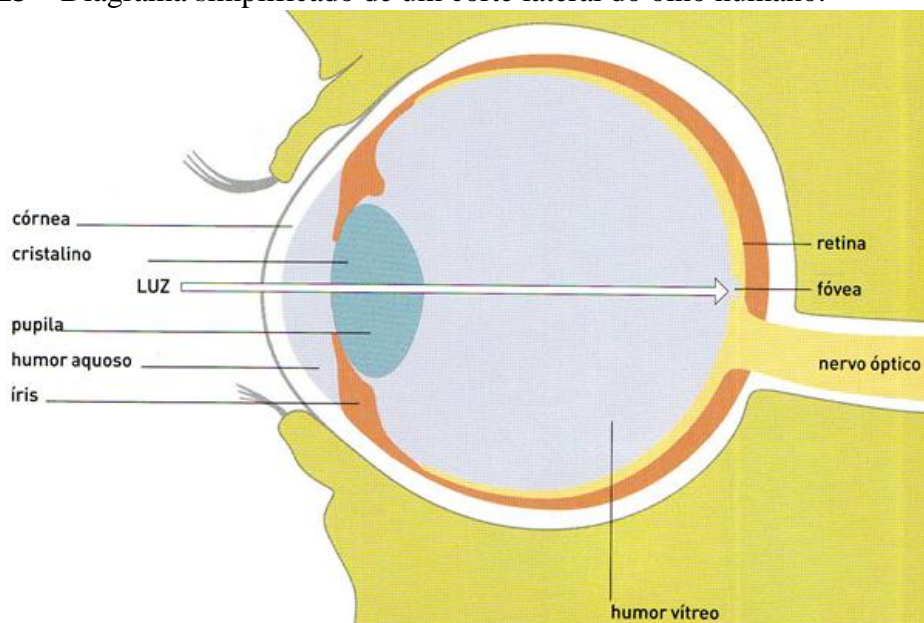
Fonte: adaptado de Fraser; Murphy; Bunting (2005).

Na retina, estão localizados milhões de receptores de luz, pertencentes a duas classes: os cones e os bastonetes (FRASER; BANKS, 2013; GONZALES; WOODS, 2013). Na **Figura 3.13** é mostrado um esquema simplificado de um corte lateral do olho humano.

Os cones são de três tipos, sensíveis a comprimentos de ondas específicos, correspondentes às cores vermelhas, verdes e azuis, sendo que praticamente todos – cerca de 6 a 7 milhões em cada olho – estão localizados numa pequena área, na parte central da retina, chamada de fóvea: uma depressão circular, com cerca de 1,5 mm de diâmetro, com uma densidade de cones de aproximadamente 150.000 elementos (cones) por milímetro quadrado, equivalente a uma matriz sensora quadrada medindo 1,5 mm x 1,5 mm, e são responsáveis pela visão clara ou visão fotópica (FRASER; MURPHY; BUNTING, 2005; GONZALES; WOODS, 2013).

Todos os bastonetes são essencialmente iguais entre si, estão presentes em quantidade muito maior que os cones, entre 75 a 150 milhões de unidades, distribuídos por toda a retina, e não são capazes de identificar as cores, mas possuem boa sensibilidade mesmo diante de baixas luminosidades e são responsáveis pela *visão escura* ou *visão escotópica* (FRASER; BANKS, 2013, GONZALES; WOODS, 2013).

Figura 3.13 – Diagrama simplificado de um corte lateral do olho humano.

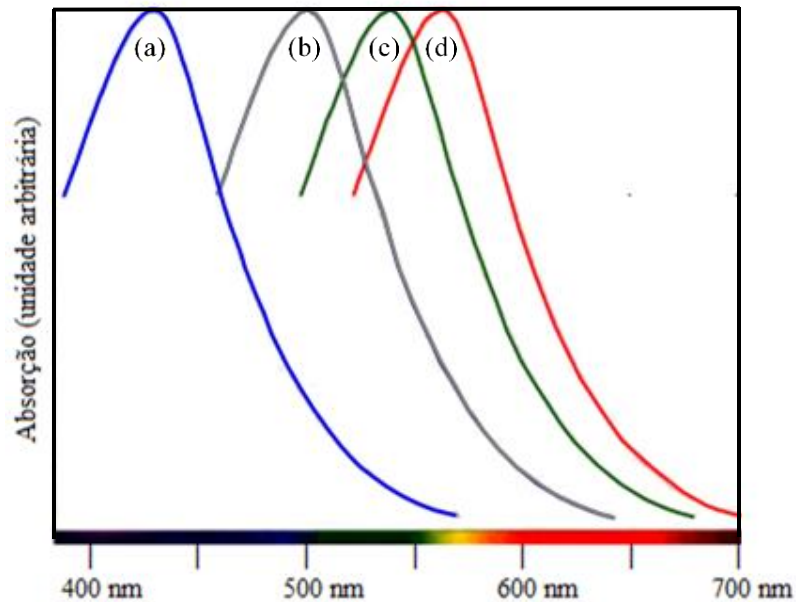


Fonte: adaptado de Fraser; Banks (2013, p. 24).

As faixas de absorção pelos fotorreceptores do olho humano (cones e bastonetes) são mostradas na **Figura 3.14**, na qual tem-se que as curvas de absorção dos cones azuis, com um pico de absorção a 420 nm **(a)**, dos bastonetes, com um pico de absorção a 499 nm **(b)**, dos

cones verdes, com um pico de absorção a 530 nm (c) e dos cones vermelhos, com um pico de absorção a 565 nm (d) (FRASER; MURPHY; BUNTING, 2005).

Figura 3.14 – Absorção da luz pelos cones azuis (a), bastonetes (b), cones verdes (c) e cones vermelhos (d).



Fonte: modificado de Fraser; Murphy; Bunting (2005, p. 18).

Embora os receptores do olho humano, especificamente os cones, só consigam detectar diretamente as cores vermelho, verde e azul, que, por isso, recebem a denominação de cores primárias, as demais cores são percebidas pela combinação das sensações das cores detectadas, conforme foi proposto pelo inglês Thomas Young (1773-1829) e aperfeiçoado pelo alemão Herman von Helmholtz (1821-1894), no século XIX, e que hoje é conhecida como mistura aditiva de cores e constitui o cerne do sistema de cores RGB que será descrito adiante (FRASER; BANKS, 2013, GONZALES; WOODS, 2013).

3.3.4 Sistemas de cores

Sistemas de cores são especificações de um sistema de coordenadas e um subespaço dentro desse sistema, no qual cor é representado por um único ponto e sua construção visa facilitar a especificação de cores de forma padronizada e permitir a comunicação entre usuários, na manipulação de cores e nos sistemas de processamento de imagens (GONZALES; WOODS, 2013). É frequente o uso das expressões modelos de cores e espaços de cores como sinônimas de sistemas de cores.

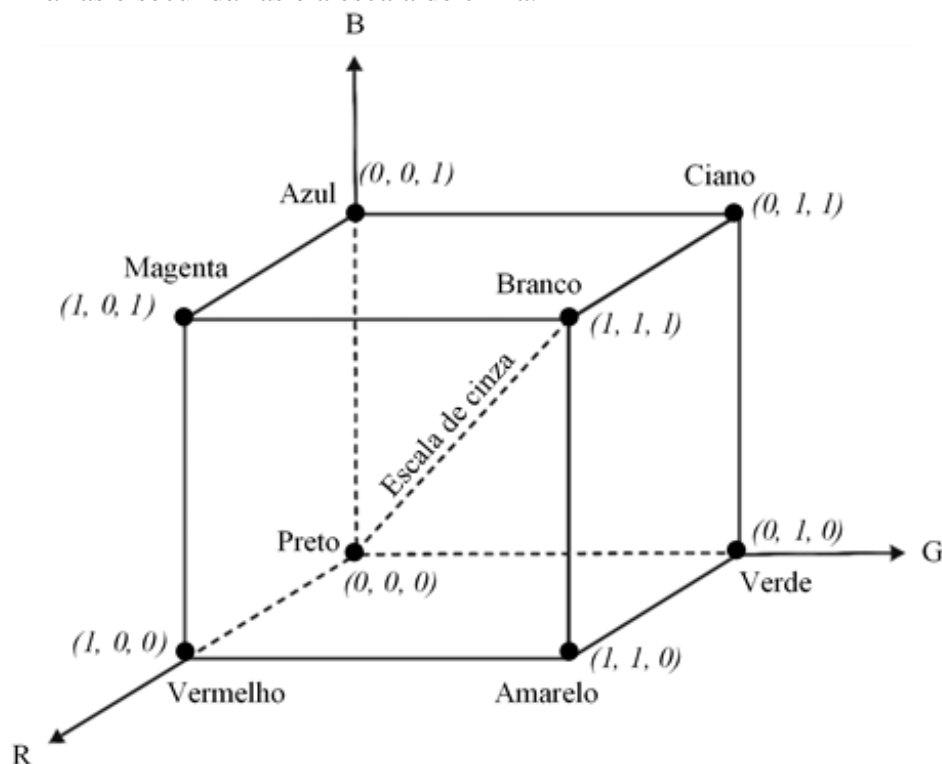
Segundo Gonzales e Woods (2013), é grande o número de sistemas de cores usados atualmente, sendo mais largamente usados os sistemas RGB (do inglês *red, green, blue*), CMY (do inglês *cyan, magenta, yellow*), CMYK (do inglês *cyan, magenta, yellow, black*) e HSI (do inglês *hue, saturation, intensity*).

3.3.4.1 Sistema de cores RGB

O sistema RGB tem seu uso mais voltado a *hardwares*, destacando-se especialmente em câmeras fotográficas e telas coloridas, sendo concebido a partir das três cores primárias: vermelho, verde e azul. Baseia-se no sistema aditivo de cores, organizadas sob um sistema de coordenadas cartesianas que tem como subespaço um cubo (GONZALES; WOODS, 2013).

Na **Figura 3.15** é mostrada uma representação do cubo do sistema RGB, no qual os oito vértices representam o preto $(0,0,0)$, branco $(1, 1, 1)$ e as máximas intensidades das cores primárias: vermelho $(1, 0, 0)$, verde $(0, 1, 0)$ e azul $(0, 0, 1)$, alternadas com as máximas intensidades das cores secundárias: amarelo $(1, 1, 0)$, magenta $(1, 0, 1)$ e ciano $(0, 1, 1)$. As três arestas que nascem na origem representam as intensidades das cores primárias puras e a diagonal principal, segmento de reta que vai do preto ao branco, é a escala de cinza.

Figura 3.15 – Representação gráfica do sistema RGB com indicações para o preto, o branco, as cores primárias e secundárias e a escala de cinza.



Fonte: modificado de Gonzales e Woods (2013).

Fica também evidente que as cores secundárias são obtidas pela mistura de duas cores primárias, o magenta (vermelho + azul), ciano (azul + verde) e amarelo (vermelho + verde). Os valores de cinza, então, podem ser calculados a partir das intensidades de vermelho, verde e azul.

Nessa representação, os valores de intensidade estão normalizados e variam de 0 a 1 em cada eixo, mas essa variação dependerá da quantização da imagem em questão e para os sistemas atuais é comum se ter imagens coloridas de 24 bits, sendo 8 bits para cada uma das cores primárias, que faz cada cor primária variar suas intensidades de 0 a 255.

3.3.4.2 Sistema de cores HSI

Diferente do sistema RGB, o sistema HSI não é voltado para uso em hardware, sendo mais apropriado a designação de características visuais humanas, por isso é conhecido como sistema de percepção humana. Não faz sentido, no cotidiano, um indivíduo tentar definir a cor de um objeto em suas componentes RGB, uma vez que as características facilmente percebidas para uma cor são matiz, saturação e brilho.

Segundo Gonzales e Woods (2013), o matiz está relacionado ao comprimento de onda dominante que é emitido por um objeto numa cena, enquanto a saturação está relacionada à diluição de uma cor pela cor branca, de modo que, na ausência da cor branca, uma cor é totalmente saturada e, por fim, o brilho, que é um descritor subjetivo praticamente impossível de ser mensurado está relacionado à percepção acromática de intensidade (nível de cinza) que é facilmente mensurável e se refere às imagens acromáticas.

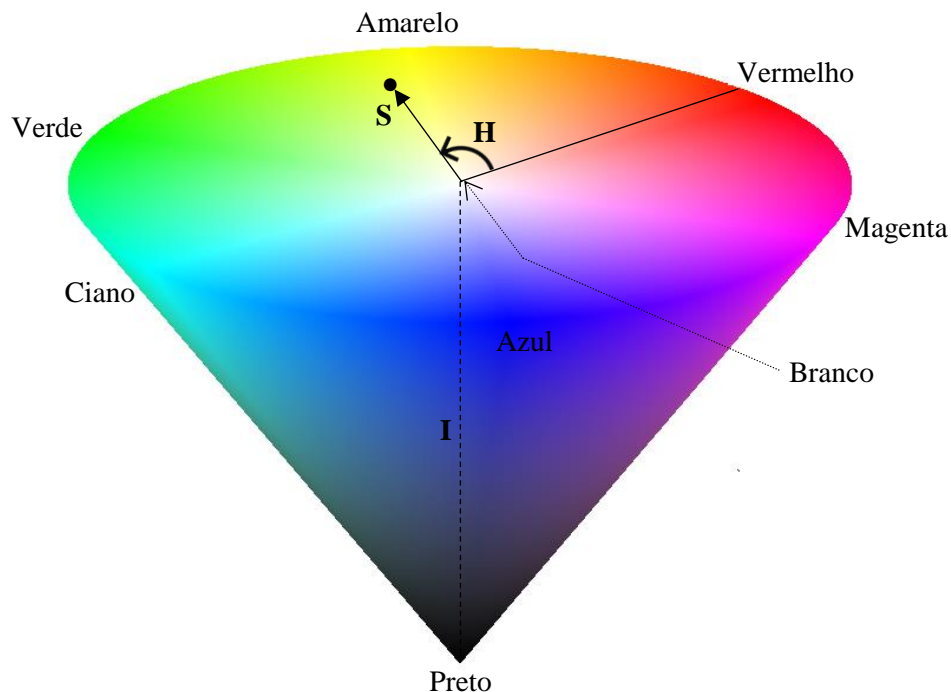
É possível representar o sistema HSI usando-se diversos subespaços, como o hexaedro, o octaedro, o cilindro, dentre outros. Usaremos o cone reto como subespaço, com o preto no vértice e o branco no centro da base circular, e, portanto, a *intensidade* (escala de cinza) variando ao longo da altura do cone, como mostrado na **Figura 3.16**. Por conveniência, o vértice será mantido apontado para baixo, e as cores primárias podem ser facilmente percebidas separadas por ângulos de 120° , alternadas pelas cores secundárias, isto é, o ângulo entre cada cor secundária será, também, de 120° e o ângulo entre uma cor primária e um cor secundária consecutiva será de 60° . Os atributos matiz, saturação e intensidade são atribuídos da seguinte maneira:

- O matiz (H) é dado pelo ângulo formado pelo segmento de reta que parte do eixo principal e que contém o ponto representativo da cor arbitrária e o segmento de reta que contém uma cor de referência, normalmente o vermelho, medido no sentido anti-

horário. Assim, o matiz de uma cor tem forte ligação com o comprimento de onda que define esta cor;

- A saturação (S) é dada pelo comprimento do vetor cuja origem é a intersecção do eixo principal com o plano perpendicular a este eixo e que contém o ponto de cor em questão e cuja extremidade é o ponto da cor em questão;
- A intensidade (I) é dada pela distância entre o vértice e o plano perpendicular ao eixo principal e que contém o ponto de cor em questão.

Figura 3.16 – Representação do espaço de cores do sistema de cores HSI baseado no cone reto. O ponto marcado em preto representa uma cor arbitrária.



Fonte: modificado de Hobbies, Hobbits and Hobos (2011).

3.3.4.3 Conversão entre os sistemas RGB e HSI

Como os sensores de captura de imagens comumente se baseiam em filtros de cores RGB, sendo este o sistema de cores mais comuns em máquinas fotográficas e escâneres, para se trabalhar com o sistema de cores HSI é necessário realizar operações de conversão entre os dois sistemas e tal conversão precisa ser realizada *pixel a pixel*.

De acordo com Gonzales e Woods (2013), para a conversão do sistema RGB para o sistema HSI, as Equações de 3 a 6 devem ser usadas.

$$H = \begin{cases} \theta & \text{se } B \leq G \\ 360 - \theta & \text{se } B > G \end{cases} \quad (3)$$

Em que:

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}(R-G) + (R-B)}{[(R-G)^2 + (R-B)(G-B)^{1/2}]} \right\} \quad (4)$$

$$S = 1 - \frac{3}{(R+G+B)} [\text{mín}(R, G, B)] \quad (5)$$

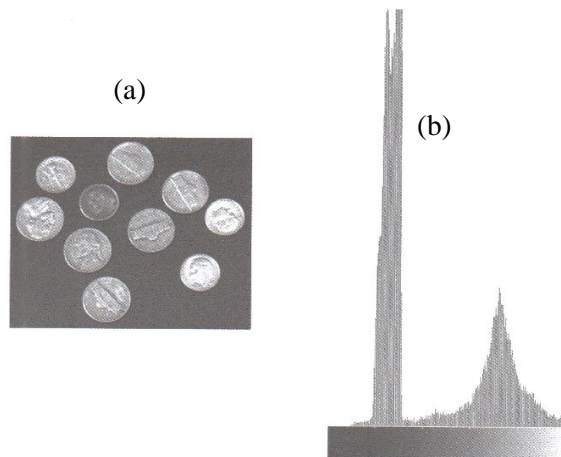
$$I = \frac{1}{3}(R + G + B) \quad (6)$$

De acordo com a Equação 3, os valores de H variam de 0 a 360. No entanto, é possível escalonar para valores entre 0 e 255. Outras equações devem ser utilizadas para a reconversão do sistema HSI para o sistema RGB. No entanto, não as abordaremos aqui, visto não serem objetos deste estudo.

3.3.5 Distribuição de pixels: histogramas

Histograma é um gráfico de barras retangulares que se usa para representar distribuições de frequência com dados agrupados (GIOVANNI; BONJORNIO; GIOVANNI JR, 2000). Segundo Solomon e Breckon (2013), o histograma de uma imagem é um gráfico de distribuição de frequências dos valores das intensidades de todos os *pixels* que a compõem.

Figura 3.17 – Imagem de dez moedas em fundo escuro (a) e seu histograma (b).



Fonte: SOLOMON; BRECKON (2013, p. 59).

Na **Figura 3.17** é mostrada uma imagem em escala de cinza (a) e seu respectivo histograma (b). Percebem-se dois picos no histograma, o primeiro e mais elevado fica numa região de baixa intensidade (mais próximo ao preto) e que se refere ao fundo escuro da imagem

e o segundo pico, de menor altura e na região de maior intensidade (mais próxima ao branco), referente às faces das moedas (SOLOMON; BRACKON, 2013).

3.4 Técnicas para o controle de autenticidade de origem geográfica de vinhos

As técnicas usadas no controle de autenticidade de vinhos, quanto a sua origem geográfica, há muito tempo têm grande relevância na comunidade acadêmica e as mais variadas técnicas analíticas, incluindo as espectrométricas e as cromatográficas, têm sido utilizadas (VERSARI et al., 2014). Espectrometria de absorção atômica (AAS, do inglês *Atomic Absorption Spectrometry*), espectrometria de emissão atômica com plasma indutivamente acoplado (ICP-AES, do inglês *Inductively Coupled Plasma-Atomic Emission Spectrometry*), espectrometria de massa com plasma indutivamente acoplado (ICP-MS *Inductively Coupled Plasma-Mass Spectrometry*) e cromatografia líquida de alta e ultra alta eficiência (HPLC e UPLC, das expressões em inglês *High Performance Liquid Chromatography* e *Ultra Performance Liquid Chromatography*), geralmente associadas a ferramentas quimiométricas, têm sido utilizadas na determinação de origem geográfica de vinhos, com obtenção de respostas satisfatórias (ANDRADE, R. H. S. et al. 2013; CARUSO et al., 2012; GALGANO et al., 2008; RAŽIĆ; ONJIA, 2010; FABANI al., 2010; BENTLIN et al., 2011; RODRIGUES et al., 2011; COETZEE et al., 2014; FRAIGE et al., 2014; MUCCILLO et al., 2014; CUADROS-INOSTROZA, 2010). Essas técnicas analíticas, no entanto, apresentam custo elevado, o que limita o seu uso aos laboratórios mais sofisticados e a grandes produtores, não sendo utilizadas para o controle de vinhos comerciais, de menor preço de mercado.

Por outro lado, nas últimas décadas, estudos baseados em imagens digitais (DIB, do inglês *Digital Image Based*) foram realizados com diversas matrizes e demonstraram capacidade de rastreamento de origem geográfica de sementes de quinoa (MEDINA; SKURTYS; AGUILERA, 2010), de chás (DINIZ et al., 2012) de mel (DOMINGUEZ et al.; 2014) e de própolis (PIERINI et. al., 2016), de controle de fermentação de chás (BORAH; BHUYAN, 2005) e de biomassa de leveduras em fermentações (ACEVEDO et al., 2009), de rastreamento de adulterações em leite de vaca (SANTOS; WENTZELL; PEREIRA-FILHO, 2012), café (SOUTO et al., 2014), azeite de oliva (MILANEZ; PONTES, 2015), classificação de amido em função dos seus grãos de origem (TORRENCE; WRIGHT; CONWAY, 2004), biodiesel (COSTA et al., 2015; COSTA et al., 2016), de taninos (GRASEL; MARCELO; FERRÃO, 2016), classificação de espécies de bactérias (ALMEIDA et al., 2014) e de validade e fontes de óleos vegetais comestíveis (MILANEZ; PONTES, 2014).

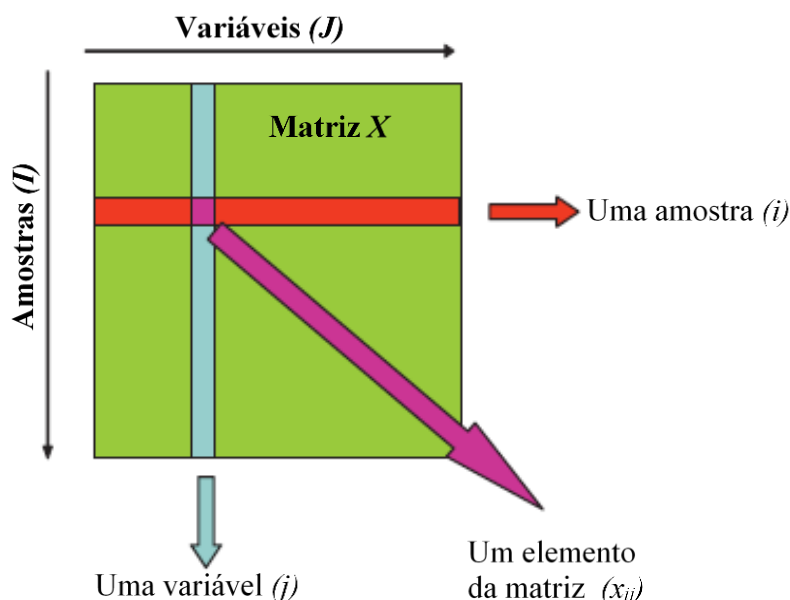
Não foi encontrado na literatura nenhum método analítico baseado em imagens digitais para classificação de vinhos em razão da sua origem geográfica.

3.5 Ferramentas quimiométricas

Quimiometria é a área química que usa métodos matemáticos, estatísticos e outros que empregam lógica formal para planejar ou selecionar procedimentos e experimentos de medição ótimos e fornecer informações químicas relevantes máximas através da análise de dados químicos (MASSART et al., 2003). A quimiometria encontrou aplicação generalizada na química analítica, mas não se restringe a ela, e vem sendo fortemente usada no controle da qualidade de produtos das indústrias químicas, farmacêuticas e de alimentos, na área forense e química medicinal (MASSART et al., 2003; FERREIRA, 2015).

Os dados experimentais, multivariados, normalmente adquiridos em instrumentos como espectrômetros, cromatógrafos, dentre outros, são dispostos em uma matriz, $\mathbf{X}(I \times J)$, onde as I linhas representam as amostras e as J colunas representam as variáveis (por exemplo, as intensidades de cores no histograma de uma imagem) e cujos elementos representam os valores das medições nessas amostras, conforme representado na **Figura 3.18** (BRERETON, 2009).

Figura 3.18 – Representação genérica de dados químicos analíticos em forma de matriz.



Fonte: Adaptado de Brereton (2009).

3.5.1 Técnicas de reconhecimento de padrões

Os seres humanos são muito bons em realizar classificação, tarefa que é repetida inúmeras vezes a cada dia, baseada na existência de similaridades entre os objetos para os colocar num mesmo grupo, e diferenças que os coloquem em grupos distintos (BEEBE, 1998; FERREIRA, 2015). Segundo Brereton (2003), a classificação ocupa lugar especial na Química, sendo usada para agrupar os elementos ou as substâncias em função das suas propriedades, como é exemplo da Classificação Periódica dos Elementos.

De acordo com Ferreira (2015), os métodos de reconhecimento de padrões podem ser supervisionados ou não supervisionados. No primeiro grupo, amostras (objetos) conhecidamente pertencentes a cada uma das classes são usadas para construção de modelos de classificação, enquanto no segundo grupo (métodos não supervisionados) não há conhecimento prévio a respeito da classificação das amostras, que são agrupadas de acordo com as informações contidas nos seus dados experimentais. Os métodos não supervisionados costumam ser chamados de métodos de análise exploratória de dados (FERREIRA, 2015).

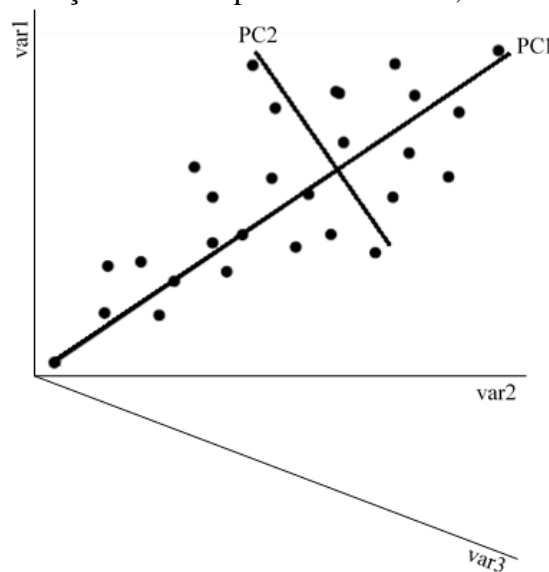
Dentre as diversas técnicas de reconhecimento de padrões descritas na literatura, algumas das mais populares e difundidas no tratamento de dados químicos são a Análise por Componentes Principais (PCA, do inglês *Principal Component Analysis*), a Análise Discriminante Linear (LDA, do inglês *Linear Discriminant Analysis*) com seleção de variáveis pelo Algoritmo das Projeções Sucessivas (SPA, do inglês *Successive Projections Algorithm*), SPA-LDA, e Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA, do inglês *Partial Least Squares Discriminant Analysis*).

A PCA é uma técnica não supervisionada de reconhecimento de padrões que visa reduzir a dimensionalidade de dados multivariados, transferindo as informações contidas em muitas variáveis para um número pequeno de novas variáveis, as Componentes Principais (PC, do inglês, *Principal Component*), facilitando a visualização de agrupamentos de amostras e de amostras anômalas (FERREIRA, 2015).

Cada PC consiste numa combinação linear de variáveis correlacionadas e, havendo num conjunto de dados mais de uma PC, as informações contidas numa delas não estarão presentes nas demais, isto é, as PCs de um mesmo conjunto de dados são ortogonais, a primeira componente principal, PC1, descreverá maior variância dos dados que a PC2, que descreverá maior variância que a PC3 e assim subsequentemente até que toda a variância dos dados tenha sido explicada (BEEBE, 1998; PONTES, 2009; FERREIRA, 2015; LAVINE; DAVIDSON, 2006).

Na **Figura 3.19** é mostrada uma representação gráfica da transformação no espaço de três variáveis (X, Y, Z) para o espaço de duas componentes principais (PC1 e PC2) de um conjunto de 28 amostras. Uma terceira componente principal, PC3, poderia ser calculada, mas conteria apenas ruído e pode ser considerada como um resíduo de modelagem (LAVINE; DAVIDSON, 2006).

Figura 3.19 – Eixos das componentes principais que definem um novo conjunto de vetores de base para o espaço de medição definido pelas variáveis X, Y e Z.



Fonte: Lavine e Davidson (2006).

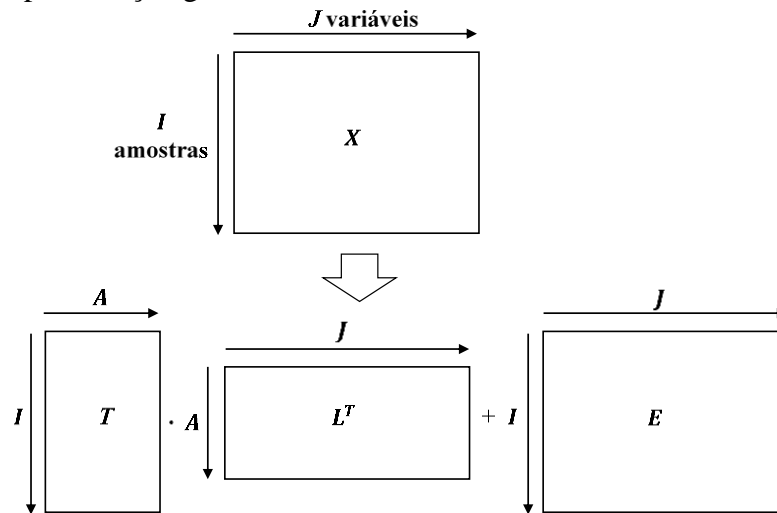
Matematicamente, a matriz de dados \mathbf{X} é decomposta em duas matrizes, uma matriz de escores \mathbf{T} , que exprime as relações entre as amostras, e uma matriz de *loadings* \mathbf{L} , que exprime as relações entre as variáveis, conforme representado pela Equação 7 (BRERETON, 2009).

$$\mathbf{X}_{(I \times J)} = \mathbf{T}_{(I \times A)} * \mathbf{L}_{(A \times J)}^T + \mathbf{E}_{(I \times J)} \quad (7)$$

Da **Equação 7**, temos que a matriz $\mathbf{L}_{(J \times A)}$ corresponde a matriz de projeções (subespaço ortogonal) também denominada de *loadings* e $\mathbf{T}_{(I \times A)}$ são os vetores projeções, ou seja, as coordenadas das amostras neste novo subespaço (escores). O termo A, corresponde a nova dimensionalidade dos dados, em outras palavras, A, é o número de PC usadas para explicar a informação contida em \mathbf{X} . A matriz \mathbf{E} , representa a informação não modelada pela PCA e tem as mesmas dimensões da matriz \mathbf{X} (BRERETON, 2009).

Na **Figura 3.20** é apresentada uma ilustração da decomposição por componentes principais, onde é possível observar que a decomposição em PCs corresponde a projetar as amostras em um subespaço ortogonalizado de baixa dimensionalidade na direção de máxima variância.

Figura 3.20 – Representação gráfica do método PCA.



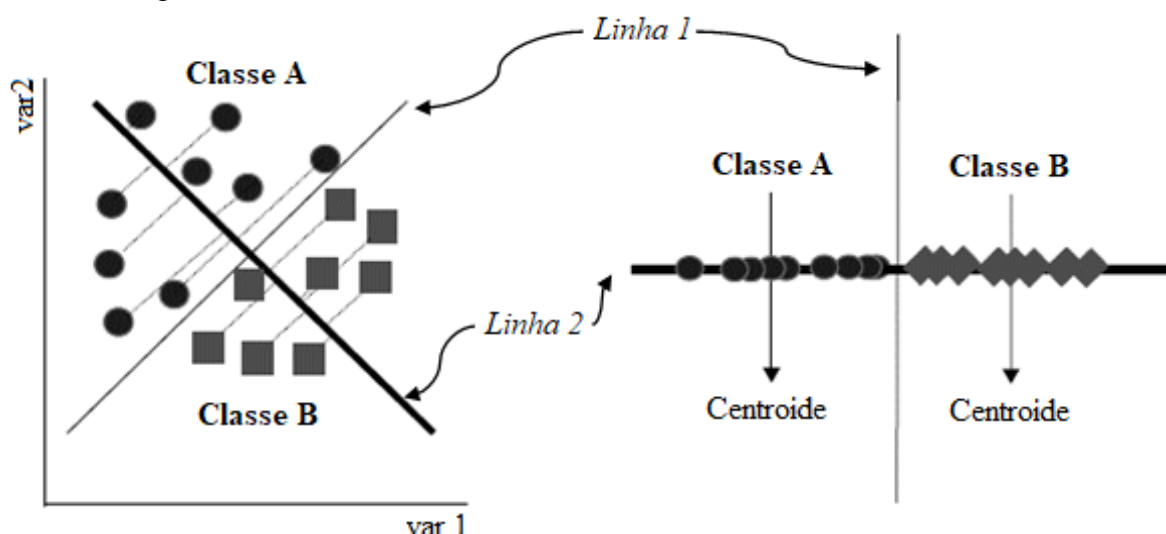
Fonte: adaptado de Brereton (2009).

Tão importante quanto usar ferramentas para detectar presença de argumentos entre as amostras, é usar uma estrutura de classes previamente conhecidas para fazer a predição de novas amostras. Neste contexto, a LDA é uma técnica de reconhecimento de padrões supervisionados, ou seja, que usa um conjunto de amostras com suas medidas e classes conhecidas para se construir ou treinar um modelo de decisão ou regra, enquanto outro conjunto de amostras, também com medidas e classes conhecidas, será usado para testar o desempenho do modelo construído em prever ou estimar a classificação de amostras desconhecidas (MASSART et al., 1988; FERREIRA, 2015). O primeiro conjunto é chamado conjunto de treinamento e o segundo recebe o nome de conjunto de teste. Semelhante a PCA, a técnica de LDA também usa combinações lineares para reduzir o número de variáveis do conjunto de dados. No entanto, as funções discriminantes visam o máximo de separação entre as classes (MASSART et al., 1988).

Na **Figura 3.21** é mostrado um esquema de discriminação entre duas classes A (círculos pretos) e B (quadrados cinza) em duas variáveis (var1 e var2). Percebe-se que os objetos representados por círculos são claramente distintos dos representados por quadrados, mas nenhuma das duas variáveis por si só pode discriminar completamente esses dois grupos. Entretanto, é possível separar os dois grupos ao traçar uma linha (*linha 1*) que seja uma combinação linear das duas variáveis, de modo que os objetos acima desta linha pertencem à classe A e os objetos que ficam abaixo pertencem à classe B. Projetando os objetos numa segunda linha (*linha 2*), perpendicular à linha discriminante (*linha 1*), temos que os objetos com valores de projeção mais baixos pertencem à classe A, enquanto aqueles com valores superiores pertencem à classe B. É possível determinar a associação de classe simplesmente de acordo

com o valor acima ou abaixo de um divisor. Alternativamente, é possível determinar o centroide (média das medidas) de cada classe ao longo da projeção e, se a distância ao centro da classe A for menor que a distância ao centro da classe B, o objeto é colocado na classe A, e vice-versa, mas isso depende de as classes serem quase igualmente difusas, ou seja, possuírem aproximadamente a mesma dispersão (BRERETON, 2003; SILVA, 2013).

Figura 3.21 – Esquema de discriminação entre duas classes (A e B), baseando-se nas variáveis originais (var1 e var2) e em linha discriminante.



Fonte: adaptado de Brereton (2003).

O poder de discriminação das variáveis será elevado quando os centróides das classes de amostras estiverem suficientemente distantes uns dos outros e quando as amostras das classes possuírem agrupamentos densos. Em termos matemáticos, isso significa que a variância entre as classes é grande em comparação com as variações dentro de cada classe. Deste modo, verifica-se que os três parâmetros matemáticos que determinam o efeito discriminatório de duas variáveis químicas são a variância entre classes, a variância dentro da classe e a correlação entre as variáveis químicas (MASSART et al., 1988).

No reconhecimento de padrões supervisionados, o objetivo é definir a distância de cada objeto ao centro de uma classe (as variâncias interclasses e intraclasses) e várias métricas de distância podem ser usadas para calcular os agrupamentos de objetos, ou *clusters*, e, por conseguinte, poderiam ser usadas para o cálculo de LDA. As mais comuns incluem as distâncias euclidiana, de Manhattan e de Mahalanobis. A distância de Mahalanobis é, destacadamente, a mais usada em dados químicos, porque considera as diferentes dispersões das classes e, também, pondera as influências das variáveis (BRERETON, 2003; MASSART et al., 1988).

A distância de Mahalanobis de uma amostra (i) ao centróide de uma classe é calculada de acordo com a Equação 8.

$$d_{iA} = \sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_A) \cdot \mathbf{S}_A^{-1} \cdot (\mathbf{x}_i - \bar{\mathbf{x}}_A)^T} \quad (8)$$

Onde \mathbf{x}_i é um vetor linha para a amostra i , $\bar{\mathbf{x}}_A$ é a medida média (ou centróide) para uma classe A e \mathbf{S}_A é a matriz de variância-covariância para esta classe (cujos elementos diagonais correspondem à variância de cada variável e os elementos fora da diagonal a covariância).

As limitações da LDA incluem as necessidades do número de amostras ser maior que o número de medidas, isto é, para J variáveis deve haver, ao menos, J + 2 amostras para que haja alguma discriminação, e que as variáveis sejam ao máximo independentes (não colineares) (BRERETON, 2003; FERREIRA, 2015; MASSART et al., 1988). Ocorre que, em análise multivariada, especialmente com uso de técnicas espectroanalíticas, o número de variáveis costuma ser muito grande e tal fato demandaria um número igualmente grande de amostras, aspecto que inviabiliza o estudo. Também, é muito frequente que existam num espectro um grande número de comprimentos de onda que tenham respostas instrumentais colineares. Quando a técnica analítica se baseia em imagens digitais, surge outra limitação da LDA, que é o grande número de variáveis com resposta instrumental nula.

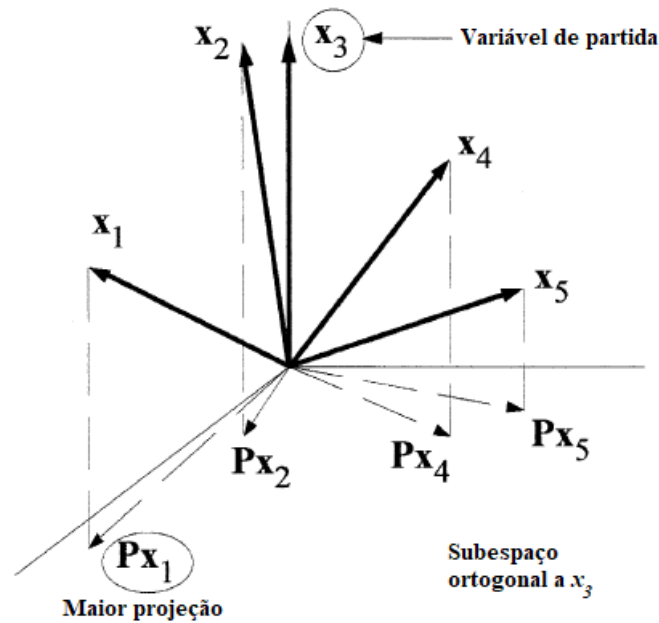
Para se contornar essas limitações, recorre-se a algoritmos adequados a realizar seleção de variáveis. Os métodos de seleção de variáveis devem selecionar um subconjunto de variáveis que retenham a informação discriminante das classes e exibam a menor colinearidade possível, favorecendo o alto poder preditivo por obtenção de uma matriz inversa da matriz de variância-covariância das variáveis (S) sem problemas de instabilidade numérica. Diferentes algoritmos para seleção de variáveis são conhecidos na literatura, como Algoritmo da Projeções Sucessivas (SPA), Algoritmo Genético (GA, do inglês *Genetic Algorithm*) e o *Stepwise* (SW) (ARAÚJO et al., 2001; FERNANDES, 2013, SILVA, 2013).

O SPA surgiu em 2001 como estratégia para selecionar variáveis para uso em calibração multivariada e é tido como um método de seleção que parte de uma variável arbitrária e, em seguida, realiza iterações até um número N_{max} previamente definido pelo usuário (ARAÚJO et al., 2001). O SPA na seleção de variáveis em LDA é composto de duas etapas: projeção (geração das cadeia de variáveis) e avaliação do poder preditivo. Na etapa de projeções, dada uma matrix de respostas instrumentais $\mathbf{X}(I \times J)$, o SPA partindo de uma variável \mathbf{x}_j , projeta todas as demais variáveis no complemento ortogonal (P^j de dimensões $I \times I$) do subespaço definido por \mathbf{x}_j . Baseado no comprimento dos vetores projeções $P^j \mathbf{x}$, são selecionadas variáveis que

apresentem a menor multicolinearidade mútua (ARAÚJO et al., 2001). A quantidade de variáveis que serão selecionadas será definida pelo melhor resultado em relação ao critério que avalia a habilidade de previsão do modelo de calibração multivariada, a Raiz Quadrada do Erro Médio Quadrático de Validação (RMSEV, do inglês Root Mean Square Error of Validation) (ARAÚJO et al., 2001; FERNANDES, 2013).

Na **Figura 3.22** é mostrado um exemplo de como seria a primeira iteração e mostra a projeção de quatro variáveis (\mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_4 e \mathbf{x}_5) no subespaço de \mathbf{x}_3 (variável de partida), onde a projeção da variável 1 (\mathbf{Px}_1) é a maior e será incorporada a \mathbf{x}_3 (ARAÚJO et al., 2001).

Figura 3.22 – Exemplo da seleção de variáveis com SPA.



Fonte: adaptado de Araújo et al. (2001).

Pontes et al. (2005) apresentaram uma versão do SPA com uma modificação que consistiu na incorporação de uma função de custo associada ao risco médio (G) de classificação incorreta por LDA, portanto para fins de classificação, o SPA-LDA. O custo médio (G) é calculado de acordo com a Equação 9.

$$G = \frac{1}{I_{\text{val}}} \sum_{i=1}^{I_{\text{val}}} g_i \quad (9)$$

em que g_i , o risco de classificação errada da i -ésima amostra de validação (\mathbf{x}_i), é calculado de acordo com a Equação 10.

$$g_i = \frac{d_{i\bar{A}}^2}{d_{iN}^2} \quad (10)$$

Na Equação 9, o termo I_{val} corresponde ao número de amostras do conjunto de validação do modelo. Na Equação 10, o termo do numerador, $d_{i\bar{A}}^2$, é o quadrado da distância de Mahalanobis da amostra \mathbf{x}_i ao centróide da sua classe correta e o termo do denominador, $d_{i\bar{N}}^2$, é o quadrado da distância de Mahalanobis da amostra \mathbf{x}_i ao centróide da classe errada mais próxima.

A Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA) é uma técnica de reconhecimento de padrão supervisionado baseado no algoritmo PLS, que leva à formação de um modelo de regressão que envolve duas matrizes: uma matriz independente $\mathbf{X}(I \times J)$, contendo respostas instrumentais de I amostras em J variáveis, e uma matriz simulada \mathbf{Y} , dependente de \mathbf{X} e contendo os índices de classes das amostras (BALLABIO; CONSONNI, 2013; BEVILACQUA et al., 2013, BRERETON; LLOYD, 2014, BRERETON, 2003) .

A matriz dependente \mathbf{Y} terá tantas colunas quanto o número de classes existentes e tantas linhas quanto o número de amostras presentes no conjunto de treinamento amostras (BEVILACQUA et al., 2013, BRERETON; LLOYD, 2014). Com isso, as informações sobre a classe a que pertença uma amostra do conjunto de treinamento, \mathbf{x}_i , são codificadas usando uma representação binária em que todas as entradas do vetor-linha desta amostra \mathbf{y}_i (na matriz dependente \mathbf{Y}) são iguais a zero, com exceção da coluna correspondente à categoria a qual a amostra pertence, cujo valor será igual a 1, conforme representado na **Figura 3.23** (BEVILACQUA et al., 2013; SILVA, 2013).

Figura 3.23 – Modelo de análise discriminante de mínimos quadrados parciais (PLS-DA) para mais de duas classes.

$\mathbf{X}(I \times J)$	$\mathbf{Y}(I \times 4)$
Classe 1	1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0
Classe 2	0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0
Classe 3	0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0
Classe 4	0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1

Fonte: adaptado de Silva (2013).

O PLS-DA busca encontrar onde uma função que relaciona a matriz \mathbf{X} , contendo as variáveis medidas (ou calculadas) nas amostras, para a matriz simbólica \mathbf{Y} ,

$$\mathbf{Y} = f(\mathbf{X}) \quad (11)$$

Assumindo-se que há uma relação funcional linear, pode-se reescrever a equação como:

$$\mathbf{Y} = \mathbf{X} * \mathbf{B} \quad (12)$$

Na Equação 12, \mathbf{B} é uma matriz de coeficientes de regressão, calculado pelo algoritmo PLS de regressão e será o modelo de classificação (BEVILACQUA et al., 2013).

A regressão PLS assume que tanto a matriz independente \mathbf{X} quanto a matriz dependente \mathbf{Y} podem ser projetadas em um espaço de fatores de baixa dimensão e que existe uma relação linear entre os escores das duas matrizes. Matematicamente, as matrizes dependentes e independentes são decompostas em escores e *loadings* de acordo com

$$\mathbf{X} = \mathbf{T} * \mathbf{P}^T + \mathbf{E}_X \quad (13)$$

onde \mathbf{T} , \mathbf{P} e \mathbf{E}_X são, respectivamente, as matrizes de escores, *loadings* e resíduos da matriz de dados originais \mathbf{X} ,

$$\mathbf{Y} = \mathbf{U} * \mathbf{Q}^T + \mathbf{E}_Y \quad (14)$$

onde \mathbf{U} , \mathbf{Q} e \mathbf{E}_Y são, respectivamente, as matrizes de escores, *loadings* e resíduos de \mathbf{Y} , e

$$\mathbf{U} = \mathbf{T} * \mathbf{C} \quad (15)$$

onde \mathbf{C} é uma matriz diagonal de coeficientes (BEVILACQUA et al., 2013, BRERETON; LLOYD, 2014, BRERETON, 2003).

Com base nas relações descritas nas Equações 14 e 15, é possível calcular a matriz dos coeficientes de regressão \mathbf{B} , que permite prever os valores da matriz dependente do conjunto de teste $\hat{\mathbf{Y}}_{test}$, nas quais as variáveis independentes \mathbf{X}_{test} foram medidas:

$$\hat{\mathbf{Y}}_{test} = \mathbf{X}_{test} * \mathbf{B} \quad (16)$$

Enquanto a matriz dependente \mathbf{Y} para as amostras de treinamento é codificada em binário, os vetores preditos para as amostras desconhecidas $\hat{\mathbf{y}}_{desc}$ conterão números reais e a regra de classificação para uma amostra x_m é realizada atribuindo-a à categoria correspondente a posição do componente do vetor de \mathbf{Y} previsto para esta amostra, $\hat{\mathbf{y}}_m$, cujo valor seja o mais alto. Por exemplo, se quatro classes estão envolvidas no problema de classificação, como no exemplo descrito na **Figura 3.23**, quando aplicado a uma amostra genérica desconhecida, o modelo na Equação 16 resultaria em um vetor-linha de quatro valores reais predispostos dimensionalmente que poderia levar valores de, por exemplo, [0,01 -0,04 **0,87** 0,06] e neste caso, a amostra x_m seria classificada como pertencente a classe 3 (BEVILACQUA et al., 2013).

4 EXPERIMENTAL

4.1 Amostras

Neste trabalho, foram usadas amostras de vinhos tintos, finos e tranquilos, comprados em supermercados de Natal/RN e de João Pessoa/PB ou doadas por vitivinicultores e instituições de pesquisa localizadas no Vale do São Francisco, Centro de Pesquisa Agropecuária do Trópico Semiárido, da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA Semiárido) e Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano (IF SERTÃO-PE). As amostras doadas pelas instituições de pesquisa não têm caráter comercial, sendo rotuladas como experimentais e sobre sua composição, podem ser tidos como monovarietais, isto é, não possuem qualquer corte. Contudo, não há informações acerca das técnicas enológicas empregadas na sua vinificação. As amostras foram adquiridas no período entre maio de 2015 e março de 2017, rotuladas e mantidas em temperatura ambiente, em armário fechado e na posição horizontal.

Três abordagens de classificação foram consideradas, em função da origem geográfica, dos vitivinicultores e das variedades de uvas.

Em relação à origem geográfica, foram utilizadas 100 amostras divididas, conforme informações contidas em seus rótulos, em duas classes: a classe Vale do São Francisco (**VSF**), e a classe de outras regiões, brasileiras e de outros países (**Mundo**). As informações contidas nos rótulos das amostras usadas nesta abordagem estão mostradas no ANEXO A, em que se percebe que as 64 amostras da classe VSF, embora sejam todas de uma mesma região de origem, apresentam uma grande variedade de informações, sendo produzidas por 8 diferentes viticultores, com 13 diferentes composições de uvas, 5 diferentes graduações alcoólicas e em 6 diferentes anos. Enquanto as amostras da classe Mundo apresentam 8 diferentes composições, 5 diferentes graduações alcoólicas e são produzidas por 21 diferentes viticultores, em 15 diferentes regiões e em 5 ou 6 diferentes anos.

A classificação relativa aos vitivinicultores incluiu apenas amostras produzidas no Vale do São Francisco e, devido à necessidade de número adequado de amostras, três classes foram formadas: **1**, formada pelos vinhos experimentais, **2**, formada por vinhos produzidos pelo vitivinicultor A, e **3**, formada por vinhos produzidos pelo vitivinicultor B. Embora tenha havido doação de seis vitivinicultores, as quantidades de amostras de quatro deles foram insuficientes para uma modelagem segura. No ANEXO B estão indicadas as informações contidas nos rótulos das amostras usadas na abordagem de classificação em função dos vitivinicultores

Para a classificação com foco nas variedades de uvas indicadas nos rótulos dos vinhos, três classes foram formadas: classe **CS**, formada por vinhos varietais de Cabernet Sauvignon, classe **Sy**, formada por vinhos varietais de Syrah ou Shiraz e classe **TN**, formada por vinhos varietais de Touriga Nacional (**TN**). Nesta abordagem, apenas amostras produzidas no Vale do São Francisco foram utilizadas. O detalhamento das amostras utilizadas nesta abordagem de classificação está disponível no ANEXO C.

As quantidades de amostras, em função das classes levantadas em cada uma das abordagens de classificação são apresentadas na **Tabela 4.1**.

Tabela 4.1 – Quantidades de amostras, em função das classes levantadas, em cada uma das abordagens de classificação utilizadas neste trabalho.

Abordagem de classificação	Classes	Quantidades
Origem geográfica	Vale do São Francisco (VSF)	64
	Outras regiões (Mundo)	36
Vitivinicultor	Experimentais (1)	20
	Vitivinicultor A (2)	24
	Vitivinicultor B (3)	26
Variedades de uvas	Cabernet Sauvignon (CS)	18
	Syrah/Shiraz (Sy)	14
	Touriga Nacional (TN)	16

Fonte: elaborada pelo autor, a partir das amostras utilizadas neste trabalho.

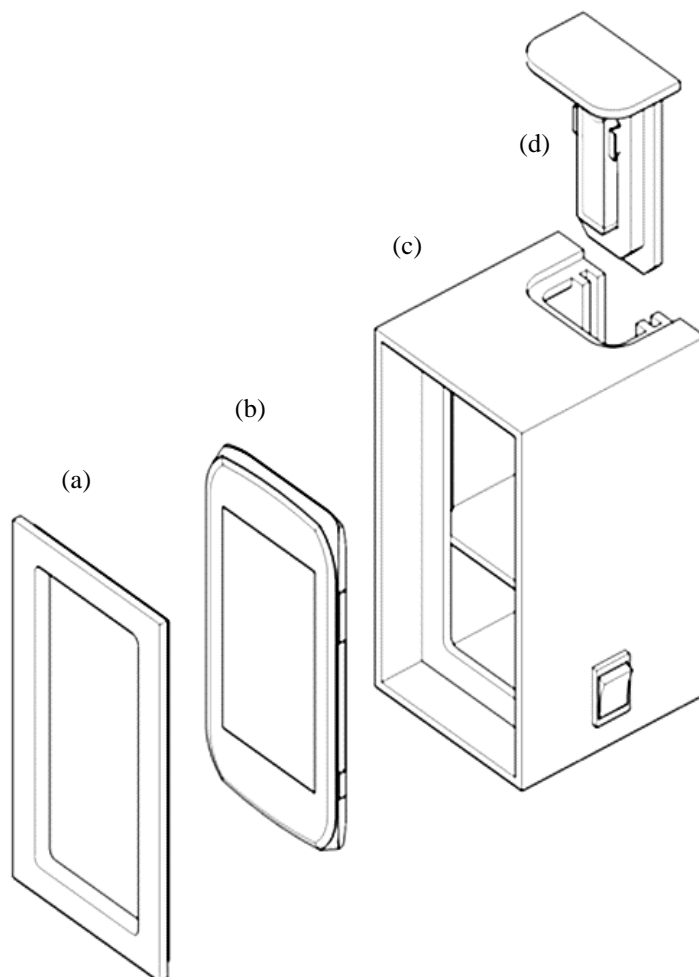
4.2 Aparelho para aquisição das imagens digitais

Para aquisição das imagens digitais, um aparelho simples e de baixo custo foi projetado e construído. O dispositivo para medições colorimétricas baseado em imagens digitais (DMCdib) está depositado no Instituto Nacional de Propriedade Industrial (INPI) com requerimento de Patente de Invenção sob o número BR1020170114490 (UNIVERSIDADE FEDERAL DA PARAÍBA, 2017).

O protótipo do DMCdib usado neste trabalho consistiu de uma caixa, com dimensões aproximadas de $16\text{ cm} \times 7\text{ cm} \times 7\text{ cm}$, construída em madeira cortada a *laser* e com as instalações adequadas (iluminação, acoplamento a um aparelho *smartphone*, apresentação das amostras,

etc.). Um esboço do DMCdib é apresentado na **Figura 4.1**, em que se observa suas partes principais: a tampa fixadora do *smartphone* (a), aparelho *smartphone* (b), caixa principal (c) e gaveta com suporte para cubeta (d).

Figura 4.1 – Esboço do DMCdib e suas partes principais: a tampa fixadora do *smartphone* (a), aparelho *smartphone* (b), caixa principal (c) e gaveta com suporte para cubeta (d).



Fonte: adaptado de UNIVERSIDADE FEDERAL DA PARAÍBA (2017).

O aparelho *smartphone* acoplado ao DMCdib se manteve fixo frontalmente às amostras analisadas e a uma distância constante de 3,5 cm destas. O DMCdib foi dotado de um sistema de iluminação com 6 lâmpadas LED brancas, posicionadas acima das amostras, nos extremos laterais da cavidade interna da caixa principal, logo abaixo do aparelho *smartphone*, alimentadas eletricamente por uma bateria de 9V, mantendo sua iluminação interna constante e proveniente, exclusivamente, desse sistema de iluminação. As paredes internas da caixa principal do DMCdib foram cobertas com tinta acrílica de coloração branca. A transferência das imagens do DMCdib para o computador foi realizada com uso de cabo de dados.

4.3 Aquisição dos dados baseados em imagens digitais

Um aparelho *smartphone* *Nokia Lumia 710*, com sistema operacional *Windows Phone* 7.5, foi acoplado ao DMCdib e usado para captura das imagens digitais das amostras de vinho. O *smartphone* teve o seu *flash* desligado e a seguinte configuração da câmera ajustada e usada em todas as leituras:

- a. Dimensões das imagens: 1.944 *pixels* x 2.592 *pixels* (largura x altura);
- b. Intensidade de bits: 24;
- c. Representação de cores: 16,7 milhões de cores;
- d. Formato da imagem: jpeg;
- e. Distância focal: 35 mm.

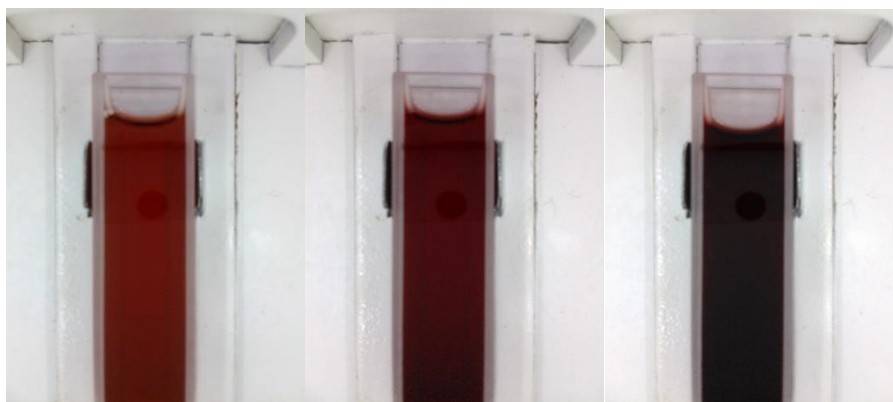
Uma cubeta de vidro óptico, com duas janelas polidas, medindo externamente 12,5 mm x 7,5 mm x 45 mm, com caminho óptico de 5 mm e com volume interno de 1,5 mL foi utilizada.

Alíquotas das amostras foram colhidas sem a retirada dos vedantes (rolhas) das garrafas, utilizando-se uma agulha de 50 mm de extensão, com seringa de 5 mL para perfurar as rolhas e colher as alíquotas. Nenhum processamento foi realizado nas amostras.

Imagens digitais coloridas, de 24 bits, foram registradas para cada amostra, medidas em três replicatas, usando volume fixo de 1,3 mL. As imagens digitais foram registradas entre os dias 1º de fevereiro e 03 de abril de 2017, totalizando trezentas imagens digitais.

Na **Figura 4.2** estão mostradas três imagens de diferentes amostras de vinho, todas do Vale do São Francisco, registradas pelo DMCdib. Nesta figura, observa-se o interior do DMCdib com suas paredes brancas e a cubeta contendo as amostras, fixada por pinças e inserida num rebaixo.

Figura 4.2 – Imagens de três amostras de vinho registradas pelo DMCdib.

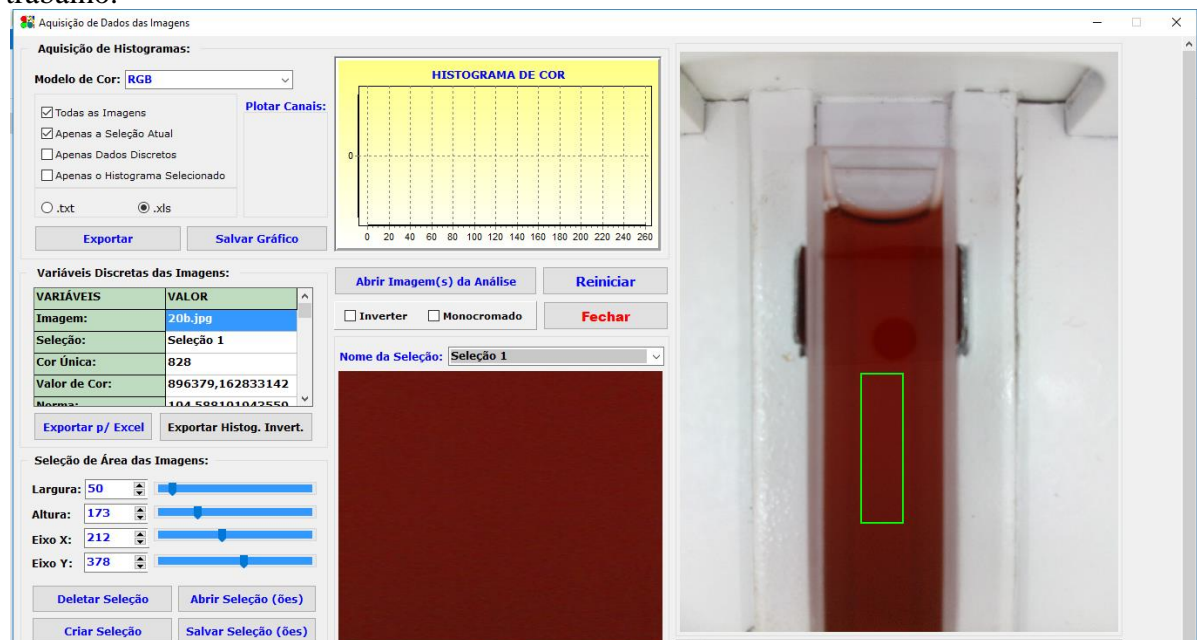


Fonte: produzida pelo autor.

Após o registro pela câmera do *smartphone* acoplado ao DMCdib, as imagens foram transferidas para um *Notebook Dell Inspiron 5557*, com sistema operacional *Windows 10 Home* de 64 *bits*, e dados de intensidades de cores, nos sistemas de cores RGB e HSI, foram obtidos com uso do *software* para aquisição de dados de imagens – QUIMAGENS (UNIVERSIDADE FEDERAL DA PARAÍBA, 2013). Uma região da imagem foi selecionada e as mesmas dimensões e localização espacial foram utilizadas para o tratamento de todas as imagens digitais.

Na **Figura 4.3** está representada uma imagem da janela de trabalho do QUIMAGENS, nas condições em que foi utilizado, em que se observa o retângulo verde, que indica a região da imagem onde os *pixels* foram utilizados para a obtenção dos dados de intensidade de cores da imagem digital. As informações das dimensões e da localização espacial da região selecionada estão informadas na parte inferior esquerda da imagem. A imagem que está sendo processada é da segunda replicata da amostra número 20.

Figura 4.3 – Imagem da janela de trabalho do *software* QUIMAGENS, como utilizado neste trabalho.



Fonte: produzida pelo autor.

Inicialmente, uma matriz de dados com 300 linhas e 1.536 colunas, $\mathbf{X}(300 \times 1536)$, foi obtida. As linhas referem-se às 100 amostras, cada uma delas com três replicatas e as colunas referem-se às variáveis (intensidades de cores RGB e HSI), sendo 256 variáveis para cada componente de cor.

4.4 Procedimentos quimiométricos e *softwares*

Uma redução no número de linhas da matriz de dados de intensidades de cores obtida inicialmente foi realizada com o cálculo das médias das replicatas de cada amostra e uma matriz de dados com 100 linhas x 1.536 colunas foi obtida. O *software* Unscrambler® X.1 (CAMO SA) foi utilizado no cálculo das médias das replicatas.

A partir da matriz com as médias das medidas das amostras, outras duas matrizes foram construídas contendo os dados das outras duas abordagens de classificação utilizadas neste trabalho (vitivinicultores e variedades de uvas). Em seguida, gráficos de distribuição de frequência *versus* intensidades de cores (histogramas) para cada uma delas foram plotados e, após inspeção visual dos histogramas, como estratégia de pré-processamento, seleções *a priori* de variáveis foram realizadas nos dados das três matrizes, sendo eliminadas as variáveis cujo sinal era nulo. Por fim, para cada estratégia de classificação adotada, as matrizes para cada sistema de cores, \mathbf{X}_{RGB} e \mathbf{X}_{HSI} , foram construídas.

Antecedendo a análise discriminante, uma análise exploratória com uso de Análise de Componentes Principais (PCA) foi realizada, separadamente, nas matrizes \mathbf{X}_{RGB} e \mathbf{X}_{HSI} de cada uma das abordagens de classificação.

Para os cálculos de análises discriminantes, com SPA-LDA e PLS-DA, os conjuntos de treinamento e de teste foram construídos por particionamento nas matrizes de dados \mathbf{X}_{RGB} e \mathbf{X}_{HSI} , separadamente, para cada uma das abordagens de classificação, utilizando-se o algoritmo Kennard-Stone (KS) (KENNARD; STONE, 1969) e não foram construídos conjuntos de validação, tendo se usado validação cruzada ou validação interna (SOARES et al., 2014). Na **Tabela 4.2** estão mostradas as quantidades de amostras em cada um dos conjuntos de treinamento e de teste para as abordagens de classificação origem geográfica, vitivinicultores e variedades de uvas. Todas as modelagens foram realizadas com os dados centrados na média.

Para os cálculos das médias das replicatas das amostras da PCA, Unscrambler® X.1 (CAMO SA) foi utilizado. Os algoritmos KS, SPA-LDA (PONTES et al., 2005; SOARES et al., 2014), PLS-DA (BALLABIO; CONSONNI, 2013) foram executados com o *software* Matlab® 9.3 (Mathworks, EUA).

Para avaliação final dos modelos, a taxa de classificação correta (TCC) das amostras e a sensibilidade e a seletividade de cada classe, dos conjuntos de treinamento e de teste, foram utilizadas, de acordo com as Equações 17, 18 e 19.

$$TCC (\%) = \left(\frac{N_{\text{acertos}}}{N_{\text{Total}}} \right) * 100\% \quad (17)$$

$$\text{Sensibilidade (\%)} = \frac{VP}{VP+FN} * 100\% \quad (18)$$

$$\text{Seletividade (\%)} = \frac{VN}{VN+FP} * 100\% \quad (19)$$

Em que VN, FP, VP e FN são os verdadeiros negativos, falsos positivos, verdadeiros positivos e falsos negativos, respectivamente.

Tabela 4.2 – Quantidades de amostras dos conjuntos de treinamento e de teste.

Origem geográfica		
Classes	Quantidades de amostras em cada conjunto	
	Treinamento	Teste
Vale do São Francisco (VSF)	44	20
Outras regiões (Mundo)	25	11
Total	69	31

Vitivinicultores		
Classes	Quantidades de amostras em cada conjunto	
	Treinamento	Teste
Experimentais (1)	14	6
Vitivinicultor A (2)	16	8
Vitivinicultor B (3)	18	8
Total	48	22

Variedades de uvas		
Classes	Quantidades de amostras em cada conjunto	
	Treinamento	Teste
Cabernet Sauvignon (CS)	12	6
Syrah/Shiraz (Sy)	9	5
Touriga Nacional (TN)	11	5
Total	32	16

Fonte: Elaborada pelo autor, a partir das amostras utilizadas neste trabalho.

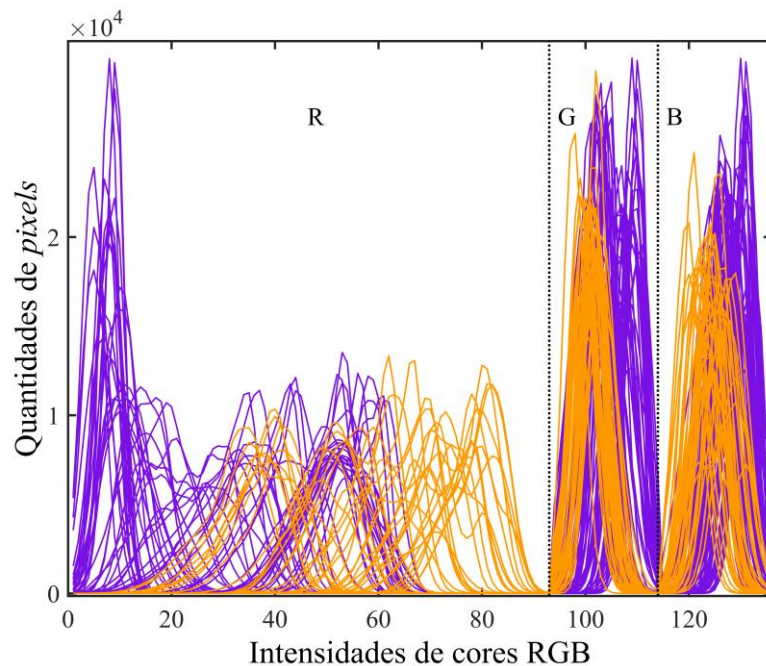
5 RESULTADOS E DISCUSSÃO

5.1 Classificação dos vinhos com base na origem geográfica

Na **Figura 5.1** são mostrados os histogramas das imagens no sistema RGB das 100 amostras usadas na modelagem de classificação de vinhos tintos em função da sua origem geográfica, após uma seleção *a priori* de variáveis em que se retirou as variáveis visivelmente nulas e 135 variáveis RGB foram mantidas. A retirada das variáveis nulas, além de permitir uma visível ampliação da figura, será útil nos cálculos de LDA, pois sua permanência não possibilitaria a transposição de vetores do cálculo das distâncias de Mahalanobis, como indicado na Equação 8.

Para permitir a sobreposição dos histogramas das 100 amostras no mesmo gráfico, com compreensão visual das amostras, os histogramas foram construídos com linhas contínuas,. As amostras das classes VSF e Mundo estão mostradas, respectivamente, em linhas roxas e laranjas.

Figura 5.1 – Histogramas das imagens no sistema RGB das 100 amostras usadas na modelagem de classificação de vinhos tintos em função da sua origem geográfica, após retirada das variáveis nulas. (—) VSF e (—) Mundo.



Fonte: elaborado pelo autor.

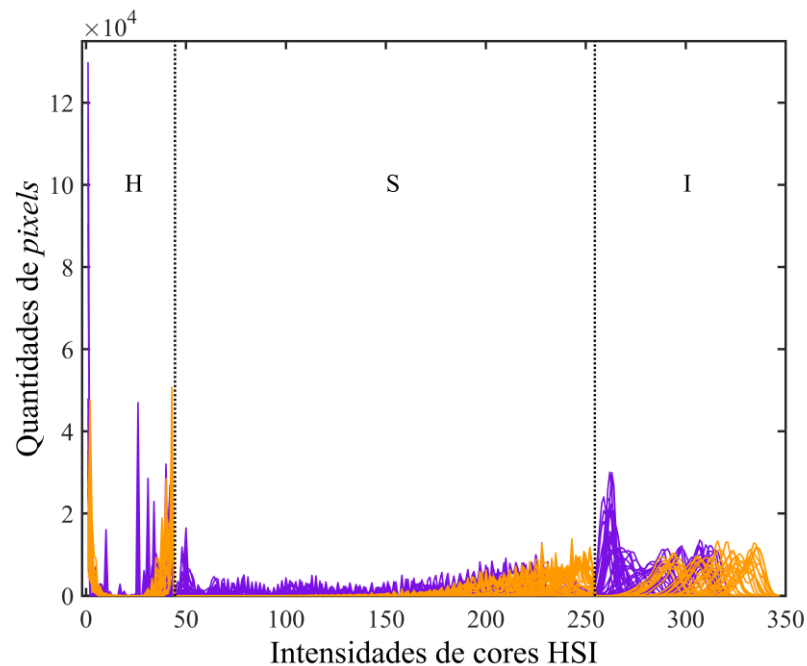
Notas: 1. Embora as intensidades de cores tenham valores discretos, para melhor visualização, aqui são utilizadas linhas contínuas.

A análise dos histogramas RGB indica uma grande dispersão dos dados no componente de cor R e picos mais estreitos nos componentes G (verde) e B (azul), indicando que as amostras em análise apresentam grande variação de tons de vermelho. A variação observada no componente R é uma característica da própria matriz, como função das variedades de uvas usadas na vinificação, bem como dos processos enológicos empregados pelos diferentes vitivinicultores. Conforme pode ser visto no ANEXO A, as amostras utilizadas nesta abordagem são produzidas com diferentes variedades de uvas, por diferentes vitivinicultores, o que sugere ter havido o emprego de diferentes técnicas enológicas. Os componentes G e B, apresentam-se como picos bem definidos, mostrando que não há grande dispersão dos dados nestas duas componentes de cores. Visualiza-se que há, nas três componentes (R, G e B), um ligeiro afastamento das amostras da classe VSF daquelas da classe Mundo. Estes afastamentos devem ser função de mudanças no perfil de antocianinas das amostras, causadas pelas diferentes características climáticas e de solo existentes entre as suas regiões de origem.

Na **Figura 5.2** são mostrados os histogramas das imagens no sistema HSI das 100 amostras usados na modelagem de classificação de vinhos tintos em função da sua origem geográfica, após seleção *a priori* de variáveis. Nas variáveis do H (matiz) é possível visualizar um afastamento entre as classes, de modo que as amostras da classe VSF apresentam respostas nas variáveis mais do centro, enquanto as da classe Mundo estão nas variáveis do extremo do matiz. A análise das intensidades de S (saturação) revela uma grande dispersão dos dados, com respostas num grande número de variáveis, sendo que as amostras da classe Mundo só têm informações entre as variáveis 150 e 250, enquanto as amostras da classe VSF se distribuem por todas as variáveis deste componente. Por fim, a análise da componente I (intensidade) mostra um grau médio de dispersão das amostras, dos valores mais baixos de intensidade de cinza, mais próximos do preto, até valores médios. Nas variáveis da componente I, as amostras da classe VSF apresentam resposta analítica nas variáveis com menores valores intensidades do que as amostras da classe Mundo, o que pode ser consequência de diferentes processos enológicos, como estabilizações fenólicas e filtragens (GUERRA, 2013) e cujas informações não estão disponíveis.

Em qualquer dos componentes HSI, as amostras de ambas as classes (VSF e Mundo) apresentam grande dispersão nos dados, o que pode ser considerado como natural devido à grande variedade da matriz e do conjunto de amostras utilizadas nesta pesquisa.

Figura 5.2 – Histogramas das imagens no sistema HSI das 100 amostras usadas na modelagem de classificação de vinhos tintos em função da sua origem geográfica, após retirada das variáveis nulas. (–) VSF e (–) Mundo.



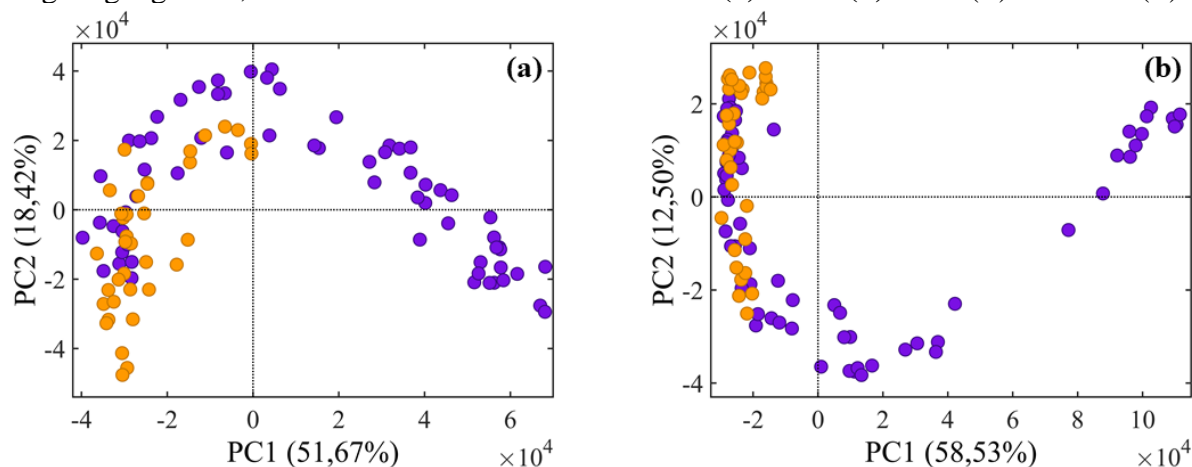
Fonte: elaborado pelo autor.

Nota: embora os dados de um histograma tenham valores discretos, para melhor visualização, aqui são utilizadas linhas contínuas.

5.1.1 Análise exploratória com PCA

Os escores de PCA revelam, tanto nos dados de intensidades de cores RGB quanto HSI, mostrados na **Figura 5.3**, uma grande dispersão das amostras, sendo que as amostras da classe VSF estão ainda mais dispersas que a classe formada pelas outras regiões do mundo, diferente do que se poderia inadvertidamente esperar. Essa grande dispersão nos escores de PCA para amostras da classe VSF pode ser atribuída a composição varietal dos vinhos, uma vez que as amostras desta classe são de 10 variedades e mais 3 cortes (*assemblages*) distintos, enquanto as amostras da classe Mundo são, em sua maioria, de uma varietal (Cabernet Sauvignon) ou corte em esta participe majoritariamente, como pode ser visto no ANEXO A.

Figura 5.3 – Gráfico de escores PC1 x PC2 de classificação de vinhos em função da sua origem geográfica, baseado nos sistemas de cores RGB (a) e HSI (b). VSF (●) e Mundo (●).



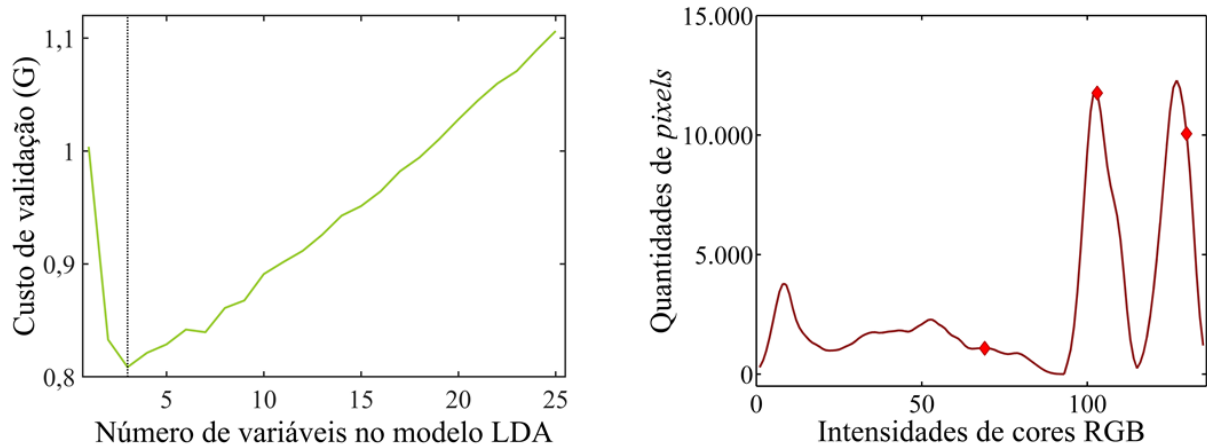
Fonte: elaborados pelo autor.

Na **Figura 5.3 (a)**, vê-se que amostras da classe Mundo estão no lado negativo da PC1, onde têm menor variância. Muitas amostras da classe VSF se separam completamente da classe Mundo, assumindo escores positivos na PC1, no entanto, parte das amostras da classe VSF se misturam às da classe Mundo. O mesmo comportamento é avistado na **Figura 5.3 (b)**, com várias amostras da classe VSF bem separadas das da classe Mundo e com escores positivos na PC1, enquanto outra parte das amostras da classe VSF aparecem com escores negativos e sobrepostas às da classe Mundo. Um pequeno grupo com 13 amostras da classe VSF está mais afastado das demais, com escores superiores a 60.000. Praticamente todas as amostras deste grupo são de vinhos experimentais elaborados com a variedade Touriga Nacional e sugerem uma boa capacidade da técnica analítica baseada em imagens digitais em agrupar, também, vinhos em função das suas composições varietais.

5.1.2 Análise discriminante com SPA-LDA e PLS-DA

Na **Figura 5.4** são mostrados os gráficos dos custos de validação (G) em função dos números de variáveis (a) e das variáveis selecionadas no modelo SPA-LDA (b), com os dados de cores RGB. Esses gráficos indicam a necessidade de um baixo número de variáveis para as modelagens. Para o modelo RGB, o menor custo de validação é obtido com 3 variáveis, que estão mostradas no gráfico da **Figura 5.4(b)**, destacadas sobre a média dos valores de intensidades de cores RGB das amostras de treinamento, sendo que foi selecionada uma variável em cada componente de cor (R, G e B), confirmando o que foi observado na análise dos histogramas acerca do pequeno afastamento existente entre os picos das duas classes.

Figura 5.4 – Gráficos dos custos de validação (G) em função do número de variáveis incluídas no modelo (a) e das variáveis selecionadas (b) pelo algoritmo SPA-LDA, com dados RGB.

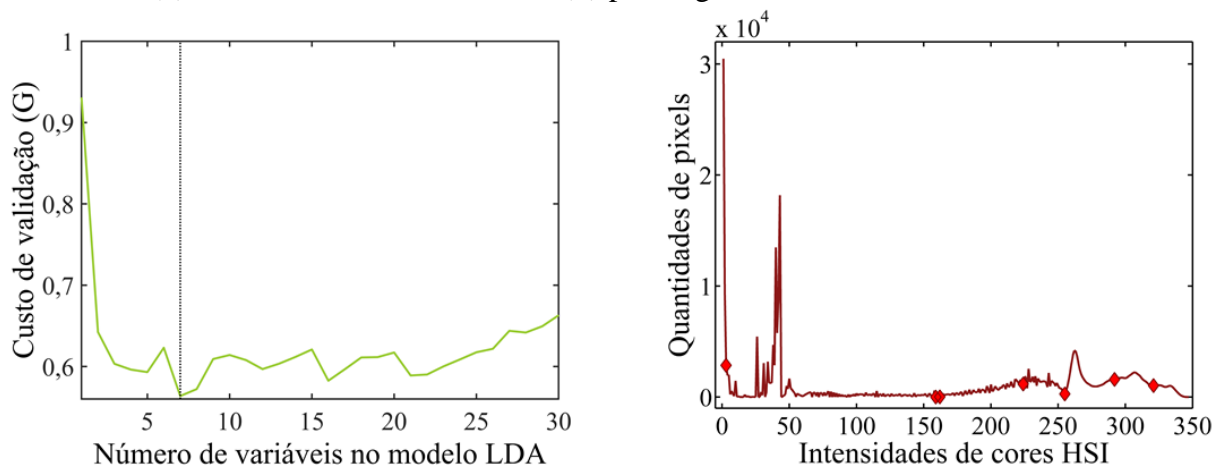


Fonte: elaborados pelo autor.

Nota: embora as intensidades de cores tenham valores discretos, para melhor visualização, aqui são utilizadas linhas contínuas.

Na **Figura 5.5** são mostrados os gráficos dos custos de validação (G) em função dos números de variáveis (a) e das variáveis selecionadas no modelo SPA-LDA (b), com os dados de cores HSI. O menor custo de validação é obtido quando 7 variáveis são incluídas no modelo. As variáveis selecionadas são mostradas sobre o histograma médio das amostras de classificação, em que se verifica que uma variável pertence à componente H, três pertencem à componente S e as outras três pertencem à componente I. Essa distribuição das variáveis sugere que a saturação (S) e a intensidade (I) têm maior poder de discriminação que o matiz de cor (H).

Figura 5.5 – Gráficos dos custos de validação (G) em função do número de variáveis incluídas no modelo (a) e das variáveis selecionadas (b) pelo algoritmo SPA-LDA com dados HSI.



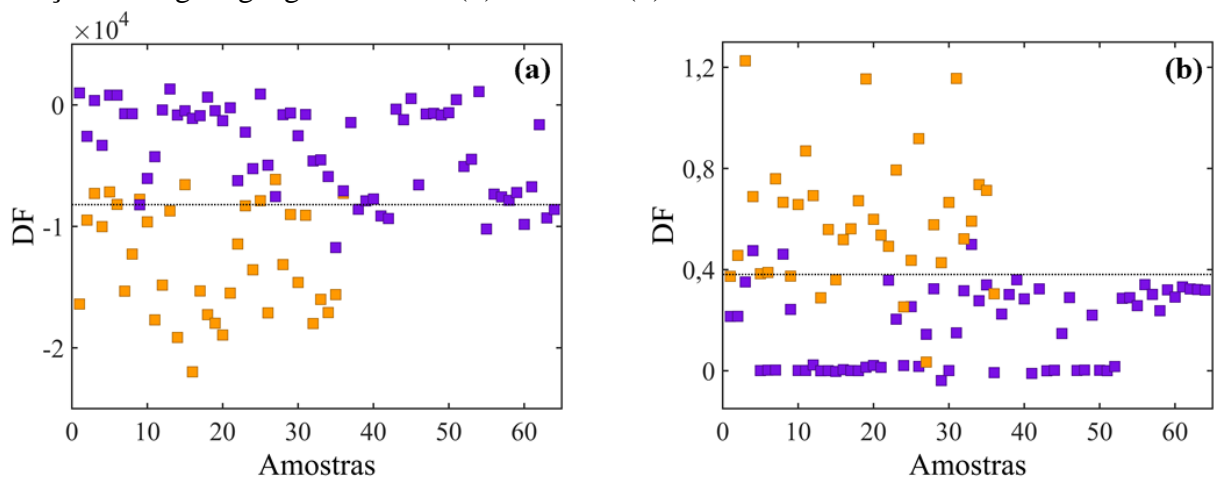
Fonte: elaborados pelo autor.

Nota: embora as intensidades de cores tenham valores discretos, para melhor visualização, aqui são utilizadas linhas contínuas.

Na **Figura 5.6** são mostrados os gráficos dos escores das Funções Discriminantes (DF, do inglês *Discriminant Function*) calculados com o SPA-LDA, usando os dados de cores RGB **(a)** e HSI **(b)** para classificação de vinhos tintos em função da origem geográfica, com os quadrados roxos indicando as amostras verdadeiramente da classe VSF e os quadrados laranjas representando as amostras verdadeiramente da classe Mundo. Na **Figura 5.6 (a)**, a linha discriminante traçada é a fronteira entre as duas classes, com as amostras classificadas como pertencentes à classe VSF localizadas acima dela e as amostras classificadas como pertencentes à classe Mundo situadas abaixo, enquanto na **Figura 5.6 (b)** essas posições estão invertidas.

A leitura da **Figura 5.6 (a)** sugere que a modelagem realizada com SPA-LDA sobre os dados RGB apresenta alta taxa de acerto na classificação, com oito falsos positivos da classe VSF e seis falsos positivos da classe Mundo, além de outras três amostras que se localizam muito próximas à fronteira das classes. O que se vê na **Figura 5.6 (b)**, com os escores da Função Discriminante calculada com os dados HSI, é indicativo de uma maior taxa de classificação correta, com apenas nove falsos negativos (três na classe VSF e seis na classe Mundo), além de três outras amostras que estão muito próximas à fronteira das classes. Os números de erros e de acertos em cada um dos conjuntos de treinamento e de testes, bem como os valores da taxa de classificação correta (TCC), das sensibilidades e especificidades estão resumidos na matriz de confusão mostrada na **Tabela 5.1**, assim como os resultados obtidos com o algoritmo PLS-DA.

Figura 5.6 – Gráficos dos escores das Funções Discriminantes (DF) calculadas com o SPA-LDA, usando os dados de cores RGB **(a)** e HSI **(b)**, para classificação de vinhos tintos em função da origem geográfica: VSF (■) e Mundo (■).



Fonte: elaborados pelo autor.

Tabela 5.1 – Matriz de confusão dos modelos de classificação construídos com os algoritmos SPA-LDA e PLS-DA, baseado em dados de imagens digitais nos sistemas de cores RGB e HSI.

SPA-LDA									
Classe assinalada		RGB				HSI			
		Validação cruzada		Teste		Validação cruzada		Teste	
		VSF	Mundo	VSF	Mundo	VSF	Mundo	VSF	Mundo
Classe VSF		39	5	16	4	41	3	20	0
verdadeira Mundo		5	20	2	9	3	22	2	9
Sensibilidade (%)		89	80	80	82	93	88	100	82
Seletividade (%)		80	89	82	80	88	93	82	100
TCC (%)		85,5		80,7		91,3		93,4	

PLS-DA									
Classe assinalada		RGB				HSI			
		Validação cruzada		Teste		Validação cruzada		Teste	
		VSF	Mundo	VSF	Mundo	VSF	Mundo	VSF	Mundo
Classe VSF		30	14	9	11	37	7	16	4
verdadeira Mundo		1	24	1	10	2	23	1	10
Sensibilidade (%)		68	96	90	48	84	92	80	91
Seletividade (%)		96	68	48	90	92	84	91	80
TCC (%)		78,3		61,3		87,0		83,9	

Fonte: elaborada pelo autor.

Os modelos obtidos com o algoritmo SPA-LDA mostram bons resultados, com baixos números de variáveis selecionadas (3 e 7, respectivamente, para os dados RGB e HSI), elevadas sensibilidades, especificidades e TCC, que em qualquer dos conjuntos (treinamento/validação cruzada ou teste) e em ambos os sistemas de cores usados, superam 80%. Contudo, o modelo baseado nos dados HSI apresentou melhores resultados, com TCC de 91,3% na validação do modelo e de 93,4% na estimação das classes das amostras do conjunto de teste, valores tão bons quanto os alcançados por outros estudos que usaram espectrometria de absorção atômica e cromatografia gasosa, técnicas muito mais robustas e caras que a que se aplica neste trabalho (MORET; SCARPONI; CESCO, 1994; GARCIA-PARRILLA et al., 1997). Há que se destacar a alta sensibilidade da classe VSF, tanto do modelo quanto da estimação, obtidos com os dados HSI, indicando alto desempenho do modelo em reconhecer as amostras desta classe (VSF) frente a amostras de outras classes, com sensibilidade de 100% no conjunto de teste. A modelagem realizada com SPA-LDA somou 16 erros de classificação com os dados RGB, sendo 10 erros de classificação no conjunto de treinamento e 6 no conjunto de teste. Usando os

dados HSI, apenas 8 erros de classificação foram cometidos, 6 nas amostras de treinamento e, somente, 2 no conjunto de teste.

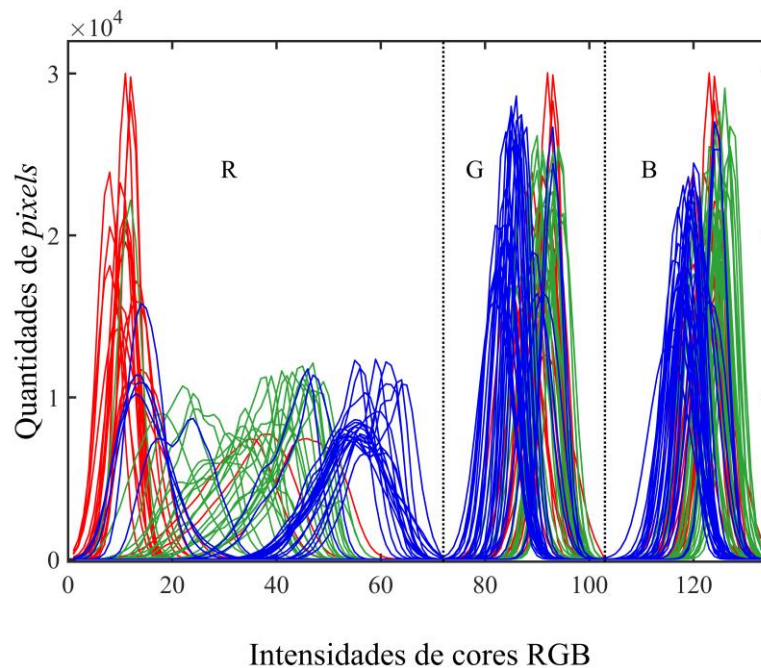
Uma análise discriminante com regressão pelo método dos mínimos quadrados parciais (PLS-DA) foi realizada, independentemente, sobre os dados RGB e HSI. Os cálculos indicaram que uma variável latente seria suficiente para a modelagem dos dados RGB, enquanto 5 variáveis latentes foram requeridas para o modelo de classificação dos vinhos com os dados HSI.

A modelagem PLS-DA também apresentou bons resultados com os dados HSI, onde todas as métricas de avaliação do modelo (especificidade, sensibilidade e TCC) apresentaram valores superiores a 80%. Contudo, o modelo obtido com os dados RGB não apresentou resultados tão bons, com 15 erros na validação e outros 12 na estimação do conjunto de teste, com TCC, respectivamente, igual a 78,3% e 61,3%, sendo seu desempenho também inferior ao modelo obtido com os dados de cores RGB pelo algoritmo SPA-LDA.

5.2 Classificação com base nos vitivinicultores

Os histogramas das 70 amostras usadas na abordagem de classificação por vitivinicultores são mostrados na **Figura 5.7**, com os dados de cores RGB e seleção *à priori* de variáveis. As amostras das classes 1, 2 e 3 estão representadas, respectivamente, pelas linhas vermelhas, verdes e azuis. Regiões de separação das classes são visíveis nas variáveis da componente de cor R, na faixa entre as variáveis 1 e 20 destacam-se mais fortemente as amostras da classe 1, na faixa entre 20 e 50 destacam-se mais claramente as amostras da classe 2 e as amostras da classe 3 mostram-se mais destacadas na faixa entre as variáveis 50 e 80. Nas variáveis referentes às componentes de cores G e B também há afastamentos entre as classes, contudo em menor proporção. Na mesma figura, é possível observar ainda que uma amostra da classe 2 parece sobreposta às amostras da classe 1, enquanto duas amostras da classe 1 aparecem sobrepostas às da classe 2 e algumas amostras da classe 3 aparecem distantes das demais amostras de sua classe.

Figura 5.7 – Histogramas de cores das 70 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função dos seus produtores, após seleção *a priori* de variáveis, com dados RGB. Classes 1 (—), 2 (—) e 3 (—).



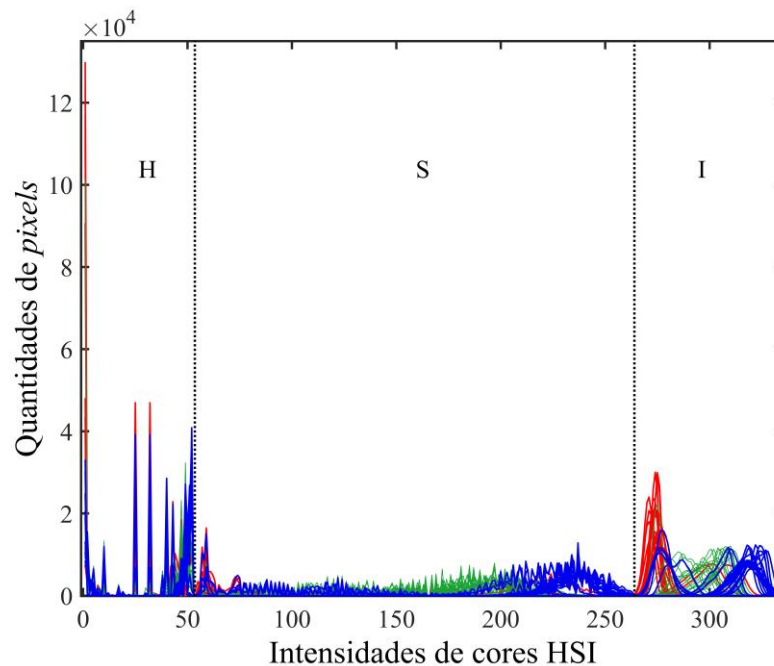
Fonte: elaborado pelo autor.

Nota: embora os dados de um histograma sejam discretos, para melhor visualização, aqui estão representados por linhas contínuas.

Os histogramas de cores das 70 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função dos seus produtores, com dados HSI após seleção *a priori* de variáveis são mostrados na **Figura 5.8**, onde é possível se observar uma sobreposição das classes, nas variáveis das componentes H e S. Nas variáveis da componente I, percebe-se uma separação das classes, semelhante ao que foi observado na componente R do sistema de cores RGB.

Estas observações guardam concordância com o conhecimento de que a composição da matéria corante é função das técnicas enológicas empregadas pelos vitivinicultores, bem como da influência das condições de cultivo das videiras (MONAGAS; BARTOLOMÉ, 2009). Não se pode, entretanto, fazer referência a qual antocianina poderia causar as alterações de cores, uma vez que a técnica analítica utilizada não mede diretamente o comprimento de onda emitido, já que usa filtros de cores (FRASER; BANKS, 2013; GONZALES; WOODS, 2013; SOLOMON; BRECKON, 2013).

Figura 5.8 – Histogramas de cores das 70 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função dos seus produtores, após seleção *a priori* de variáveis, com dados HSI. Classes 1 (—), 2 (—) e 3 (—).



Fonte: elaborado pelo autor.

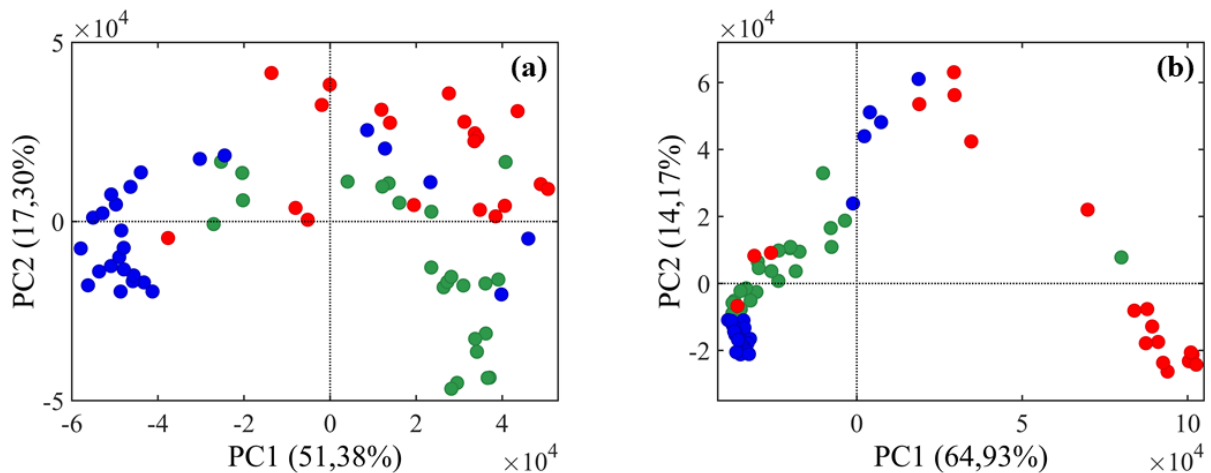
Nota: embora os dados de um histograma sejam discretos, para melhor visualização, aqui estão representados por linhas contínuas.

5.2.1 Análise exploratória com PCA

A análise exploratória com PCA foi realizada e os gráficos dos escores da primeira componente principal (PC1) *versus* os escores da segunda componente principal (PC2) são mostrados na **Figura 5.9**, calculados a partir dos histogramas de intensidades de cores RGB (a) e HSI (b), onde é possível notar tendências de agrupamentos entre as amostras, conforme as classes a que efetivamente pertencem.

Conforme se observa na **Figura 5.9 (a)** a variância explicada da PC1 é de 51,4% e da PC2 17,3%. Esta figura mostra forte tendência de agrupamentos, com amostras da classe 3 dominando a extensão negativa da PC1, mais intensamente na faixa entre -60.000 e -40.000, e as classes 1 e 2 se sobrepondo na faixa positiva, enquanto a PC2 parece melhor separar as amostras da classe 1, mais presente na região positiva, das amostras da classe 2 que dispõem melhor na região negativa.

Figura 5.9 – Gráfico de escores PC1 \times PC2 de classificação de vinhos do Vale do São Francisco em função dos seus produtores, baseado no sistema de cores RGB (a) e HSI (b). Classes 1 (●), 2 (●) e 3 (●).



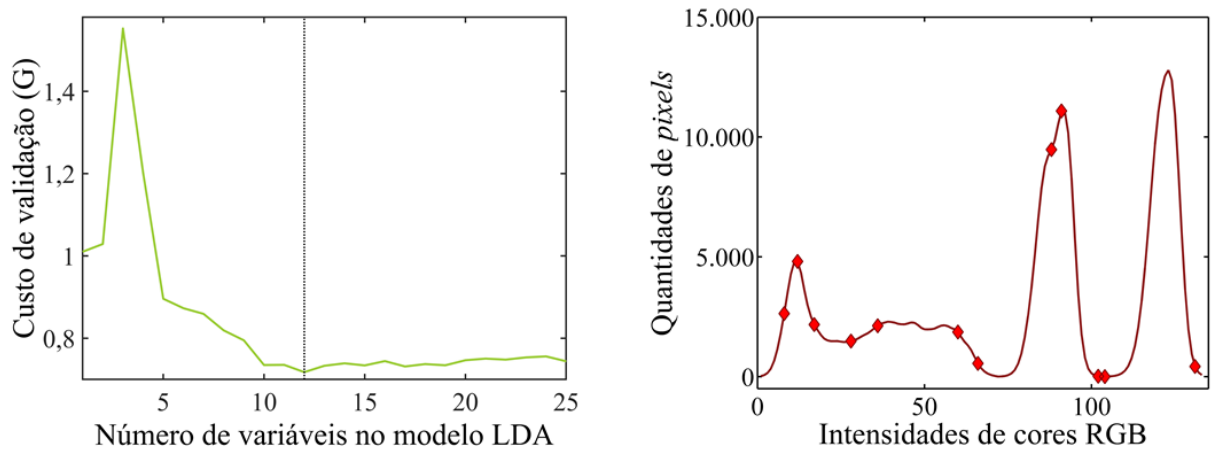
Fonte: elaborados pelo autor.

Tendências de agrupamentos são bem visíveis na **Figura 5.9 (b)**, com as amostras da classe 1 mais agrupadas sobre a PC1, com escores acima de 20.000, e afastadas das outras duas classes que se mostram sobrepostas nesta componente principal. A PC2 parece mais útil como indicadora de separação entre as classes 2 e 3, enquanto as amostras da classe 2 têm escores maiores que -10.000, a maioria das amostras da classe 3 estão bem agrupadas com escores entre -25.000 e -10.000. Uma amostra da classe 2 se encontra afastada das demais amostras desta classe, estando muito próxima das amostras da classe 1, enquanto duas amostras da classe 1 estão bem sobrepostas na classe 2, reforçando o que foi observado na análise dos histogramas. Outras cinco amostras da classe 3 se encontram mais afastadas do seu *cluster*, se posicionando no lado positivo da PC2 e entre as classes 1 e 2 na PC1 (com escores abaixo de 20.000). As variâncias explicadas da PC1 e da PC2 são iguais a 64,9% e 14,2%, respectivamente.

5.2.2 Análise discriminante com SPA-LDA e com PLS-DA

A **Figura 5.10** e mostra os gráficos do custo de validação (G) em função do número de variáveis incluídas no modelo de classificação (a) e as variáveis selecionadas para inclusão no modelo SPA-LDA (b), usando os histogramas de intensidades de cores RGB. A análise da **Figura 5.10 (a)** mostra que o menor custo de validação é obtido com 12 variáveis e se percebe na **Figura 5.10 (b)** que 7 dessas variáveis pertencem à componente R, corroborando com o que foi observado quando da discussão acerca dos histogramas de cores RGB. Quanto às demais variáveis, 4 são intensidades de cores da componente G e apenas 1 é da componente B.

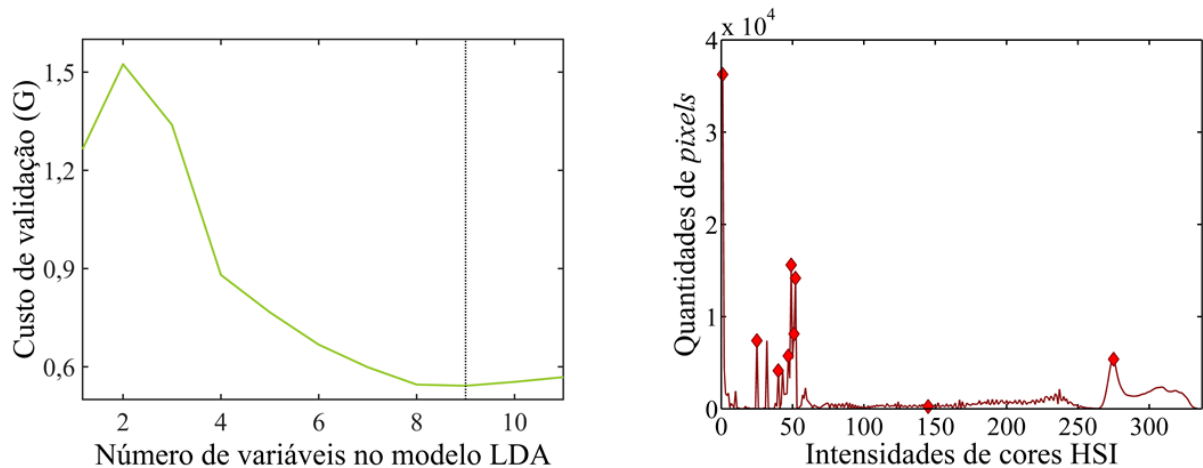
Figura 5.10 – Gráficos dos custos de validação (G) em função do número de variáveis (a) e das variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras (b), usando dados RGB.



Fonte: elaborados pelo autor.

Na **Figura 5.11** são mostrados os gráficos do custo de validação (G) em função do número de variáveis incluídas no modelo de classificação (a) e as variáveis selecionadas para inclusão no modelo SPA-LDA (b), usando os histogramas de intensidades de cores do sistema HSI.

Figura 5.11 – Gráficos dos custos de validação (G) em função do número de variáveis (a) e das variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras (b), usando dados HSI.



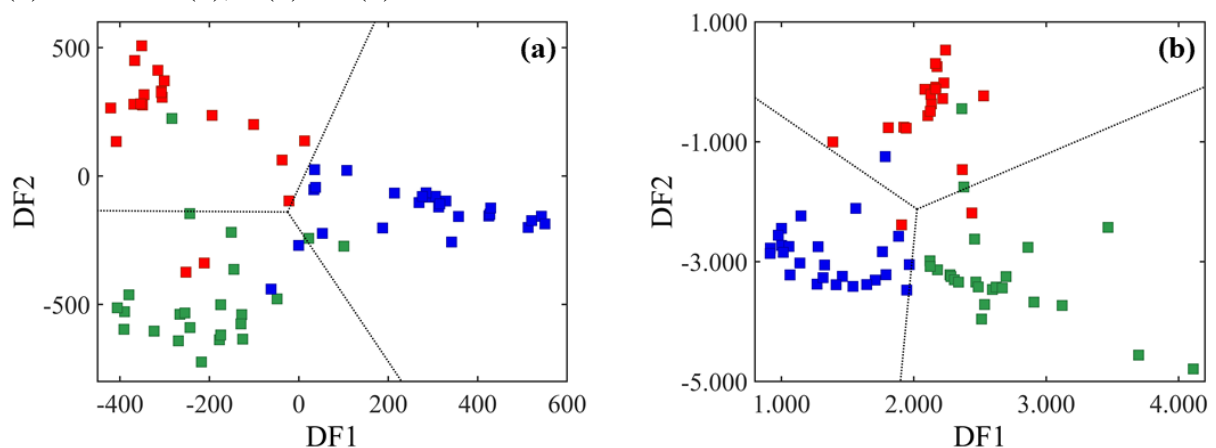
Fonte: elaborados pelo autor.

Mais parcimoniosa que a modelagem dos dados RGB, o menor custo de validação obtido com os dados HSI é obtido com 9 variáveis incluídas no modelo SPA-LDA, como se ver na **Figura 5.11 (a)**, com 7 delas pertencentes à componente H, 1 pertencente à componente S e 1 pertencente à componente I, mostradas na **Figura 5.11 (b)**. É possível inferir que o tom ou matiz de vermelho dos vinhos do VSF influencia fortemente na discriminação entre os

viticultores estudados nesta pesquisa, havendo também significativa diferença na saturação e no nos níveis de cinza deste matiz, indicando que dependendo do vitivinicultor que produza, o vinho poderá ser mais claro ou mais escuro, isto é, com diferentes intensidades de cinza, o que poderia ser provocado pelo nível da etapa de filtração adotada ou pela ausência dela.

A **Figura 5.12** mostra os gráficos dos escores das amostras nas funções discriminantes calculados com o SPA-LDA, usando os dados de cores RGB **(a)** e HSI **(b)**, as cores dos quadrados indicam a classe verdadeira a que pertence cada amostra, sendo vermelho para as amostras da classe 1, verde para as amostras da classe 2 e azul para as amostras da classe 3. Percebe-se que duas DF foram calculadas para cada uma das modelagens (com os dados RGB e HSI), portanto, três linhas discriminantes são traçadas de modo a formar a fronteira entre as três classes.

Figura 5.12 – DF SPA-LDA, com sistema de cores RGB, para classificação de vinhos do Vale do São Francisco em função dos seus produtores, baseado no sistema de cores RGB **(a)** e HSI **(b)**. Classes 1 (■), 2 (■) e 3 (■).



Fonte: elaborados pelo autor.

Na **Figura 5.12 (a)**, é possível observar que três amostras da classe 1 (vermelho) estão classificadas erradamente, duas falsamente pertencentes à classe 2 e uma falsamente pertencente à classe 3. Na mesma figura, percebe-se que duas amostras da classe 2 são falsamente classificadas como pertencentes à classe 3 e uma amostra pertencente à classe 3 é falsamente classificada como pertencente à classe 1. Também se percebe que duas amostras da classe 3 são falsamente classificadas como pertencentes à classe 2 e todas as demais estão corretamente classificadas. Com base nesta figura, é possível afirmar que, ao menos, 8 erros de classificação são cometidos pelo modelo SPA-LDA construído com os dados de cores RGB. A análise da **Figura 5.12 (b)** mostra que um número menor de erros de classificação é cometido pelo modelo SPA-LDA construído com os dados de cores HSI, com apenas 5 erros. Dois erros

de classificação são observados com amostras pertencentes à classe 1, sendo que uma delas foi falsamente classificada como sendo da classe 2 e outra falsamente classificada como pertencente à classe 3. 2 amostras da classe 2 e 1 amostra da classe 3 foram falsamente classificadas como pertencentes à classe 1. Estes resultados estão alinhados ao que se observou quando da análise dos histogramas, onde foi percebido que algumas amostras se afastavam das demais da sua classe.

Na **Tabela 5.2**, é mostrada a matriz de confusão com os resultados dos modelos de classificação construídos com os algoritmos SPA-LDA, para a classificação de vinhos do Vale do São Francisco em função dos vitivinicultores e baseado em imagens digitais em sistemas de cores RGB e HSI. Esta tabela mostra também os resultados dos modelos de classificação obtidos com PLS-DA.

Tabela 5.2 – Matriz de confusão dos modelos de classificação construídos com os algoritmos SPA-LDA e PLS-DA para a classificação de vinhos do Vale do São Francisco, em função dos vitivinicultores e baseados em imagens digitais, em sistemas de cores RGB e HSI.

SPA-LDA													
Classe assinalada		RGB						HSI					
		Validação cruzada			Teste			Validação cruzada			Teste		
		1	2	3	1	2	3	1	2	3	1	2	3
Classe verdadeira	1	12	2	-	6	-	-	12	1	1	5	-	-
	2	1	13	2	-	8	-	2	14	-	-	8	-
	3	-	2	16	-	-	8	1	-	17	-	-	8
Sensibilidade (%)		86	81	89	100	100	100	86	88	94	100	100	100
Seletividade (%)		97	88	93	100	100	100	91	97	97	100	100	100
TCC (%)		85,4			100			89,6			100		

PLS-DA													
Classe assinalada		RGB						HSI					
		Validação cruzada			Teste			Validação cruzada			Teste		
		1	2	3	1	2	3	1	2	3	1	2	3
Classe verdadeira	1	11	2	1	6	-	-	12	2	-	5	1	-
	2	2	14	-	-	8	-	1	13	2	-	8	-
	3	3	3	12	1	-	7	1	3	14	-	-	8
Sensibilidade (%)		79	88	67	100	100	88	86	81	78	100	89	100
Seletividade (%)		85	84	97	94	100	100	94	84	93	100	93	100
TCC (%)		77,1			95,5			81,3			95,5		

Fonte: elaborada pelo autor.

Na modelagem realizada com SPA-LDA, os dois sistemas de cores usados neste estudo, RGB e HSI, mostraram desenvolvimentos semelhantes, sendo que o modelo HSI cometeu 2 erros a

menos e foi mais parcimonioso com relação ao número de variáveis incluídas no modelo (9). Ao todo, sete erros de classificação foram observados quando da modelagem com os dados de cores RGB, contra 5 erros quando a modelagem ocorreu com dados HSI. Quando se compara os erros dos conjuntos de treinamento e de teste, as modelagens mostraram melhores respostas no conjunto de testes, em que os dois modelos obtiveram TCC de 100%, enquanto no conjunto de treinamento, a taxa de classificação correta foi de 85,4% e 89,6%, respectivamente para os dados RGB e HSI.

Os modelos de classificação obtidos com PLS-DA indicaram a necessidade de 3 e 4 variáveis latentes, respectivamente, para os dados RGB e HSI. As variâncias explicadas acumuladas foram de 79% para os dados RGB e 91% para os dados HSI.

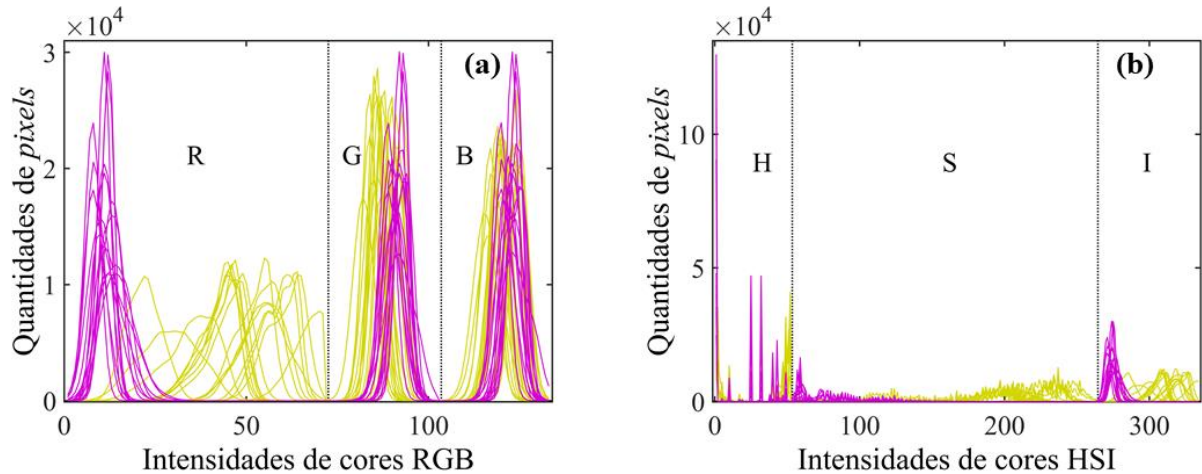
Os resultados mostrados na **Tabela 5.2** indicam que os modelos PLS-DA foram menos robustos que os obtidos com SPA-LDA, mais notadamente aqueles obtidos para os conjuntos de treinamento, com total de 11 e 10 erros, respectivamente, com as modelagens nos dados RGB e HSI. Analisando-se apenas os conjuntos de teste, os resultados foram iguais entre si, para as modelagens com dados de cores RGB e HSI, com TCC igual a 95,5% e muito próximo ao desempenho obtido com SPA-LDA. A maior diferença entre os dois algoritmos se mostrou mais intensamente nos conjuntos de treinamento, onde as taxas de classificação corretas obtidas foram de 77,1% e 81,3%, respectivamente para as modelagens dos dados RGB e HSI, ainda que inferiores aqueles obtidos com SPA-LDA, podem servir como triagem preliminar, já que foram capazes de prever corretamente a classificação das amostras dos conjuntos de teste.

5.3 Classificação baseada nas variedades de uvas da composição dos vinhos varietais

5.3.1 *Cabernet Sauvignon x Touriga Nacional*

A **Figura 5.13** mostra os histogramas de intensidades de cores, nos sistemas de cores RGB (**a**) e HSI (**b**), de 34 amostras de vinhos varietais de diferentes produtores do Vale do São Francisco, sendo 18 produzidos com uvas da variedade Cabernet Sauvignon (CS) e 16 produzidos com uvas da variedade Touriga Nacional (TN). Os histogramas das amostras de vinhos CS estão indicados pelas linhas amarelas, enquanto os histogramas de TN estão indicados pelas linhas magentas, após seleção *a priori* de variáveis.

Figura 5.13 – Histogramas de intensidades de cores RGB das 34 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função das variedades das uvas, Cabernet Sauvignon (—) vs Touriga Nacional (---), com dados RGB **(a)** e HSI **(b)**, após seleção *a priori* de variáveis.



Fonte: elaborado pelo autor.

Nota: embora os dados de um histograma sejam discretos, para melhor visualização, aqui estão representados por linhas contínuas.

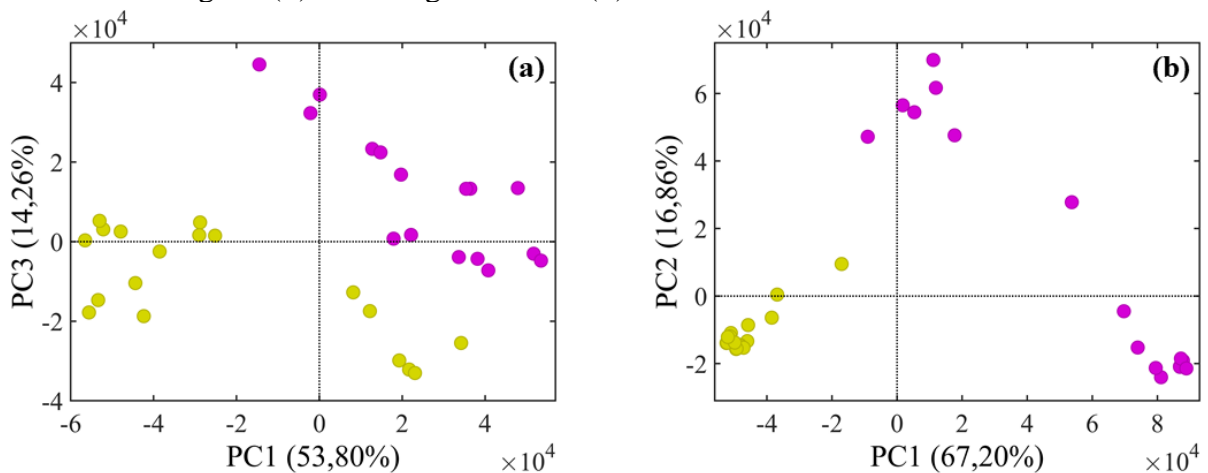
As primeiras variáveis da **Figura 5.13 (a)**, referentes à componente R do sistema de cores RGB, sugerem boa separação das amostras em suas classes CS e TN, com as amostras de TN se mostrando com menores intensidades de vermelho (R) que as CS. Também se observa, na própria componente R, que respostas analíticas (distribuição de frequência) referentes às amostras da classe CS se mostram espalhadas, como se formassem *clusters* ou subclasses, que pode ser consequência da produção por diferentes vitivinicultores, como analisado no item 5.2 deste trabalho. Os componentes G e B, ainda que possuam menor número de variáveis significativas, isto é, se apresentam com picos mais estreitos, também parecem úteis para discriminar as amostras em suas classes CS e TN.

Os histogramas de intensidades de cores HSI, mostrados na **Figura 5.13 (b)**, sugerem também boa separação entre as classes CS e TN. Nas variáveis referentes à componente intensidade (I), mais uma vez as amostras da classe TN se mostram posicionadas com menores intensidades, mais próximas do preto, do que as amostras da classe CS, que também se apresentam com maiores valores de saturação. De fato, os vinhos produzidos com uvas da variedade Touriga Nacional tendem a apresentar com coloração escura, como descreve Amarante (2015).

5.3.1.1 *Análise exploratória com PCA*

Os gráficos dos escores da PCA são mostrados na **Figura 5.14**, PC1 *versus* PC3, para os dados RGB **(a)** e PC1 *versus* PC2, para os dados HSI **(b)**. Como mostrado na **Figura 5.14 (a)**, na PCA realizada sobre os dados RGB, apesar de se mostrarem dispersas, as amostras estão visivelmente separadas em classes, com as amostras da classe CS formando por pelo menos dois *clusters*, como foi observado na análise dos histogramas, um deles dominando a parte negativa da PC1 (com escores abaixo de -20.000) e o outro na parte positiva da PC1 e próximos da Classe TN, mas com escores negativos na PC3 (abaixo de -10.000). As amostras da variedade TN também se encontram espalhadas, com escores acima de -15.000 na PC1 e de -10.000 na PC3. A variância explicada acumulada nas três primeiras PC soma 82,6%, indicando boa modelagem.

Figura 5.14 – Gráficos dos escores de PCA baseado nos dados de cores RGB **(a)** e HSI **(b)** para a classificação de vinhos do Vale do São Francisco em função das variedades das uvas Cabernet Sauvignon (●) vs Touriga Nacional (●).



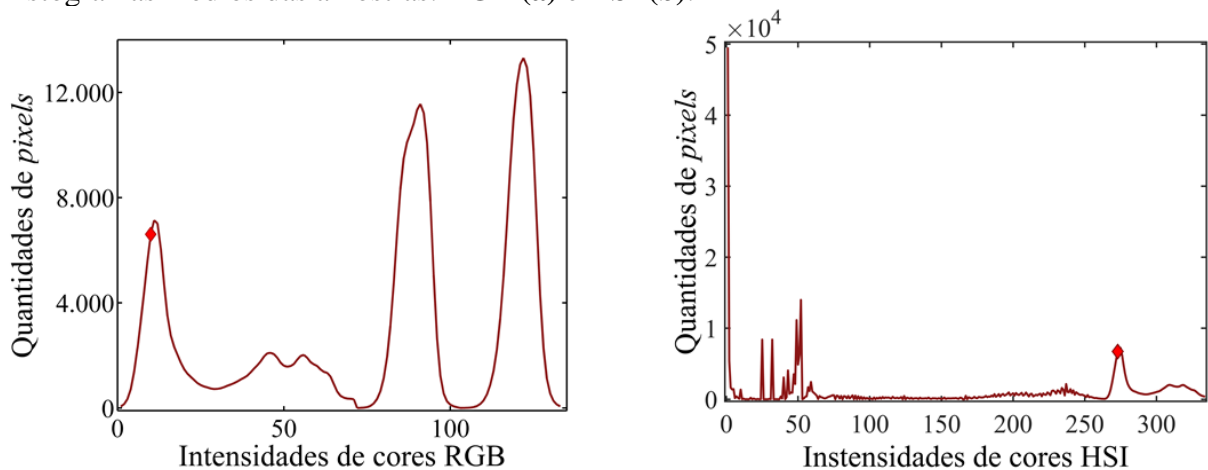
Fonte: elaborado pelo autor.

Para os dados HSI, mostrados na **Figura 5.14 (b)**, as amostras da Classe CS se mostram melhor agrupadas que as amostras da Classe TN, que tendem a formar duas subclasses que podem ser consequências da origem relativa aos vitivinicultores. Contudo, é perceptível a tendência de separação das classes, com as amostras da Classe CS mais posicionadas a esquerda da PC1, com escores abaixo de -10.000, e entre -20.000 e 8.000 na PC2, enquanto as amostras da Classe TN têm escores acima de -10.000 na PC1. De fato, os escores da PC1, com variância explicada igual a 67,2 %, já seriam suficientes para separar as duas classes.

5.3.1.2 Análise discriminante com SPA-LDA e PLS-DA

Os gráficos das variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras, são mostrados na **Figura 5.15**, os dados RGB (**a**) e HSI (**b**). Na **Figura 5.15 (a)**, percebe-se que apenas uma variável foi requerida para modelar os dados de cores RGB, ou seja, uma única intensidade da componente R foi usada pelo SPA-LDA na construção do modelo para discriminar as duas classes (CS e TN), como sugere a análise dos histogramas. Isto sugere que os tons de vermelho são suficientes para discriminar amostras de vinho entre as castas Cabernet Sauvignon e Touriga Nacional. A análise da **Figura 5.15 (b)** indica que, se usado o sistema de cores HSI, também uma única variável, da componente I, é requerida na construção do modelo.

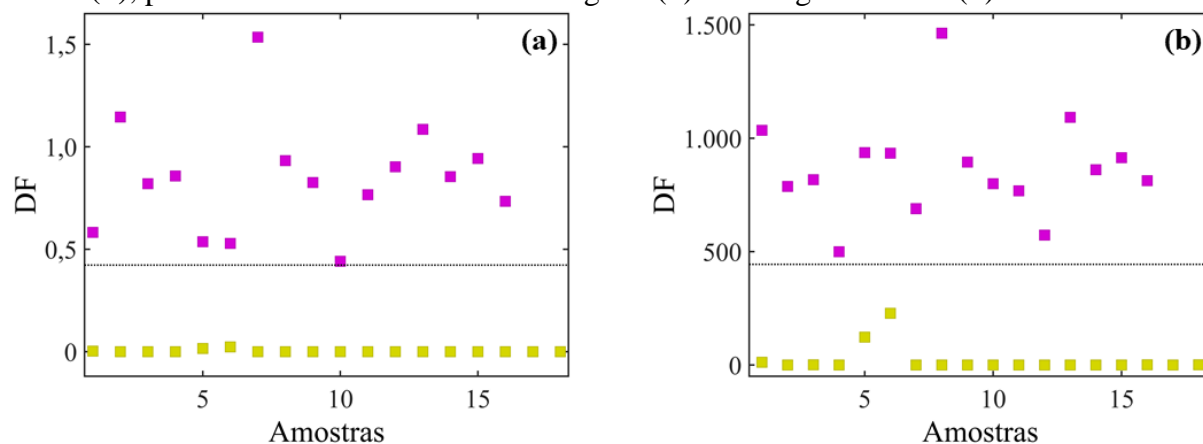
Figura 5.15 – Gráficos das variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras: RGB (**a**) e HSI (**b**).



Fonte: elaborados pelo autor.

Os gráficos dos escores das funções discriminantes são mostrados na **Figura 5.16** para as modelagens realizadas com SPA-LDA sobre os dados RGB (**a**) e HSI (**b**). As amostras da classe CS são representadas por quadrados amarelos e são classificadas corretamente acima da linha discriminante. As amostras da classe TN são representadas pelos quadrados magentas e sua classificação é indicada abaixo da linha discriminante. A análise da figura mostra que todas amostras foram corretamente classificadas, ou seja, a classificação obteve taxa de classificação correta de 100%, sugerindo que, tanto o sistema de cores RGB quanto o HSI são capazes de discriminar vinhos varietais produzidos com uvas Cabernet Sauvignon de vinhos produzidos com a casta Touriga Nacional.

Figura 5.16 – Gráficos das Funções Discriminantes (DF) da classificação de vinhos do Vale do São Francisco em função das variedades das uvas baseado nos sistemas de cores RGB (a) e HSI (b), para as variedades Cabernet Sauvignon (■) e Touriga Nacional (■).



Fonte: elaborados pelo autor.

A análise discriminante com PLS-DA requereu uma variável latente quando usados os dados RGB e duas variáveis para modelar os dados de cores HSI.

A **Tabela 5.3** é a matriz de confusão dos modelos construídos com o algoritmo PLS-DA para a classificação de 34 amostras de vinhos varietais de diferentes produtores do Vale do São Francisco, produzidos com as variedades Cabernet Sauvignon (CS) e Touriga Nacional (TN). Esta tabela mostra que, foi possível classificar corretamente todas as amostras de vinhos varietais das variedades Cabernet Sauvignon e Touriga Nacional, isto é, com taxa de classificação correta igual a 100%, baseando-se em histogramas de cores dos sistemas RGB ou HSI e PLS-DA.

Tabela 5.3 – Matriz de confusão do modelo de classificação de vinhos do Vale do São Francisco em função das variedades de uvas Cabernet Sauvignon e Touriga Nacional, baseado em imagens digitais em sistemas de cores RGB e HSI e construído com o algoritmo PLS-DA.

		RGB				HSI			
		Validação cruzada		Teste		Validação cruzada		Teste	
Classe assinalada		CS	TN	CS	TN	CS	TN	CS	TN
Classe verdadeira	CS	12	-	6	-	12	-	6	-
	TN	-	11	-	5	-	11	-	5
Sensibilidade (%)		100	100	100	100	100	100	100	100
Seletividade (%)		100	100	100	100	100	100	100	100
TCC (%)		100		100		100		100	

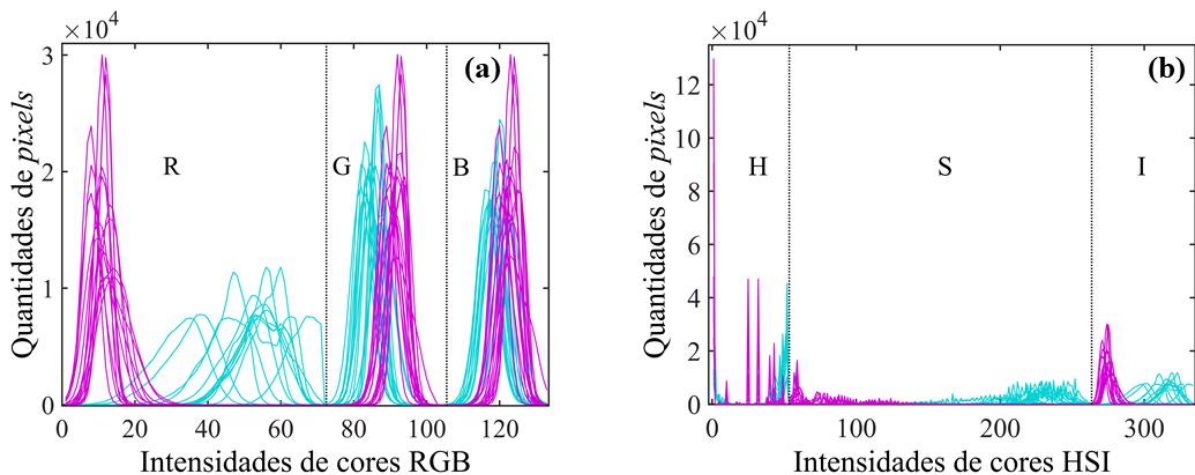
Fonte: elaborada pelo autor.

Os resultados obtidos com os algoritmos SPA-LDA e PLS-DA para a classificação de vinhos do VSF em função das uvas CS e TN foram idênticos.

1.3.2 Syrah x Touriga Nacional

Na **Figura 5.17**, são mostrados os histogramas de intensidades de cores das 30 amostras de vinhos varietais de diferentes produtores do Vale do São Francisco, sendo 14 produzidos com uvas da variedade Syrah (Sy) e 16 produzidos com uvas da variedade Touriga Nacional (TN), com dados de cores RGB **(a)** e HSI **(b)** após seleção *a priori* de variáveis com respostas nulas. Os histogramas das amostras da classe Sy estão indicadas pelas linhas ciano e as da classe TN pelas linhas magentas. Nesta figura, observa-se boa separação entre as duas classes de amostras, em qualquer das componentes de cores dos sistemas RGB e HSI. Provavelmente, uma ou duas variáveis, em cada um dos sistemas de cores, serão suficientes para discriminar as duas classes de amostras.

Figura 5.17 – Histogramas de cores das 30 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função das variedades das uvas Syrah (–) vs Touriga Nacional (–), baseado nos sistemas de cores RGB **(a)** e HSI **(b)**, após seleção *a priori* de variáveis.



Fonte: elaborados pelo autor.

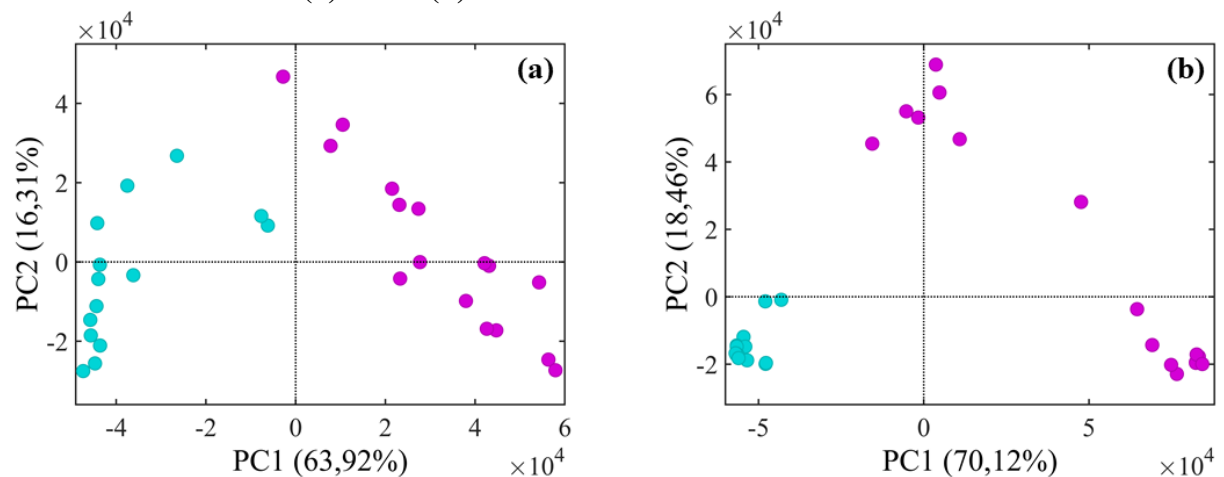
Nota: embora os dados de um histograma sejam discretos, para melhor visualização, aqui estão representados por linhas contínuas.

1.3.2.1 Análise exploratória com PCA

A **Figura 5.18** mostra os gráficos de escores das PCA realizadas, independente e respectivamente, com os dados RGB **(a)** e HSI **(b)**, nas quais é possível visualizar perfeita separação entre as amostras da casta Syrah daquelas produzidas com a casta Cabernet Sauvignon. Na **Figura 5.18 (a)**, vê-se que as amostras da Classe Sy estão posicionadas com escores inferiores a -7.000, na PC1, enquanto as amostras da Classe TN apresentam escores

superiores a -5.000. Logo, é possível inferir que uma única PC é suficiente para sugerir uma boa discricionariiedade das duas classes. Já na **Figura 5.18 (b)**, observa-se que as amostras da Classe Sy estão bem agrupadas e posicionadas na PC1, com escores entre -60.000 e -40.000, enquanto as amostras da Classe TN apresentam escores acima de -20.000. Isto é, com os dados HSI também é possível inferir que uma única PC é suficiente para sugerir uma boa discricionariiedade das duas classes. As variâncias explicadas das duas primeiras PC são, respectivamente, 70,1% e 18,5%.

Figura 5.18 – Gráfico de escores PC1 x PC2 da classificação de vinhos do Vale do São Francisco em função das variedades das uvas Syrah (●) e Touriga Nacional (●), baseado no sistema de cores RGB (a) e HSI (b).

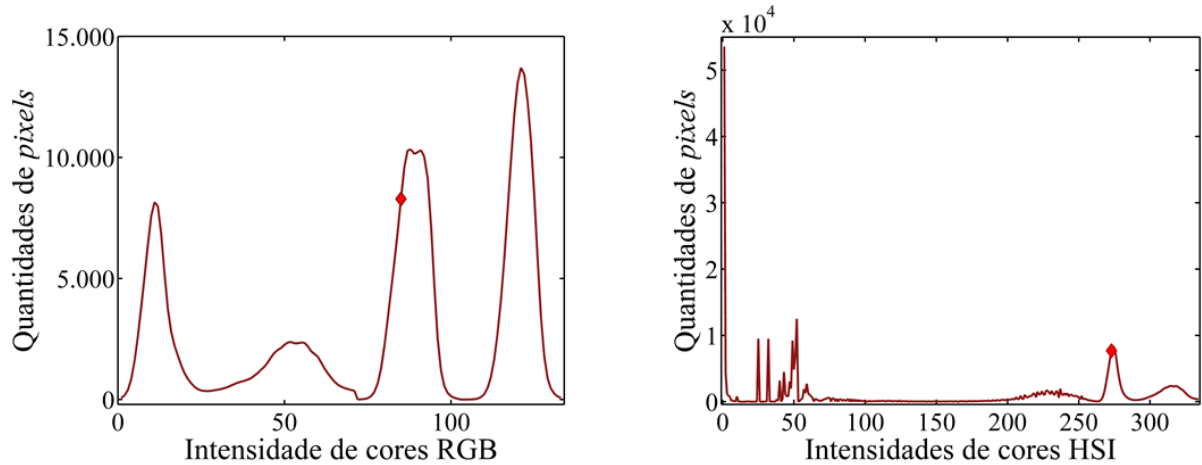


Fonte: elaborados pelo autor.

1.3.2.2 Análises discriminantes com SPA-LDA e PLS-DA

Conforme pode ser visto na **Figura 5.19** a modelagem com SPA-LDA indicou que, para discriminar as amostras entre as classes Syrah e Touriga Nacional, é necessária única variável igualmente para os modelos construídos com os dados RGB (a) e HSI (b). O número de variáveis incluídas para esta modelagem coincidiu com o que foi utilizado na discriminação das amostras de Cabernet Sauvignon e Touriga Nacional, isto é, o SPA-LDA necessitou de uma única variável para os dados RGB, da componente G, e outra para os dados HSI, da componente I.

Figura 5.19 – Gráficos das variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras, dos dados RGB (a) e HSI (b).

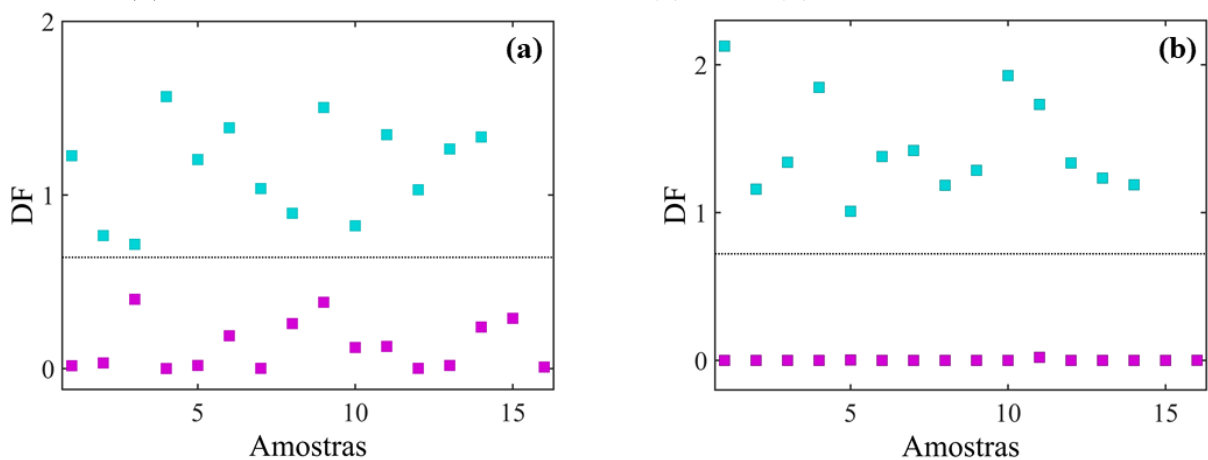


Fonte: elaborados pelo autor.

Os gráficos dos escores das funções discriminantes são mostrados na **Figura 5.20**, para as modelagens realizadas com SPA-LDA sobre os dados RGB (a) e HSI (b). As amostras das classes Sy e TN são representadas, respectivamente, por quadrados cianos e magentas, as amostras da classe Sy são corretamente classificadas acima da linha discriminante e as da classe TN são corretamente classificadas abaixo.

A análise da figura indica que todas amostras foram corretamente classificadas, isto é, a TCC obtida com pelo SPA-LDA com os dados RGB e HSI foram iguais a 100%, indicando que, tanto o sistema de cores RGB quanto o HSI oferecem dados suficientes para classificar vinhos varietais entre Syrah e Touriga Nacional.

Figura 5.20 – Gráfico das Funções Discriminantes (DF) da classificação de vinhos do Vale do São Francisco em função das variedades das uvas, para as variedades Syrah (■) e Touriga Nacional (■), baseado nos sistemas de cores RGB (a) e HSI (b).



Fonte: elaborados pelo autor.

A matriz de confusão dos modelos de classificação construídos com o algoritmo PLS-DA para a classificação de 30 amostras de vinhos varietais de diferentes produtores do Vale do São Francisco, produzidos com as castas Syrah (Sy) e Touriga Nacional (TN) é mostrada na **Tabela 5.4**. Como se pode observar nesta tabela, tanto com os dados RGB quando os dados HSI, os modelos construídos com PLS-DA foram capazes de classificar sem cometer nenhum erro de classificação, isto é, com TCC igual a 100%, as 30 amostras de vinhos varietais das castas Syrah e Touriga Nacional.

Tabela 5.4 – Matriz de confusão do modelo de classificação de vinhos do Vale do São Francisco em função das variedades de uvas Syrah e Touriga Nacional, baseado em imagens digitais em sistemas de cores RGB e HSI e construído com o algoritmo PLS-DA.

Classe Assinalada		RGB				HSI			
		Validação cruzada		Teste		Validação cruzada		Teste	
		Sy	TN	Sy	TN	Sy	TN	Sy	TN
Classe verdadeira	Sy	9	-	5	-	9	-	5	-
	TN	-	11	-	5	-	11	-	5
Sensibilidade (%)		100	100	100	100	100	100	100	100
Seletividade (%)		100	100	100	100	100	100	100	100
TCC (%)		100		100		100		100	

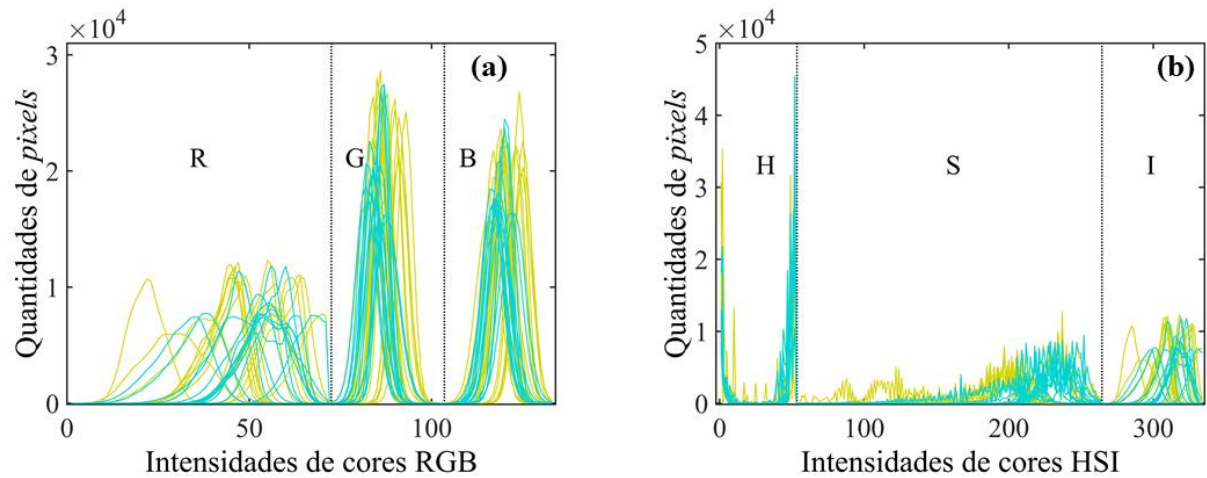
Fonte: elaborada pelo autor.

A diferença entre as duas técnicas quimiométricas foi o número de variáveis utilizadas: enquanto os modelos construídos com SPA-LDA utilizaram uma única variável, os modelos construídos com PLS-DA fizeram uso de duas variáveis latentes.

5.3.4 Cabernet Sauvignon versus Syrah

Na **Figura 5.21**, são mostrados os histogramas de 32 amostras de vinhos varietais do Vale do São Francisco, 18 da variedade Cabernet Sauvignon e 14 da variedade Syrah, nos sistemas de cores RGB (**a**) e HSI (**b**), após realização de seleção *a priori* de variáveis e centragem na média, em que as amostras da classe CS são mostradas pelas linhas amarelas e as da classe Sy pelas linhas cianas. Nos gráficos dos dois sistemas de cores, parece haver grande sobreposição das classes, com umas poucas amostras da Classe CS separando-se das demais, especialmente nas componentes de cores R, B e S.

Figura 5.21 – Histogramas de cores das 32 amostras de vinhos do Vale do São Francisco usadas na modelagem de classificação em função das variedades das uvas Cabernet Sauvignon (—) vs Syrah (—), baseado no sistema de cores RGB (a) e HSI (b), após seleção *a priori* de variáveis.



Fonte: elaborados pelo autor.

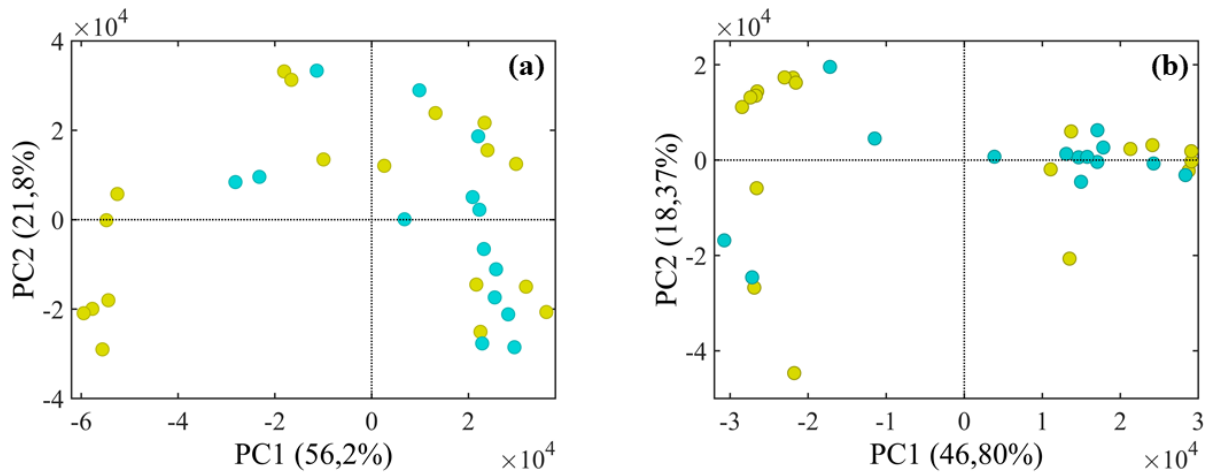
Nota: embora os dados de um histograma sejam discretos, para melhor visualização, aqui estão representados por linhas contínuas.

Diferentemente do que se observou nas comparações anteriormente realizadas, em que se comparou vinhos varietais aos pares, entre as variedades Cabernet Sauvignon *versus* Touriga Nacional e Syrah *versus* Touriga Nacional, e com histogramas das classes apresentando boa separação, quando se compara as amostras de Syrah (Sy) com as Cabernet Sauvignon (CS), parece não haver separação dos histogramas. Nos dados RGB, mostrados na **Figura 5.21 (a)**, parece haver um pequeno afastamento das classes nas variáveis das componentes G e B, mas apenas algumas amostras da classe CS se afastam para a esquerda, mantendo-se outras sobrepostas à classe Sy.

5.3.4.1 Análise exploratória com PCA

Alinhado ao que se observa nos histogramas, os escores de PCA mostrados na **Figura 5.22** para os dados de cores RGB (a) e HSI (b), não indicam tendências de separação das duas classes, exceto por umas poucas amostras da Classe CS que se posicionam no lado mais negativo da PC1, nos dois modelos de cores (seis nos dados RGB e oito com os dados HSI). Uma ou outra amostra parecem mais isoladas, contudo, no geral, há sobreposição das amostras das Classes Sy e CS.

Figura 5.22 – Gráfico de escores de PCA (PC1 x PC2) da classificação de vinhos do Vale do São Francisco em função das variedades das uvas Cabernet Sauvignon (●) vs Syrah (●), baseado no sistema de cores RGB (a) e HSI (b).

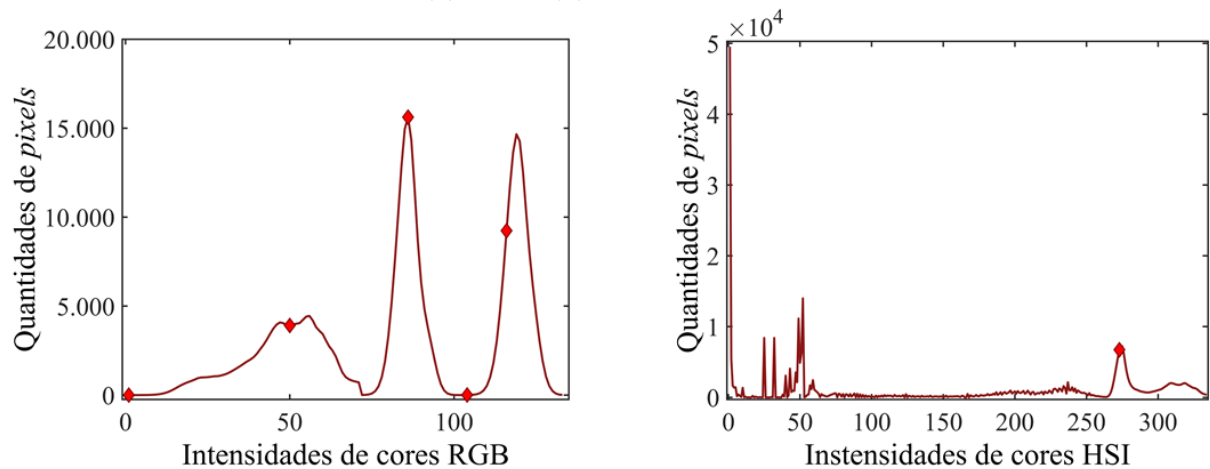


Fonte: elaborados pelo autor.

5.3.4.2 *Análise discriminante com SPA-LDA e PLS-DA*

Na modelagem realizada com SPA-LDA, o menor custo de validação foi obtido quando 5 variáveis foram inseridas no modelo com os dados RGB de cores. Quando os dados HSI foram usados na modelagem, o menor custo de validação foi obtido com 7 variáveis. Na **Figura 5.23**, são mostradas quais variáveis foram incluídas nos modelos SPA-LDA, sobre os histogramas dos sistemas RGB (a) e HSI (b).

Figura 5.23 – Variáveis selecionadas pelo algoritmo SPA-LDA sobre os histogramas médios das amostras dos sistemas RGB (a) e HSI (b).



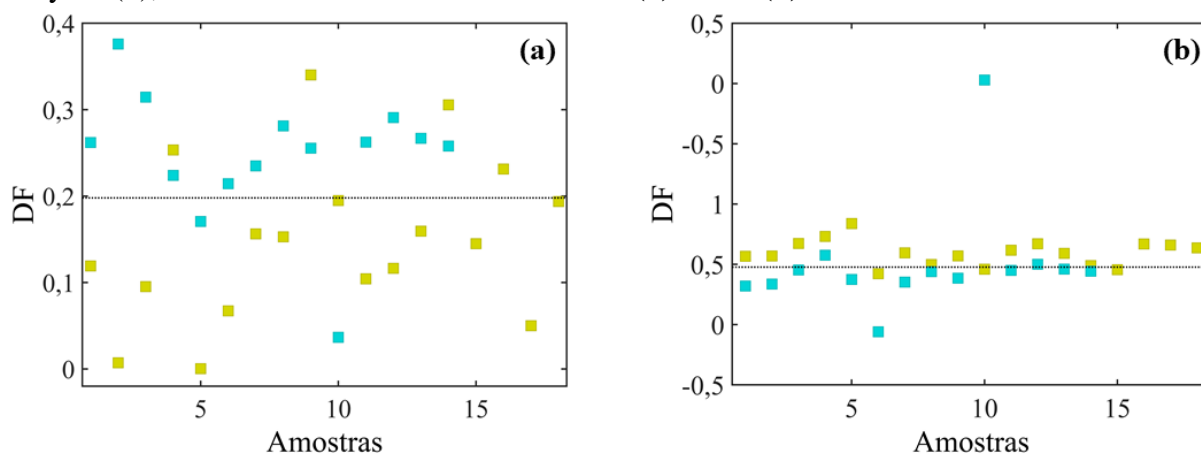
Fonte: elaborados pelo autor.

Na **Figura 5.23 (a)**, observa-se que foram selecionadas 2 variáveis da componente de cores R, 1 da componente G e 2 da componente B, enquanto a **Figura 5.23 (b)** mostra que 5

variáveis da componente H e 2 variáveis da componente S foram seleccionadas, sugerindo que a diferenças entre os vinhos das duas variedades devem ser nos seus matizes de cor e na sua saturação, sendo que os vinhos das duas classes não apresentam diferenças de brilho ou estas diferenças não são úteis para discriminar as duas classes.

Os gráficos dos escores das funções discriminantes são mostrados na **Figura 5.24**, para as modelagens realizadas com SPA-LDA sobre os dados RGB **(a)** e HSI **(b)**, amostras das classes CS e Sy são representadas por quadrados, respectivamente, amarelos e cianos. Na **Figura 5.24 (a)**, a linha discriminante é a fronteira entre as duas Classes, com CS verdadeiramente classificada abaixo e Sy verdadeiramente classificada acima dela. Surpreendentemente, apenas 6 amostras foram falsamente classificadas (2 da Classe Sy e 4 da Classe CS), com um total de acertos correspondente a 81,3%. A análise da **Figura 5.24 (b)**, na qual as amostras da Classe CS estão verdadeiramente posicionadas acima da linha discriminante, com as Sy abaixo, indica que o modelo traçado com os histogramas de cores HSI oferece a mesma taxa de acertos, com 6 amostras falsamente classificadas, aqui 3 erros de classificação ocorreram em cada uma das classes.

Figura 5.24 – Gráfico das Funções Discriminantes (DF) da classificação de vinhos do Vale do São Francisco em função das variedades das uvas, para as variedades Cabernet Sauvignon (■) e Syrah (■), baseado nos sistemas de cores RGB **(a)** e HSI **(b)**.



Fonte: elaborados pelo autor.

A matriz de confusão dos modelos de classificação construídos com os algoritmos SPA-LDA e PLS-DA para a classificação de 32 amostras de vinhos varietais de diferentes produtores do Vale do São Francisco, produzidos com as variedades Cabernet Sauvignon e Syrah é mostrada na **Tabela 5.5**. Como se pode observar nesta tabela, tanto com os dados RGB quanto os HSI, os modelos construídos com SPA-LDA foram capazes de classificar com poucos erros de classificação, com TCC iguais 85,7% e 72,7%, respectivamente, para os conjuntos de

treinamento e de testes, tanto para os dados RGB quanto os HSI, com sensibilidades e especificidades que alcançaram até 89%, dependendo da classe escolhida.

O uso do algoritmo PLS-DA indicou melhor modelagem com 1 e 6 variáveis latentes, respectivamente, para os dados de cores RGB e HSI. Embora os resultados obtidos nos conjuntos de testes tenham sido iguais ou superiores aos obtidos com SPA-LDA, bons resultados não foram obtidos no conjunto de treinamento, cujas taxas de classificação correta ficaram abaixo de 60%.

Tabela 5.5 – Matriz de confusão dos modelos de classificação de vinhos do Vale do São Francisco em função das variedades de uvas Cabernet Sauvignon e Syrah, baseados em imagens digitais em sistemas de cores RGB e HSI.

SPA-LDA									
Classe assinalada		RGB				HSI			
		Validação cruzada		Teste		Validação cruzada		Teste	
		CS	Sy	CS	Sy	CS	Sy	CS	Sy
Classe verdadeira	CS	10	2	4	2	10	2	5	1
	Sy	1	8	1	4	1	8	2	3
Sensibilidade (%)		83	89	67	80	83	89	83	60
Seletividade (%)		89	83	80	67	89	83	60	83
TCC (%)		85,7		72,7		85,7		72,7	

PLS-DA									
Classe assinalada		RGB				HSI			
		Validação cruzada		Teste		Validação cruzada		Teste	
		CS	Sy	CS	Sy	CS	Sy	CS	Sy
Classe verdadeira	CS	5	7	4	2	6	6	5	1
	Sy	2	7	1	4	5	4	1	4
Sensibilidade (%)		42	78	67	80	50	44	83	80
Seletividade (%)		78	42	80	67	44	50	80	83
TCC (%)		57,1		72,7		47,6		81,8	

Fonte: elaborada pelo autor.

Os resultados de classificação inferiores obtidos para a discriminação destas duas classes, comparados aos resultados anteriores, podem se dar em decorrência da grande quantidade de matéria corante que é comum às uvas das variedades Cabernet Sauvignon e Syrah (AMARANTE, 2015) ou ser consequência dos cortes existentes nos vinhos. Sobre a segunda possibilidade, embora se trate de vinhos com indicações varietais nos rótulos, não há informações sobre a composição total das amostras. Sabendo que a legislação brasileira admite que um vinho fino receba no rótulo a designação de uma variedade de uvas *Vitis vinifera* se contiver, ao menos, 75% em volume daquela espécie e que não precisa indicar o restante da

composição (BRASIL, 1990) e, ainda, que as duas variedades de uvas tintas *Vitis vinifera* mais plantadas no VSF são Syrah e Cabernet Sauvignon (AMARANTE, 2015; PEREIRA; VANDERLINDE; LIMA, 2011), poderia se supor que uma variedade estaria complementando, em até 25% em volume, a composição da outra e vice-versa. Caso se confirme a segunda possibilidade, com a presença de cortes de Syrah nos varietais Cabernet Sauvignon e desta no varietais de Syrah, seria provável uma grande proximidade de cores entre as duas classes e uma modelagem com número muito maior de amostras seria demandada. De qualquer modo, uma modelagem mais completa, com um maior número de amostras para cada uma das classes, poderia indicar melhores resultados.

6 CONCLUSÕES

Um instrumento analítico, de baixo custo de montagem e de manutenção, baseado em imagens digitais e com potencial para ser levado à indústria, foi construído e seu pedido de patente foi depositado junto ao Instituto Nacional de Propriedade Industrial (INPI), sob o nome de Dispositivo para Medições Colorimétricas baseado em imagens digitais (DMCdib) (ANEXO D). O instrumento demonstrou capacidade satisfatória para resolver problemas de classificação de vinhos quando usado em bancada no laboratório.

Vinhos do Vale do São Francisco e de outras regiões do mundo foram analisados, com uso do DMCdib. Estudos de classificação foram realizados com ferramentas quimiométricas, através dos quais um estudo exploratório empregando a análise de componentes principais (PCA) sugeriu a existência de padrões que possibilitam diferenciar os vinhos do Vale do São Francisco (VSF) daqueles provenientes de outras regiões e outros padrões que diferenciam os vinhos do VSF em relação aos vitivinicultores e das variedades de uvas usadas na produção de vinhos varietais.

Modelos de classificação de vinhos em função de sua origem geográfica, dos vitivinicultores e das variedades de uvas usadas na produção de vinhos varietais foram construídos usando Análise Discriminante Linear com seleção de variáveis usando o Algoritmo das Projeções Parciais (SPA-LDA) e Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA), usando dados dos modelos de cores RGB e HSI.

Na abordagem de classificação em função da origem geográfica dos vinhos, o modelo SPA-LDA apresentou taxas de classificação corretas (TCC) nos conjuntos de treinamento e de teste, respectivamente, de 85,5% e 80,7%, com os dados RGB e de 91,3% e 93,6%, com os dados HSI. O modelo construído com PLS-DA apresentou TCC nos conjuntos de treinamento e de teste, respectivamente, de 87,0% e 61,3%, com os dados RGB e de 85,6% e 83,9%, com os dados RGB. Os melhores resultados foram obtidos com os modelos construídos usando o algoritmo SPA-LDA e dados do sistema de cores HSI.

Na abordagem de classificação em função dos vitivinicultores, os melhores resultados foram obtidos com os modelos construídos usando o algoritmo SPA-LDA e dados do sistema de cores HSI. O modelo SPA-LDA apresentou taxas de classificação corretas (TCC) nos conjuntos de treinamento e de teste, respectivamente, de 85,4% e 100%, com os dados RGB e de 89,9% e 100%, com os dados HSI. O modelo construído com PLS-DA apresentou TCC nos conjuntos de treinamento e de teste, respectivamente, de 77,0% e 95,5%, com os dados RGB e de 85,3% e 95,5%, com os dados HSI.

Na abordagem de classificação considerando-se as variedades de uvas presentes na composição dos vinhos varietais, comparadas aos pares, os modelos construídos com SPA-LDA e PLS-DA obtiveram resultados idênticos na discriminação entre os pares Cabernet Sauvignon *versus* Touriga Nacional e Syrah *versus* Touriga Nacional, com TCC de 100% nos conjuntos de treinamento e de teste. Na comparação Cabernet Sauvignon *versus* Syrah, divergências foram observadas nos resultados produzidos pelos dois algoritmos usados, sendo que os modelos construídos com SPA-LDA alcançaram iguais resultados com os dados RGB e HSI, com TCC de 85,7% e 72,7%, respectivamente, para os conjuntos de treinamento e de teste, enquanto os modelos obtidos com PLS-DA alcançaram TCC de modestos 57,1% no treinamento e 72,7% no teste, com os dados RGB e 47,6% no treinamento e 81,8% no teste, com os dados HSI.

Os resultados demonstram a viabilidade do uso imagens digitais associadas a ferramentas quimiométricas para classificação de vinhos em função de sua origem geográfica, vitivinicultor e composição varietal, de forma simples, rápida, com baixo consumo de amostras, sem usar qualquer pré-tratamento, reagentes químicos ou diluição das amostras e com baixa geração de resíduos.

REFERÊNCIAS

- ACEVEDO, C. A. et al. **A non-destructive digital imaging method to predict immobilized yeast-biomass.** Food Sci Technol, v. 42. p. 1444-1449, 2009.
- ALEJANDRE, M. A. L. **Manual de viticultura, enología y cata.** 2. ed. Córdoba, España: Almuzara, 2010. E-Book.
- ALMEIDA, V. E. et al. **Using color histograms and SPA-LDA to classify bacteria.** Analytical and Bioanalytical Chemistry (Print), v. 406. p. 5989-5995, 2014.
- AMARANTE, J. O. A. **Os segredos do vinho para iniciantes e iniciados.** 4. ed. São Paulo, SP: Mescla, 2015. E-Book.
- ANDRADE, R. H. S. et al. **Anthocyanic composition of Brazilian red wines and use of HPLC-UV-Vis associated to chemometrics to distinguish wines from different regions.** Microchemical Journal 110. p. 256-262, 2013.
- ARAÚJO, M. C. U. et al. **The Successive Projections Algorithm for Variable Selection in Spectroscopy Multicomponent Analysis.** Chemometrics and Intelligent Laboratory Models, v. 57, n. 2. p. 65-73, 2001.
- BALLABIO, D.; CONSONNI, V. **Classification tools in chemistry.** Part 1: Linear models. PLS-DA. Analytical Methods, 5. p. 3790-3798, 2013.
- BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, B. **Chemometrics – A Pratical Guide.** New York: Wiley, 1998.
- BENTLIN, F. R. S. et al. **Elemental analysis of wines from South America and their classification according to country.** Journal of the Brazilian Chemical Society, 22(2) . p. 327-336, 2011.
- BEVILACQUA, M. et al. **Classification and Class-Modelling.** In: MARINI, F. (Ed.). Chemometrics in Food Chemistry. Amsterdam, Netherlands: Elsevier Science, 2013. (Data Handling in Science and Technology, 28).
- BORAH, S.; BHUYAN, M. **A computer based model for matching colours during the monitoring of tea fermentation.** International Journal of Food Science & Technology, Vol.40(6). p. 675-682, 2005.
- BRASIL, **Lei n. 9.279, de 14 de maio de 1996.** Regula direitos e obrigações relativos à propriedade industrial. Disponível em <www.planalto.gov.br/ccivil_03/leis/19279.htm>. Acesso em: 23 jul. 2014.
- BRASIL. **Decreto Nº 8.198, de 20 de fevereiro de 2014.** Regulamenta a Lei no 7.678, de 8 de novembro de 1988, que dispõe sobre a produção, circulação e comercialização do vinho e derivados da uva e do vinho.
- BRASIL. **Instrução Normativa no. 1, de 2 de fevereiro de 2006.** Brasília, Ministério da Agricultura, Pecuária e Abastecimento. Disponível em:

<http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?jornal=1&pagina=19&data=03/02/2006>. Acesso em: 24 nov. 2016.

BRASIL. **Lei n. 7.678, de 8 de novembro de 1988**. Dispõe sobre a produção, circulação e comercialização do vinho e derivados da uva e do vinho, e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/1980-1988/L7678.htm. Acesso em: 31 mar. 2014.

BRERETON, R. G. **Chemometrics for Pattern Recognition**. New York: John Wiley & Sons, 2009.

BRERETON, R. G. **Chemometrics: Data Analysis for the Laboratory and Chemical Plant**. West Sussex/England: John Wiley & Sons, 2003.

BRERETON, R. G.; LLOYD, G. R. **Partial least squares discriminant analysis: taking the magic away**. J. Chemometrics, 28. p. 213-225, 2014.

CAMARGO, U. A. **Técnicas de produção vitícola com ciclos sucessivos em condições tropicais**. In: Workshop Internacional de Pesquisa: *A Produção de Vinhos em Regiões Tropicais*, 1. 2004, Petrolina/PE. **Anais...** Bento Gonçalves: EMBRAPA, 2009, p85-95. Disponível em: <http://www.cnpuv.embrapa.br/publica/anais/wip2004/41.pdf>. Acesso em: 31 mar. 2014.

CARUSO, M. et al. **Chemical Profile of White Wines Produced from ‘Greco bianco’ Grape Variety in Different Italian Areas by Nuclear Magnetic Resonance (NMR) and Conventional Physicochemical Analyses**. J. Agric. Food Chem. 60. p. 7-15, 2012.

COETZEE, P. P.; VAN JAARSVELD, F. P., VANHAECKE, F. **Intraregional Classification of Wine via ICP-MS elemental Fingerprinting**. Food Chemistry, 164. p. 485-492, 2014.

COMITÊ DA BACIA HIDROGRÁFICA DO RIO SÃO FRANCISCO. **Plano de recursos hídricos da bacia hidrográfica do rio São Francisco 2016-2025: RP2 – Diagnóstico Consolidado da Bacia Hidrográfica do Rio São Francisco**. V. 1 – Relatório de diagnóstico. 2015. Disponível em: <http://cbhsaofrancisco.org.br/planoderecursoshidricos/wp-content/uploads/2015/04/RP2_V1_Diagnostico_Consolidado.pdf>. Acesso em: 24 mar. 2017.

COSTA, G. B. et al. **Digital image-based classification of biodiesel**. Talanta (Oxford), v. 139. p. 50-55, 2015.

COSTA, G. B. et al. **Identification of biodiesel feedstock in biodiesel/diesel blends using digital images and chemometric methods**. Analytical Methods (Print), v. 8. p. 4949-4954, 2016.

CUADROS-INOSTROZA, A. et al. **Discrimination of wine attributes by metabolome analysis**. Analytical Chemistry 82. p. 3573-3580, 2010.

CVEJIĆ, J.; GOJKOVIC-BUKARICA, L. **Wine Phenolics: Clinical Trials**. In: SOTON, L. (Ed.). **Red wine consumption and health**. New York, USA: Nova Science Publishers, 2016. (Food and beverage consumption and health). E-Book.

DARDEAU, R. **Vinho Fino Brasileiro**. 1. ed. Rio de Janeiro: Mauad X. 2015. E-Book.

DINIZ, P. H. G. D. et al. **Using a simple digital camera and SPA-LDA modeling to screen teas.** *Analytical Methods*, 4. p. 2648-2652, 2012.

DOMÍNGUEZ, M. A. et al. **Geographical origin classification of Argentinean honeys using a digital image-based flow-batch model.** *Microchemical Journal (Print)*, v. 112. p. 104-108, 2014.

FABANI, M. P. F. et al. **Evaluation of elemental profile coupled to chemometrics to assess the geographical origin of Argentinean wines.** *Food Chemistry* 119. p. 372-379, 2010.

FERNANDES, D. D. S. **Espectroscopia UV_VIS para avaliação de biodiesel e misturas biodiesel/diesel.** 2013.75 f. Dissertação (Mestrado em química) - Centro de Ciências Humanas e Agrárias, Universidade Estadual da Paraíba, Campina Grande.

FRAIGE, K.; PEREIRA-FILHO, E. R.; CARRILHO, E. **Fingerprinting of anthocyanins from grapes produced in Brazil using HPLC–DAD–MS and exploratory analysis by principal component analysis.** *Food Chemistry*, 145. p. 395-403, 2014.

FRASER, B.; MURPHY, C.; BUNTING, F. **Real world color management.** 2nd Edition. Berkeley/USA: Peachpit Press, 2005.

FRASER, T.; BANKS, A. **O guia completo da cor.** Tradução Renata Bottini. 2. ed. São Paulo: Senac, 2013.

GAUTIER, J. **Vinho.** Tradução Julia da Rosa Simões. Porto Alegre, RS: L&PM, 2013. E-Book.

GEMPERLINE, P. **Principal Component Analysis.** In: ____ (Ed.). *Practical guide to chemometrics*. 2. ed. Boca Raton, U.S.A.: CRC Press, 2006.

GONZALES, R. C.; WOODS, R. E. **Processamento digital de imagens.** 3. ed. São Paulo: Pearson Prentice Hall, 2010.

GRASEL, F. S.; MARCELO, M. C. A.; FERRÃO, M. F. **A non-destructive, rapid and inexpensive methodology based on digital images for the classification of natural tannin extracts.** *RSC Advances*, Vol.6(38). p. 32358-32364, 2016.

GUERRA, C. C. **Vinho Tinto.** In: VENTURINI FILHO, W. G. (Org.). **Bebidas Alcoólicas: Ciência e Tecnologia**, vol. 1. São Paulo: Blucher, 2010.

GUERRA, C. C.; ZANUS, M. C. Características analíticas e sensoriais de vinhos produzidos no Vale Do Submédio São Francisco, Brasil. In: **Workshop Internacional de Pesquisa: A Produção de Vinhos em Regiões Tropicais**, 1. 2004, Petrolina/PE. *Anais...* Bento Gonçalves: EMBRAPA, 2009, p. 85-95. Disponível em: <http://www.cnpuv.embrapa.br/publica/anais/wip2004/41.pdf>. Acesso em: 31 mar. 2014.

Hobbies, Hobbits, and Hobos. **Applying Simple HSV Object Tracking.** 16 June 2011. Disponível em <<http://www.hobbitsandhobos.com/2011/06/applying-simple-hsv-object-tracking/>>. Acesso em: 14 ago. 2017.

International Organization of Wine and Vine (OIV). **2017 World Vitiviniculture Situation: OIV Statistical Report on World Vitiviniculture.** Disponível em: <<http://oiv.int/public/medias/5479/oiv-en-bilan-2017.pdf>>. Acesso em: 22 jun. 2017.

JONHSON, H.; ROBINSON, F. **Atlas Mundial do Vinho**. Tradução Luiz Horta; [et al.]. 7. Ed. São Paulo: Globo Estilo. 2014.

LAVINE, B. K.; DAVIDSON, C. E. **Classification and Pattern Recognition**. In: GEMPERLINE, P. (Ed.). *Practical guide to chemometrics*. 2. ed. Boca Raton, U.S.A.: CRC Press, 2006.

MANFRÓI, V. **Vinho Branco**. In: VENTURINI FILHO, W. G. (Org.). *Bebidas Alcoólicas: Ciência e Tecnologia*, vol. 1. São Paulo: Blucher, 2010.

MASSART, D. L. et al. (Eds.). **Chemometrics: a textbook**. 2nd. ed. Amsterdam, Netherlands: Elsevier Science, 2003. (Data Handling in Science and Technology, 2).

MEDINA, W.; SKURTYS, O.; AGUILERA, J. M. **Study on image analysis application for identification Quinoa seeds (*Chenopodium quinoa* Willd) geographical provenance**. LWT - Food Science and Technology 43(2). p. 238-246, 2010.

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

MILANEZ, K. D. T. M.; PONTES, M. J. C. **Classification of extra virgin olive oil and verification of adulteration using digital image and discriminant analysis**. Analytical Methods (Print), v. 7. p. 8839-8846, 2015.

MONAGAS, M.; BARTOLOMÉ, B. General Considerations of Anthocyanins in Wine. In: MORENO-ARRIBAS, M. Victoria; POLO, M. Carmen (eds). **Wine Chemistry and Biochemistry**. New York, USA. Ed. Springer, 2009.

MUCCILLO, L. et al. **Biochemical features of native red wines and genetic diversity of the corresponding grape varieties from Campania region**. Food Chemistry, 143. p. 506-513, 2014.

PEREIRA, G. E. Os vinhos tropicais em desenvolvimento no Nordeste do Brasil. **ComCiência: Revista eletrônica de jornalismo científico**. 2013. Disponível em: <<http://www.comciencia.br/comciencia/handler.php?section=8&edicao=89&id=1093>>. Acesso em: 22 fev. 2016.

PEREIRA, G. E.; VANDERLINDE, R.; LIMA, L. L. A. **Chemical and aromatic characteristics of Brazilian tropical wines**. Acta Horticulturae, nº. 910, p. 135-140, 2011.

PIERINI, G. D. et al. **A digital image-based traceability tool of the geographical origins of Argentine propolis**. Microchemical Journal (Print), v. 128. p. 62-67, 2016.

PONTES, M. J. C. **Algoritmo das projeções sucessivas para a seleção de variáveis espectrais em problemas de classificação**. 2009.123 f. Tese (Doutorado em Química) - Centro de Ciências Exatas e Naturais, Universidade Federal da Paraíba, João Pessoa.

PONTES, M. J. C. et al. **The Successive Projections Algorithm for Spectral Variable Selection in Classification Problems**. Chemometrics and Intelligent Laboratory Models, v. 78, n.1-2. p. 11-18, 2005.

PROTAS, J. F. S.; CAMARGO, U. A.; MELLO, L. M. R. Viticultura brasileira: regiões tradicionais e polos emergentes. **Informe Agropecuário**, Belo Horizonte, v. 27, n. 234, set/out. p. 7-15, 2006.

PUCKETTE, M.; JUSTIN, H. **O guia essencial do vinho**. Tradução Lucas Cordeiro e Renato Ferreira Pires. Rio de Janeiro, RJ: Editora Intrínseca Ltda, 2016. E-Book.

RAŽIĆ, S.; ONJIA, A. **Trace element analysis and pattern recognition techniques in classification of wine from central Balkan countries**. American Journal of Enology and Viticulture, 61(4). p. 506-511, 2010.

REGIÃO ADMINISTRATIVA INTEGRADA DE DESENVOLVIMENTO DO POLO PETROLINA E JUAZEIRO (Brasil). **BahiaPernambucoRIDEPoloPetrolinaeJuazeiro.svg**. Disponível em: <https://commons.wikimedia.org/w/index.php?curid=22380953>. Acesso em: 20 jan. 2017.

Regulation (EC). **Council Regulation (EC) No 817/70 of 28 April 1970 laying down special provisions relating to quality wines produced in specified regions**. Official Journal of the European Union, No. L99/20. p. 252-257, 1970.

RIBÉREAU-GAYON, P. et al. **Handbook of Enology**, Vol. 2. 2. ed. Chichester, England: John Wiley & Sons Ltd. 2006.

RODRIGUES, S. M. et al. **Elemental analysis for categorization of wines and authentication of their certified brand of origin**. Journal of Food Composition and Analysis, 24(4–5). p. 548-562, 2011.

ROSA, S. M. **Entre vinhas e siglas**. *Adega*, No. 52, Mar/2010. Disponível em: <http://revistaadega.uol.com.br/artigo/entre-vinhas-e-siglas_1551.html#ixzz4kerE13H5>. Acesso em: 21 jun. 2017.

SANTOS, P. M.; WENTZELL, P. D.; PEREIRA-FILHO, E. R. **Scanner Digital Images Combined with Color Parameters: A Case Study to Detect Adulterations in Liquid Cow's Milk**. Food Analytical Methods, Vol.5 (1). p. 89-95, 2012.

SILVA, A. C. **Metodologias analíticas para a identificação de não conformidades em amostras de álcool combustível**. 2013. 82 f. Dissertação (Mestrado em química) – Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, João Pessoa.

SOARES, S. F. C. et al. **A New Validation Criterion for Guiding the Selection of Variables by the Successive Projections Algorithm in Classification Problems**. Journal of the Brazilian Chemical Society (Impresso), v. 25. p. 176-181, 2014.

SOLOMON, C.; BRECKON, T. **Fundamentos de processamento digital de imagens: uma abordagem prática com exemplos em Matlab**. 1. ed. Rio de Janeiro: LTC. 2013.

SOUTO, U. T. C. P. et al. **Screening for Coffee Adulteration Using Digital Images and SPA-LDA**. Food Analytical Methods (Print), v. 8, p. 1515-1521, 2014.

TERRIER, N.; PONCET-LEGRAND, C.; CHEYNIER, V. **Flavanols, Flavonols and Dihydroflavonols**. In: MORENO-ARRIBAS, M. Victoria; POLO, M. Carmen (eds). Wine Chemistry and Biochemistry. New York, USA. Ed. Springer, 2009.

TONIETTO, J.; PEREIRA, G. E. **The development of the viticulture for a high quality tropical wine production in the world.** In: International Symposium of the Group of International Experts of Vitivinicultural Models for Cooperation, 17, 2011, Asti. Proceedings. Asti: Le Progrès Agricole et Viticole. p. 25-28, 2011.

TONIETTO, J. **O conceito de denominação de origem:** uma opção para o desenvolvimento do setor vitivinícola brasileiro. Bento Gonçalves: EMBRAPA-CNPUV, 1993.

TONIETTO, J.; TEIXEIRA, A. H. C. **O clima vitícola do Submédio São Francisco e o zoneamento dos períodos de produção de uvas para elaboração de vinhos.** In: Workshop Internacional de Pesquisa: *A Produção de Vinhos em Regiões Tropicais*, 1. 2004, Petrolina/PE. **Anais...** Bento Gonçalves: EMBRAPA, 2009, p. 41. Disponível em: <http://www.cnpuv.embrapa.br/publica/anais/wip2004/41.pdf>. Acesso em: 31 mar. 2014.

TONIETTO, J.; TEIXEIRA, A. H. C. **Zonage climatique des périodes viticoles de roduction dans l'année em zonage tropicale: application de la méthodologie du Système CCM Géoviticole.** In: JOINT INTERNATIONAL CONFERENCE ON VITICULTURAL ZONING, Cape Town, South Africa. p. 193-201, 2004.

TONIETTO, J; CARBONNEAU, A. **A multicriteria climatic classification model for grape-growing regions worldwide.** In: Agricultural and Forest Meteorology, v. 124, n. 1/2. p. 81-97, 2004.

TORRENCE, R.; WRIGHT, R.; CONWAY, R. **Identification of starch granules using image analysis and multivariate techniques.** Journal of Archaeological Science, Vol.31(5). p. 519-532, 2004.

UNIVERSIDADE FEDERAL DA PARAÍBA – UFPB. Carlos Monteiro de Lima; Paulo Cesar Xavier de Lara, David Douglas de Souza Fernandes e Mario Cesar Ugulino de Araújo. **Dispositivo para medições colorimétricas baseado em imagens digitais.** BR1020170114490. 31 maio 2017. Disponível em: <https://gru.inpi.gov.br/pePI/servlet/PatenteServletController?Action=detail&CodPedido=1419726&SearchParameter=BR1020170114490 &Resumo=&Titulo=>>. Acesso em: 20 jul. 2017.

UNIVERSIDADE FEDERAL DA PARAÍBA – UFPB. **Software para aquisição de dados de imagens** - QUIMAGENS. Disponível em: <http://www.laqa.quimica.ufpb.br/index.php/downloads>. Acesso em: 28 jan. 2017.

VERSARI, A. et al. **Progress in authentication, typification and traceability of grapes and wines by chemometric approaches (Review).** Food Research International, 60, p. 2-18, 2014.

ANEXO A – Informações indicadas nos rótulos das amostras de vinhos usadas na abordagem de classificação em função da origem geográfica.

Número	Informações indicadas nos rótulos das amostras						
	Classe	Nome	Variedades	Ano	Teor alcolólico (v/v, %)	Produtor	Origem
1	VSF	Bianchetti Orgânico	Ba	2012	12	Adega Bianchetti Tedesco	VSF
2	VSF	Bianchetti Orgânico	Ba	2012	12	Adega Bianchetti Tedesco	VSF
3	VSF	Bianchetti Orgânico	Ba	2015	12	Adega Bianchetti Tedesco	VSF
4	VSF	Bianchetti	CS	2011	11	Adega Bianchetti Tedesco	VSF
5	VSF	Bianchetti	Te	2014	12	Adega Bianchetti Tedesco	VSF
6	VSF	Bianchetti Orgânico	Te	2012	12	Adega Bianchetti Tedesco	VSF
7	VSF	Château Duccos	PV	2011	13	Château Duccos	VSF
8	VSF	Château Duccos	PV	2012	13	Château Duccos	VSF
9	VSF	Vinho Experimental	AB	2014	N. I.	EMBRAPA CPATSA	VSF
10	VSF	Vinho Experimental	AB	2015	N. I.	EMBRAPA CPATSA	VSF
11	VSF	Vinho Experimental	RC	2014	N. I.	EMBRAPA CPATSA	VSF
12	VSF	Vinho Experimental	Sy	2015	N. I.	EMBRAPA CPATSA	VSF
13	VSF	Vinho Experimental	Sy	2014	N. I.	EMBRAPA CPATSA	VSF
14	VSF	Vinho Experimental	Sy	2014	N. I.	EMBRAPA CPATSA	VSF
15	VSF	Vinho Experimental	TN	2014	N. I.	EMBRAPA CPATSA	VSF
16	VSF	Vinho Experimental	TN	2015	N. I.	EMBRAPA CPATSA	VSF
17	VSF	Vinho Experimental	TN	2015	N. I.	EMBRAPA CPATSA	VSF
18	VSF	Vinho Experimental	TN	2014	N. I.	EMBRAPA CPATSA	VSF
19	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO- PE	VSF

20	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO-PE	VSF
21	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO-PE	VSF
22	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO-PE	VSF
23	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO-PE	VSF
24	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO-PE	VSF
25	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO-PE	VSF
26	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO-PE	VSF
27	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO-PE	VSF
28	VSF	Vinho Experimental	TN	2014	N. I.	IF SERTÃO-PE	VSF
29	VSF	Almadén	Sy	2011	13	Miolo	VSF
30	VSF	Almadén	Sy	2013	13	Miolo	VSF
31	VSF	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
32	VSF	Rio Sol	CS	2012	13	Vinícola Santa Maria	VSF
33	VSF	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
34	VSF	Paralelo 8 Premium	CS/ Sy/ AB/ TN/Ar	2011	13,5	Vinícola Santa Maria	VSF
35	VSF	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
36	VSF	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
37	VSF	Botticelli Varietais	CS	2013	12	Vinícola do Vale do São Francisco	VSF
38	VSF	Botticelli Coleção	OS	2014	12	Vinícola do Vale do São Francisco	VSF
39	VSF	Botticelli Coleção	RC	2011	12	Vinícola do Vale do São Francisco	VSF

40	VSF	Botticelli Coleção	RC	2012	12	Vinícola do Vale do São Francisco	VSF
41	VSF	Botticelli Coleção	RC	2012	12	Vinícola do Vale do São Francisco	VSF
42	VSF	Botticelli Coleção	Ta	2012	12	Vinícola do Vale do São Francisco	VSF
43	VSF	Garziera	CS	2014	12	Vinícola Lagoa Grande	VSF
44	VSF	Garziera	PS	2014	12	Vinícola Lagoa Grande	VSF
45	VSF	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
46	VSF	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
47	VSF	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
48	VSF	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
49	VSF	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
50	VSF	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
51	VSF	Rio Sol Reserva	CS/ Sy/ AB	2011	13,5	Vinícola Santa Maria	VSF
52	VSF	Rio Sol Reserva	CS/ Sy / AB	2013	13,5	Vinícola Santa Maria	VSF
53	VSF	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
54	VSF	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
55	VSF	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
56	VSF	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
57	VSF	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
58	VSF	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
59	VSF	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
60	VSF	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
61	VSF	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
62	VSF	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
63	VSF	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF

64	VSF	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
65	VSF	Rio Sol Winemakers Selection	TN	2009	13	Vinícola Santa Maria	VSF
66	Mundo	Aurora Varietal	CS	2015	12	Companhia Vinícola Aurora	Serra Gaúcha/RS
67	Mundo	Dom Matias Reserva	CS	2014	14	Viña Cousiño Macul	Valle del Maipo/Chile
68	Mundo	Viñas Del Mar Stilus	CS	2014	13	Viña La Rosa	Valle do Cachapoal/ CI
69	Mundo	Santa Helena Reservado	CS	2013	13,5	Viña San Pedro Tarapacá	Valle Central – CI
70	Mundo	Root 1	CS	2013	13,5	Viña Ventisquero	Valle de Colchagua – CI
71	Mundo	Solar del Paso	CS	2015	12	Vinícola Aurora S. A.	Canelones/ Uruguai
72	Mundo	Classic Salton Reserva Especial	CS	2015	13	Vinícola Salton	Bento Gonçalves/RS
73	Mundo	Lunar	CS	N. I.	12	Vinícola Serra Gaúcha/RS	São Marcos/RS
74	Mundo	Altue Amucar	CS	2015	12,5	Vitivinícola Cremaschi Barriga	Chile
75	Mundo	Concha e Toro Reservado	CS	2015	12,5	Viña Concha e Toro	Valle Central – CI
76	Mundo	Dehesas Viejas – Roble	CS	2013	13,5	Pagos del Rey	Ribeira del Duero – Es
77	Mundo	Club des Someliers	CS	2014	12	Miolo	Vale dos Vinhedos/RS
78	Mundo	Don Matias Reserva	Sy	2014	14	Causiño-Macul	Vale do Maipo – CI
79	Mundo	Miolo Seleção	CS/ Me	2014	12,5	Miolo Wine Group	Campanha/RS
80	Mundo	Almadén	Me	2016	12	Miolo Wine Group	Campanha/RS
81	Mundo	Almadén	Ta	2015	12	Miolo Wine Group	Campanha/RS
82	Mundo	Almadén	CS	2015	12	Miolo	Campanha/RS
83	Mundo	Château Rougier	CS/ Me/ CF	2015	13	Les Vignobles Lobre	Bordeaux/Fr
84	Mundo	Toro Loco	Te	2015	12,5	BVC Bodegas	Utiel-Requena/Espanha
85	Mundo	Don Matias Reserva	CS	2014	14	Causiño-Macul	Vale do Maipo – CI

86	Mundo	Toro Loco	Te	2015	12,5	BVC Bodegas	Utiel-Requena/Espanha
87	Mundo	Canepa Novísimo	CS	2015	12,5	Viña Canepa	Valle Central – Cl
88	Mundo	Toro Loco	Te	2015	12,5	BVC Bodegas	Utiel-Requena/Espanha
89	Mundo	Toro Loco	Te	2015	12,5	BVC Bodegas	Utiel-Requena/Espanha
90	Mundo	Toro Loco	Te	2015	12,5	BVC Bodegas	Utiel-Requena/Espanha
91	Mundo	Canepa Novísimo	CS	2015	12,5	Viña Canepa	Valle Central – Cl
92	Mundo	Terralis	CS/ Me	2016	12	Premier	Mendoza/Argentina
93	Mundo	Terralis	CS/ Me	2016	12	Premier	Mendoza/Argentina
94	Mundo	Terralis	CS/ Me	2016	12	Premier	Mendoza/Argentina
95	Mundo	Terralis	CS/ Me	2016	12	Premier	Mendoza/Argentina
96	Mundo	Vinhateiros do Vale	Me	2014	12,5	Vinícola Aurora Real	Vale dos Vinhedos/RS
97	Mundo	Porca de Murça	TN/ TR/ TB	2013	13	Companhia Velha Bodegas	Douro/Portugal
98	Mundo	Trapiche Vineyards	Ma	2015	13	Trapiche	Mendoza/Argentina
99	Mundo	Fleur du Cup – Unfiltered	CS	2012	14	Die Bergkelder	Stellenbosch/África do Sul
100	Mundo	Manto Blanco	Ca	2015	13	Viña Luis Felipe Edwards	Valle Central – Cl

Fonte: produzida pelo autor.

Legendas: AB = Alicante Bouchet, Ar = Aragonez, Ba = Barbera, Ca = Carmenère, CF = Cabernet Franc, CS = Cabernet Sauvignon, Ma = Malbec, Me = Merlot, PS = Petit Syrah, PV = Petit Verdot, RB = Rubi Cabernet, Sy = Syrah, Ta = Tanat, TB = Tinta Barroca, Te = Tempanillo, TN = Touriga Nacional, TR = Tinta Roriz, N.I. = teor alcóolico não informado.

ANEXO B – Informações indicadas nos rótulos das amostras de vinhos usadas na abordagem de classificação em função dos vitivinicultores.

Número	Informações indicadas nos rótulos das amostras						
	Classe	Nome	Variedades	Ano	Teor alcolólico (v/v, %)	Produtor	Origem
1	1	Vinho Experimental	AB	2014	N.I.	EMBRAPA CPATSA	VSF
2	1	Vinho Experimental	AB	2015	N.I.	EMBRAPA CPATSA	VSF
3	1	Vinho Experimental	RC	2014	N.I.	EMBRAPA CPATSA	VSF
4	1	Vinho Experimental	Sy	2015	N.I.	EMBRAPA CPATSA	VSF
5	1	Vinho Experimental	Sy	2014	N.I.	EMBRAPA CPATSA	VSF
6	1	Vinho Experimental	Sy	2014	N.I.	EMBRAPA CPATSA	VSF
7	1	Vinho Experimental	Sy	2014	N.I.	EMBRAPA CPATSA	VSF
8	1	Vinho Experimental	Sy	2015	N.I.	EMBRAPA CPATSA	VSF
9	1	Vinho Experimental	TN	2015	N.I.	EMBRAPA CPATSA	VSF
10	1	Vinho Experimental	TN	2014	N.I.	EMBRAPA CPATSA	VSF
11	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
12	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
13	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
14	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
15	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
16	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
17	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
18	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF

19	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
20	1	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
21	2	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
22	2	Botticelli Varietais	CS	2013	12	Vinícola do Vale do São Francisco	VSF
23	2	Botticelli Coleção	PS	2014	12	Vinícola do Vale do São Francisco	VSF
24	2	Botticelli Coleção	RC	2011	12	Vinícola do Vale do São Francisco	VSF
25	2	Botticelli Coleção	RC	2012	12	Vinícola do Vale do São Francisco	VSF
26	2	Botticelli Coleção	RC	2012	12	Vinícola do Vale do São Francisco	VSF
27	2	Botticelli Coleção	PS	2014	12	Vinícola do Vale do São Francisco	VSF
28	2	Botticelli Coleção	PS	2014	12	Vinícola do Vale do São Francisco	VSF
29	2	Botticelli Coleção	PS	2014	12	Vinícola do Vale do São Francisco	VSF
30	2	Botticelli Coleção	PS	2014	12	Vinícola do Vale do São Francisco	VSF
31	2	Botticelli Coleção	PS	2014	12	Vinícola do Vale do São Francisco	VSF
32	2	Botticelli Coleção	RC	2012	12	Vinícola do Vale do São Francisco	VSF
33	2	Botticelli Coleção	RC	2012	12	Vinícola do Vale do São Francisco	VSF
34	2	Botticelli Coleção	RC	2012	12	Vinícola do Vale do São Francisco	VSF
35	2	Botticelli Coleção	RC	2012	12	Vinícola do Vale do São Francisco	VSF
36	2	Botticelli Coleção	RC	2012	12	Vinícola do Vale do São Francisco	VSF

37	2	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
38	2	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
39	2	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
40	2	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
41	2	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
42	2	Botticelli Varietais	RC	2015	12	Vinícola do Vale do São Francisco	VSF
43	2	Botticelli Varietais	RC	2015	12	Vinícola do Vale do São Francisco	VSF
44	2	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
45	3	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
46	3	Rio Sol	CS	2012	13	Vinícola Santa Maria	VSF
47	3	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
48	3	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
49	3	Rio Sol Winemakers Selection	AB	2013	13	Vinícola Santa Maria	VSF
50	3	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
51	3	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
52	3	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
53	3	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
54	3	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
55	3	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
56	3	Rio Sol Reserva	CS/ Sy/ AB	2011	13,5	Vinícola Santa Maria	VSF
57	3	Rio Sol Reserva	CS/ Sy/ AB	2013	13,5	Vinícola Santa Maria	VSF
58	3	Rio Sol Reserva	CS/ Sy/ AB	2013	13,5	Vinícola Santa Maria	VSF

59	3	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
60	3	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
61	3	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
62	3	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
63	3	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
64	3	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
65	3	Rio Sol	CS/ Sy	2014	13	Vinícola Santa Maria	VSF
66	3	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
67	3	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
68	3	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
69	3	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
70	3	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF

Legendas: AB = Alicante Bouchet, Ar = Aragonez, Ba = Barbera, Ca = Carmenére, CS = Cabernet Sauvignon, PS = Petit Syrah, PV = Petit Verdot, RB = Rubi Cabernet, Sy = Syrah, Te = Tempanillo, TN = Touriga Nacional, N.I. = teor alcóolico não informado.

**ANEXO C– Informações indicadas nos rótulos das amostras de vinhos usadas na
abordagem de classificação em função das variedades de uvas.**

Número	Informações indicadas nos rótulos das amostras						
	Classe	Nome	Variedades	Ano	Teor alcolólico (v/v, %)	Produtor	Origem
1	CS	Bianchetti	CS	2011	11	Adega Bianchetti Tedesco	VSF
2	CS	Château Duccos	CS	2011	12,3	Château Duccos	VSF
3	CS	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
4	CS	Rio Sol	CS	2012	13	Vinícola Santa Maria	VSF
5	CS	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
6	CS	Botticelli Varietais	CS	2013	12	Vinícola do Vale do São Francisco	VSF
7	CS	Garziera	CS	2014	12	Vinícola Lagoa Grande*	VSF
8	CS	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
9	CS	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
10	CS	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
11	CS	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
12	CS	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
13	CS	Rio Sol	CS	2014	13	Vinícola Santa Maria	VSF
14	CS	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
15	CS	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
16	CS	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
17	CS	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF
18	CS	Botticelli Varietais	CS	2015	12	Vinícola do Vale do São Francisco	VSF

19	Sy	Vinho Experimental	Sy	2015	N.I.	EMBRAPA CPATSA	VSF
20	Sy	Vinho Experimental	Sy	2014	N.I.	EMBRAPA CPATSA	VSF
21	Sy	Vinho Experimental EMBRAPA	Sy	2014	N.I.	EMBRAPA CPATSA	VSF
22	Sy	Almadén	Sy	2013	13	Miolo	VSF
23	Sy	Almadén	Sy	2011	13	Miolo	VSF
24	Sy	Almadén	Sy	2015	13	Miolo	VSF
25	Sy	Almadén	Sy	2013	13	Miolo	VSF
26	Sy	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
27	Sy	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
28	Sy	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
29	Sy	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
30	Sy	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
31	Sy	Rio Sol	Sy	2014	13	Vinícola Santa Maria	VSF
32	Sy	Rio Sol	Sy	2014	N.I.	Vinícola Santa Maria	VSF
33	TN	Vinho Experimental	TN	2014	N.I.	EMBRAPA CPATSA	VSF
34	TN	Vinho Experimental	TN	2015	N.I.	EMBRAPA CPATSA	VSF
35	TN	Vinho Experimental	TN	2015	N.I.	EMBRAPA CPATSA	VSF
36	TN	Vinho Experimental	TN	2014	N.I.	EMBRAPA CPATSA	VSF
37	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
38	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
39	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
40	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
41	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
42	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF

43	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
44	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
45	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
46	TN	Vinho Experimental	TN	2014	N.I.	IF SERTÃO-PE	VSF
47	TN	Rio Sol Winemakers Selection	TN	2009	13	Vinícola Santa Maria	VSF
48	TN	Rio Sol Winemakers Selection	TN	2010	13	Vinícola Santa Maria	VSF

Legendas: CS = Cabernet Sauvignon, Sy = Syrah e TN = Touriga Nacional. N.I. = teor alcóolico não informado.

**ANEXO D – Pedido nacional de patente de invenção do Dispositivo de Medições
Colorimétricas baseado em imagens digitais**



31/05/2017 870170036437
02:45



**Pedido nacional de Invenção, Modelo de Utilidade, Certificado de
Adição de Invenção e entrada na fase nacional do PCT**

Número do Processo: BR 10 2017 011449 0

Dados do Depositante (71)

Depositante 1 de 2

Nome ou Razão Social: Universidade Federal da Paraíba

Tipo de Pessoa: Pessoa Jurídica

CPF/CNPJ: 24098477000110

Nacionalidade: Brasileira

Qualificação Jurídica: Instituição de Ensino e Pesquisa

Endereço: Cidade Universitária

Cidade: João Pessoa

Estado: PB

CEP: 58059-900

País: Brasil

Telefone: (83) 32167558

Fax:

Email: inova@reitoria.ufpb.br

Nome ou Razão Social: INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DO RIO GRANDE DO NORTE

Tipo de Pessoa: Pessoa Jurídica

CPF/CNPJ: 10877412000168

Nacionalidade: Brasileira

Qualificação Jurídica: Instituição de Ensino e Pesquisa

Endereço: Rua Doutor Nilo Bezerra Ramalho, 1692, Tirol

Cidade: Natal

Estado: RN

CEP: 59015-300

País: BRASIL

Telefone: (84) 996 021664

Fax:

Email: carlos.monteiro@ifrn.edu.br

Dados do Pedido

Natureza Patente: 10 - Patente de Invenção (PI)

Título da Invenção ou Modelo de DISPOSITIVO PARA MEDIÇÕES COLORIMÉTRICAS BASEADO

Utilidade (54): EM IMAGENS DIGITAIS

Resumo: A presente invenção trata de um dispositivo para medições colorimétricas baseado em imagens digitais, com aplicação na área de instrumentos analíticos, que se acopla um aparelho para aquisição de imagens digitais, com características de um estúdio fotográfico, hermeticamente fechado à entrada de luz e tem por função básica capturar e processar imagens digitais para realizar classificação de produtos nas áreas agrícola, industrial, farmacêutica, química, biológica, entre outras.

O dito dispositivo é constituído por moldura(1) que se acopla a um gabinete(3) que compreende um batente vazado(5), uma divisória(6), um ambiente de registro fotográfico(7), um ambiente de instalações elétricas(8), um sistema de iluminação interna(12)(21)(22)(23)(24) e uma abertura(9) com um anteparo batedor(10) e uma canaleta(11) para encaixar gavetas(4a)(4b)(4c) contendo suportes e recipientes(14a)(14b)(14c) para análise de amostras.