



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



TESE DE DOUTORADO

Novos métodos para autenticação e determinação de umidade e dos teores totais de proteínas e fósforo em ovos de galinha em pó usando imagens digitais, espectrometria NIR e quimiometria

Luciano Bernardo Ramo

João Pessoa – PB – Brasil

Agosto/2024



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

Novos métodos para autenticação e determinação de umidade e dos teores totais de proteínas e fósforo em ovos de galinha em pó usando imagens digitais, espectrometria NIR e quimiometria

Luciano Bernardo Ramo*

Tese de Doutorado apresentada ao Programa de Pós-graduação em Química da Universidade Federal da Paraíba como parte dos requisitos para obtenção do título de Doutor em Ciências, área de concentração: Química Analítica.

Orientador: Prof. Dr. Mário César Ugulino de Araújo

***Bolsista CAPES**



João Pessoa – PB – Brasil

Agosto/2024

Catálogo na publicação
Seção de Catalogação e Classificação

R175n Ramo, Luciano Bernardo.

Novos métodos para autenticação e determinação de umidade e dos teores totais de proteínas e fósforo em ovos de galinha em pó usando imagens digitais, espectrometria NIR e quimiometria / Luciano Bernardo Ramo. - João Pessoa, 2024.

120 f. : il.

Orientação: Mário César Ugulino de Araújo.
Tese (Doutorado) - UFPB/CCEN.

1. Espectrometria NIR. 2. Imagens digitais. 3. Calibração multivariada. 4. Classificação multivariada. 5. Ovos de galinha em pó. 6. Análise de alimentos. I. Araújo, Mário César Ugulino de. II. Título.

UFPB/BC

CDU 641.51(043)

Novos métodos para autenticação e determinação de umidade e dos teores totais de proteínas e fósforo em ovos de galinha em pó usando imagens digitais, espectrometria NIR e quimiometria.

Trabalho de Tese de Doutorado apresentada pelo discente **LUCIANO BERNARDO RAMO** e aprovada pela Comissão Examinadora composta abaixo, realizada no dia 28 de agosto de 2024.

Documento assinado digitalmente
 **MARIO CESAR UGULINO DE ARAUJO**
Data: 28/08/2024 16:24:24-0300
Verifique em <https://validar.iti.gov.br>

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

Documento assinado digitalmente
 **KASSIO MICHELL GOMES DE LIMA**
Data: 28/08/2024 17:10:53-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Kassio Michell Gomes de Lima
IQ/UFRN
Examinador externo

Documento assinado digitalmente
 **SOFACLES FIGUEREDO CARREIRO SOARES**
Data: 28/08/2024 20:09:35-0300
Verifique em <https://validar.iti.gov.br>

Dr. Sófacles Figueredo Carreiro Soares
CT/UFPB
Examinador externo

Documento assinado digitalmente
 **JOSE GERMANO VERAS NETO**
Data: 28/08/2024 22:09:07-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. José Germano Veras Neto
DQ/UEPB
Examinador interno

Documento assinado digitalmente
 **MARCIO JOSE COELHO DE PONTES**
Data: 29/08/2024 14:42:08-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Márcio José Coelho de Pontes
DQ/UFPB
Examinador interno

Dedico este trabalho primeiramente a Deus por ter me dado condições para concluir mais uma etapa importante em minha vida. Dedico também, de maneira especial, aos meus familiares, especialmente a minha esposa, cujo amor, apoio e compreensão foram a força necessária para cada passo durante minha trajetória acadêmica.

AGRADECIMENTOS

- ❖ A Deus, por sempre estar comigo, ter me abençoando em todos os momentos e me dado força e sabedoria para concluir mais uma etapa em minha vida.
- ❖ A minha esposa, Andréia Melo, cuja paciência, compreensão e apoio foram essenciais para minha perseverança na realização deste trabalho de tese.
- ❖ Aos meus pais, Luzia Bernardo e Severino do Ramo, minhas irmãs Lucélia Bernardo e Lucilene Bernardo e meus avós, em especial a vó Maria, pelo amor incondicional, encorajamento e apoio ao longo de toda a minha jornada acadêmica.
- ❖ Ao meu orientador, professor Mário Ugolino, pela orientação, dedicação, atenção, ensinamentos e paciência no desenvolvimento deste trabalho.
- ❖ A todo grupo LAQA, especialmente Amanda, Rossana, Fernanda, Tássio, Wallis, Diana, Andréia, João, David, Paulo Diniz, Edilene e Wellington pela amizade, convivência, discussões, colaboração e apoio mútuo ao longo desses anos.
- ❖ À banca examinadora, professor Márcio Coelho, Germano Vêras, Sófacles Soares e Kassio Michell pela disponibilidade em analisar este trabalho e pelas valiosas sugestões que contribuíram para a sua melhoria.
- ❖ Aos professores do PPGQ/UFPB por contribuir com a minha formação acadêmica.
- ❖ À professora Marta Madruga do LAQA/CT, pela cessão do equipamento *Kjeldahl* para realização dos experimentos e a técnica Leila, pela compreensão e suporte no método de determinação de proteínas.
- ❖ À professora Tatiane e a técnica Isabel do DBM/CCEN pela ajuda com a liofilização das amostras.
- ❖ A Capes pelas bolsas concedidas.

“Porque sou eu que conheço os planos que tenho para vocês, diz o Senhor, planos de fazê-los prosperar e não de causar dano, planos de dar a vocês esperança e um futuro.”

Jeremias 29:11

RESUMO

Os ovos de galinha em pó são empregados em várias formulações de alimentos nas indústrias de panificação e confeitaria, principalmente devido à sua longa vida útil, facilidade de armazenamento, transporte, além de possuir propriedades gelificantes, espumantes e emulsificantes. A avaliação de sua qualidade é importante para garantir a segurança alimentar dos consumidores. Sendo assim, este trabalho objetivou desenvolver métodos rápidos, não trabalhosos e acessíveis para avaliar a qualidade de ovos de galinha em pó a partir da utilização de espectros infravermelho próximo (NIR, do inglês *near-infrared*), imagens digitais, fusão de dados e técnicas de calibração e autenticação. As imagens digitais foram capturadas utilizando um *scanner*, enquanto os espectros NIR foram registrados usando espectrômetros de bancada e portátil. No primeiro estudo, novas estratégias foram desenvolvidas para determinar umidade e teores de proteínas e fósforo totais em ovos de galinha em pó. Os métodos de calibração multivariada para prever o teor de proteínas totais, em termos do coeficiente de correlação (r) e razão do desvio de desempenho da predição (RPD, do inglês *ratio performance deviation*), foram melhores utilizando imagens digitais ($r = 0,995$ e $RPD = 9,55$) do que aqueles utilizando os espectros NIR provenientes do equipamento de bancada e portátil. Por outro lado, na predição de umidade ($r = 0,974$ e $RPD = 4,35$) e teor de fósforo total ($r = 0,98$ e $RPD = 4,3$), o NIR de bancada obteve uma melhor performance analítica. Na aplicação das técnicas de fusão de dados, os resultados obtidos foram ligeiramente inferiores aos melhores métodos desenvolvidos utilizando espectros NIR pré-processados ou imagens digitais separadamente. No segundo estudo, novas estratégias foram concebidas para autenticar ovos de galinha em pó quanto a sua adulteração com proteína de soja e amido de milho. Para isso, aplicou-se técnicas de classificação utilizando a Modelagem Independente e Flexível por Analogia de Classes direcionada pelos dados (DD-SIMCA, do inglês *Data-Driven – Soft Independent Modelling of Class Analogy*). O melhor método foi aquele utilizando os espectros NIR portátil quando comparado com as imagens digitais, o qual todas as amostras adulteradas foram classificadas corretamente e apenas 1 amostra pura foi classificada como adulterada no conjunto de teste, alcançando sensibilidade de 95,7%, especificidade de 100% e acurácia 99,4%. Portanto, métodos propostos neste trabalho de tese mostraram-se uma boa ferramenta alternativa para análises químicas destrutivas, visto que amostras de ovos de galinha em pó poderiam ser analisadas sem utilização de produtos químicos e sem gerar resíduos prejudiciais à saúde e ao meio ambiente, seguindo os princípios básicos da Química Verde.

Palavras-chave: Espectrometria NIR; Imagens digitais; Calibração multivariada; Classificação multivariada; Ovos de galinha em pó; Análise de alimentos.

ABSTRACT

Powdered chicken eggs are used in various food formulations in the bakery and confectionery industries, mainly due to their long shelf life, ease of storage and transportation, and their gelling, foaming and emulsifying properties. Assessing their quality is important to ensure food safety for consumers. Therefore, this work aimed to develop fast, labor-free and accessible methods to assess the quality of powdered chicken eggs using near-infrared (NIR) spectra, digital images, data fusion and calibration and authentication techniques. Digital images were captured using a scanner, while NIR spectra were recorded using benchtop and portable spectrometers. In the first study, new strategies were developed to determine moisture and total protein and phosphorus contents in powdered chicken eggs. Multivariate calibration methods to predict total protein content, in terms of correlation coefficient (r) and ratio performance deviation (RPD), were better using digital images ($r = 0.995$ and $RPD = 9.55$) than those using NIR spectra from bench and portable equipment. On the other hand, in predicting humidity ($r = 0.974$ and $RPD = 4.35$) and total phosphorus content ($r = 0.98$ and $RPD = 4.3$), the benchtop NIR achieved better analytical performance. When applying data fusion techniques, the results obtained were slightly inferior to the best methods developed using pre-processed NIR spectra or digital images separately. In the second study, new strategies were devised to authenticate powdered chicken eggs for adulteration with soy protein and corn starch. To this end, classification techniques were applied using data-driven Independent and Flexible Modeling of Class Analogy (DD-SIMCA). The best method was the one using portable NIR spectra when compared with digital images, in which all adulterated samples were classified correctly and only one pure sample was classified as adulterated in the test set, achieving sensitivity of 95.7%, specificity of 100%, and accuracy 99.4%. Therefore, the methods proposed in this thesis work proved to be a good alternative tool for destructive chemical analysis since samples of powdered chicken eggs could be analyzed without using chemicals and without generating waste harmful to health and the environment, following the basic principles of Green Chemistry.

Keywords: NIR spectrometry; Digital images; Multivariate calibration; Multivariate classification; Powdered chicken eggs; Food analysis.

LISTA DE FIGURAS

Figura 2.1 - Esquema do Espectrofotômetro NIR convencional e portátil.....	26
Figura 2.2 - Modelo de cor RGB.	28
Figura 2.3 - Sistema de cor HSV.....	29
Figura 2.4 - Esquema da aquisição das imagens digitais e extração do histograma, os quais guardam a informação analítica.	30
Figura 2.5 - Diagrama esquemático da aquisição dos dados utilizando NIR e imagens digitais.	34
Figura 2.6 - Resumo gráfico dos três tipos de fusão de dados: nível baixo, nível médio e nível alto..	47
Figura 2.7 - Estrutura anatômica do ovo.....	49
Figura 2.8 - Produção dos fosfolipídios, colesterol e pigmentos carotenoides via metabólica não linear com anabolismo divergente.....	53
Figura 2.9 - Estrutura química (a) das proteínas, (b) lecitina e (c) colesterol.....	54
Figura 2.10 - Esquema de um liofilizador: (1) Câmara de secagem; (2) Condensador; (3) Sistema de vácuo; (4) tampa da câmara de secagem; (5) prateleira; e (6) Serpentina de resfriamento.....	57
Figura 3.1 – Diagrama esquemático do processo de tratamento da amostra, obtenção das imagens, espectros NIR e ferramentas quimiométricas utilizadas.	60
Figura 3.2 - Espectros médio NIR originais (a, b) e cortados (c, d) para as 132 amostras de ovos de galinha em pó, obtidos com os instrumentos de bancada e portátil, respectivamente e (e) espectros médios do NIR de bancada (–) e portátil (–) com <i>offset</i> que apresenta a localização da banda de P-OH presentes nas amostras de ovos de galinha em pó analisada nesta tese.....	68
Figura 3.3 – Espectro NIR médio pré-processados com SGD e primeira derivada, polinômio de segunda ordem e janela de (a) 17 (b) 19 (c) 21 pontos; (d) LBC; (e) MSC; (f) OBC + LBC; (g) OBC; e (h) SNV para o equipamento de bancada.	70
Figura 3.4 – Espectro NIR médio pré-processados com SGD com primeira derivada, polinômio de segunda ordem e janela de (a) 5 (b) 7 (c) 9 pontos; (d) LBC; (e) MSC; (f) OBC + LBC; (g) OBC; e (h) SNV para o equipamento portátil.	71
Figura 3.5 - Gráfico de Resíduo Studentizado versus Influência com o pré-processamento (a) OBC + LBC para umidade, SNV para (b) teor de proteínas e (c)	

fósforo totais do equipamento de bancada, derivada Savitzky-Golay com segunda derivada e janela de 17 pontos para (d) umidade e (e) teor de proteína total.72

Figura 3.6 - Gráfico de RMSECV versus variáveis incluídas no modelo para (a) umidade, teor de (b) proteínas e (c) fósforo totais utilizando o equipamento NIR de bancada, (d) umidade e (e) teor de proteínas totais usando o equipamento NIR portátil.73

Figura 3.7 – Variáveis espectrais selecionadas pelo modelo SPA-MLR nos espectros NIR de bancada pré-processados com LBC + OBC (a). Curva de valor predito versus valor de referência (d) e curva da região de confiança elíptica conjunta (EJCR) para determinação da umidade (g). Intervalos de variáveis espectrais selecionadas pelo modelo iSPA-PLS em espectros NIR de bancada pré-processados com SNV (b), curva de valor previsto versus valor de referência (e) e curvas EJCR para determinação do teor de fósforo total (h). Intervalos de variáveis espectrais selecionadas pelo modelo iSPA-PLS em espectros NIR portáteis pré-processados com SGD (c), curva de valor predito versus valor de referência (f) e curva EJCR para determinação do conteúdo de proteína total (i).76

Figura 3.8 - Histogramas de cor (a) originais e (b) sem as variáveis não informativas.77

Figura 3.9 - Gráfico de Resíduo Studentizado *versus* Influência para o teor de (a) proteínas e (b) fósforo totais com os modelos de cor RGB+HSV e RGB, respectivamente.78

Figura 3.10 - Gráfico de RMSECV versus variáveis incluídas no modelo para o teor de (a) proteínas e (b) fósforo totais com os modelos de cor RGB + HSV e RGB, respectivamente.79

Figura 3.11 - Intervalos de variáveis de cor selecionadas pelo modelo iSPA-PLS nos histogramas de cor dos canais RGB + HSV (a), curva do valor predito versus valor de referência (c) e curva elíptica de confiança conjunta (EJCR) (e) para determinação do teor de proteínas totais. Variáveis de cor selecionadas pelo modelo SPA-MLR nos histogramas de cores dos canais RGB (b), curva do valor predito versus valor de referência (d) e curva EJCR (f) para o teor de fósforo total.....81

Figura 3.12 - Imagens digitais e histogramas de cor das amostras de ovos de galinha em pó com quatro diferentes teores de proteínas totais (a e c) e fósforo (b e d). O círculo tracejado nas imagens digitais é a região circular de interesse selecionada e

usada pelo algoritmo IMAGENS_GUI para converter cada imagem em histogramas com os modelos de cores: GS, RGB e HSV.....82

Figura 4.1 – Espectro médio completo do NIR portátil NIR (a) e (c) cortado e histograma de cor médio completo (b) e após a remoção dos zeros (d).90

Figura 4.2 – Espectro médio do NIR portátil NIR pré-processado com (a) DSG-1^oD-2^oO com janelas de 3, 5 e 7, (b) LBC; (c) OBC; (d) OBC + LBC; (e) MSC e (f) SNV.91

Figura 4.3 – Gráfico dos escores de PC1 versus PC2 para as amostras de ovo em pó puras (●) e adulteradas (●), utilizando os espectros NIR portátil pré-processados com (a) DSG-1^oD-2^oO com janelas de 3, 5 e 7, LBC (b); OBC (c); OBC + LBC (d); MSC (e) e SNV (f).92

Figura 4.4 – Gráfico dos escores de PC1 versus PC2 para as amostras de ovo em pó puras (●) e adulteradas (●), utilizando os histogramas de cor com os modelos (a) GS, (b) RGB, (c) HSV, (d) GS+RGB, (e) GS+HSV, RGB+HSV e (f) GS+RGB+HSV.93

Figura 4.5 – Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) para o algoritmo DD-SIMCA desenvolvidos utilizando os espectros do NIR portátil (a, b) e os histogramas de cor (c, d), das amostras de ovo em pó puras (●) e adulteradas (●). A linha verde (—) representam a borda entre as amostras regulares (puras (●)) e extremas (●); a linha vermelha (—) é a fronteira da área de aceitação. 97

LISTA DE QUADROS E TABELAS

Quadro 2.1 - Proteínas da clara e suas características.....	50
Quadro 2.2 - Proteínas da gema e suas características.....	52
Quadro 2.3 - Classificação do ovo quanto ao seu peso.	55
Tabela 3.1 - Faixa de calibração, média e desvio padrão obtidos na determinação de umidade e teor de proteína e fósforo totais em amostras de ovos de galinha em pó utilizando métodos de referência.	65
Tabela 3.2 - Desempenho da calibração e predição dos modelos construídos e validados para determinação de umidade e teor de proteína e fósfor totais em amostras de ovos de galinha em pó usando espectros NIR de bancada e algoritmos de calibração multivariada.	74
Tabela 3.3 - Desempenho de calibração e predição dos modelos construídos e validados para determinação do teor de proteínas e fósforo totais em amostras de ovos de galinha em pó usando imagens digitais e algoritmos de calibração multivariada.....	80
Tabela 3.4 - Desempenho de calibração e predição dos modelos construídos e validados para determinação de umidade e teor de proteínas e fósforo totais em amostras de ovos de galinha em pó usando fusão de dados de espectros NIR e imagens digitais e calibração multivariada.	84
Tabela 3.5 - Características analíticas dos métodos quimiométricos propostos e os demais, utilizados para determinação de umidade e teores de proteínas e fósforo totais em amostras de ovos.	86

LISTA DE ABREVIATURAS E SIGLAS

ABPA	Associação Brasileira de Proteína Animal
AOAC	Associação Oficial dos Químicos Analíticos, do inglês <i>Association of Official Analytical Chemists</i>
CIE	Comissão Internacional de Iluminação, do inglês <i>Commission Internationale de l'Éclairage</i>
CMYK	Ciano-Magenta-Amarelo-Preto, do inglês <i>Cyan-Magenta-Yellow-Black</i>
DD-SIMCA	Modelagem Independente e Flexível por Analogia de Classes direcionada pelos dados, do inglês <i>Data-Driven Soft Independent Modeling of Class Analogy</i>
DMD	Dispositivo de Microespelhos Digitais, do inglês <i>Digital Micromirror Device</i>
EJCR	Região Elíptica de Confiança Conjunta, do inglês <i>Elliptical Joint Confidence Region</i>
FAO	Organização das Nações Unidas para Alimentação e Agricultura, do inglês <i>Food and Agriculture Organization</i>
FIR	Infravermelho Distante, do inglês <i>Far-Infrared</i>
FT	Transformada Fourier, do inglês <i>Fourier Transform</i>
GS	Escala de Cinzas, do inglês <i>Gray Scale</i>
HCA	Análise por Agrupamento Hierárquico, do inglês: <i>Hierarchical Cluster Analysis</i>
HDL	Lipoproteína de Alta Densidade, do inglês <i>High Density Lipoprotein</i>
HSL	Matiz-Saturação-Luminância, do inglês <i>Hue-Saturation-Luminance</i>
HSV	Matiz-Saturação-Valor, do inglês <i>Hue-Saturation-Value</i>
ID	Imagens Digitais
\bar{i} PLS	Regressão de Mínimos Quadrados Parciais por Intervalos, do inglês <i>Interval Partial Least Squares Regression</i>
\bar{i} SPA-PLS	Algoritmo das Projeções Sucessivas para seleção de Intervalos em Regressão de Mínimos Quadrados Parciais, do inglês <i>Successive Projection Algorithm for Interval selection in Partial Least Squares</i>
KS	Kennard-Stone
LBC	Correção de Linha de Base Linear, do inglês <i>Linear Baseline Correction</i>
LDA	Análise Discriminante Linear, do inglês: <i>Linear Discriminant Analysis</i>
LDL	Lipoproteína de Baixa Densidade, do inglês <i>Low Density Lipoprotein</i>
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MIR	Infravermelho Médio, do inglês <i>Mid-Infrared</i>
MLR	Regressão Linear Múltipla, do inglês <i>Multiple Linear Regression</i>
MSC	Correção Multiplicativa de Sinal, do inglês <i>Multiplicative Scatter Correction</i>
NIPALS	Mínimos Quadrados Parciais Iterativos Não Lineares, do inglês <i>Non Linear Iterative Partial Least Squares</i>

NIR	Infravermelho Próximo, do inglês <i>Near Infrared Spectroscopy</i>
OBC	Correção de Linha de Base <i>Offset</i> , do inglês <i>Offset Baseline Correction</i>
OLS	Mínimos Quadrados Ordinários, do inglês <i>Ordinary Least Squares</i>
PCA	Análise de Componente Principal, do inglês: <i>Principal Component Analysis</i>
PCs	Componentes Principais, do inglês <i>Principal Components</i>
PLS	Regressão por Mínimos Quadrados Parciais, <i>Partial Least Squares regression</i>
PLS-DA	Análise Discriminante pelos Mínimos Quadrados Parciais, do inglês: <i>Partial Least Squares-Discriminant Analysis</i>
PRESS	Soma Quadrática dos Erros de Previsão, do inglês <i>Prediction Error Sum of Squares</i>
r	Coefficiente de correlação
R ²	Coefficiente de Determinação
REP	Erro Relativo de Predição, do inglês <i>Relative Error of Prediction</i>
RGB	Vermelho-Verde-Azul, do inglês <i>Red-Green-Blue</i>
RMSECV	Raiz Quadrada do Erro Médio Quadrático da Validação Cruzada, do inglês <i>Root Mean Squared Error of Cross-Validation</i>
RMSEP	Raiz Quadrada do Erro Médio Quadrático da Predição, do inglês <i>Root Mean Square Error of Prediction</i>
ROI	Região de Interesse, do inglês <i>Region of Interest</i>
RPD	Desvios Residuais da Predição, do inglês <i>Residual Prediction Deviation</i>
SDV	Desvio Padrão dos Erros Validação, do inglês <i>Standard Deviation of Validation</i>
SGD	Derivação <i>Savitzky-Golay</i>
SNV	Varição Normal Padrão, do inglês <i>Standard Normal Variation</i>
SPA	Algoritmo das Projeções Sucessivas, do inglês <i>Successive Projection Algorithm</i>
SPXY	Partição do conjunto de amostras baseado nas distâncias conjuntas XY, do inglês <i>Sample set Partitioning based on joint XY distances</i>
SVD	Decomposição de Valor Singular, do inglês <i>Singular Value Decomposition</i>

SUMÁRIO

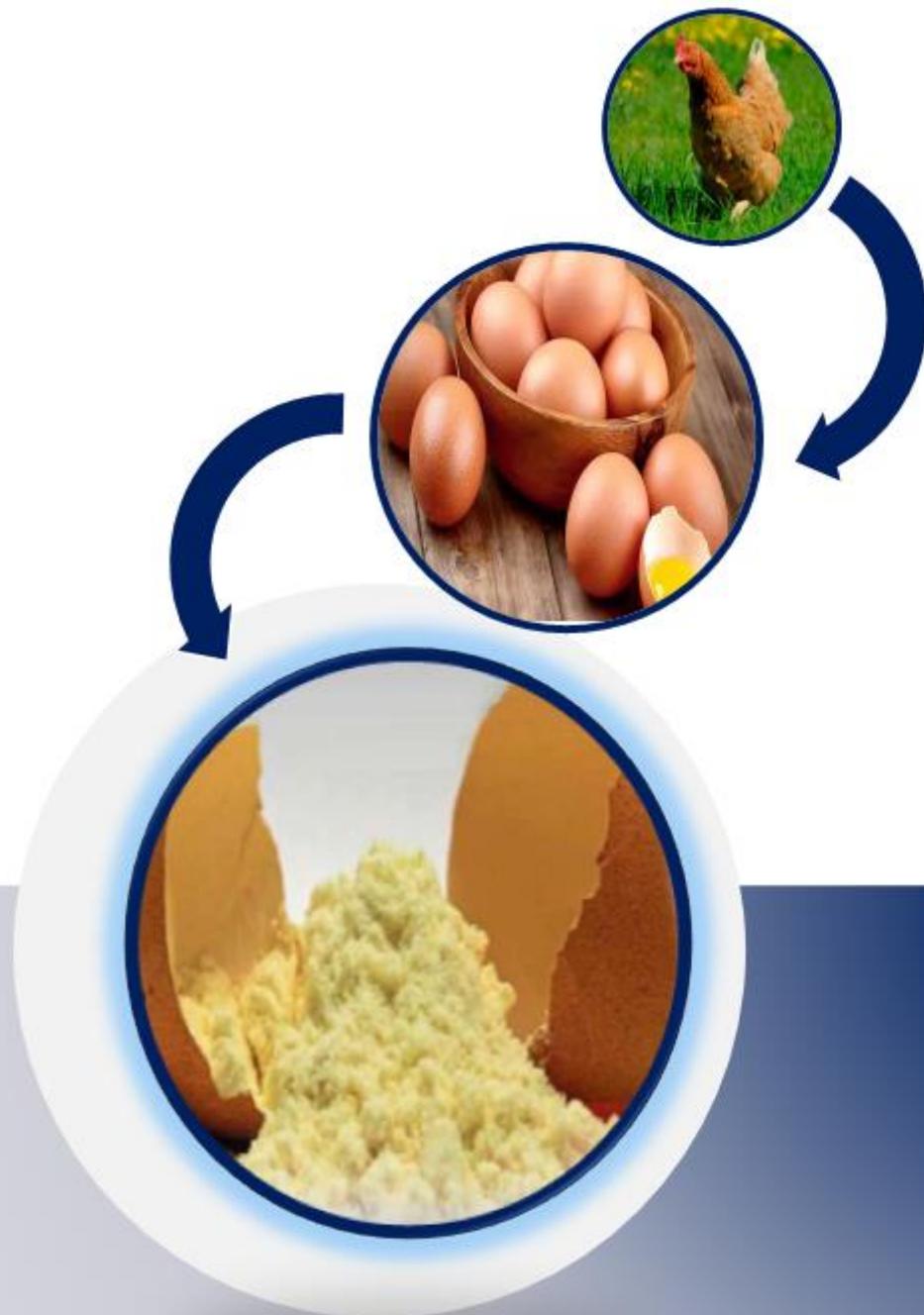
1	INTRODUÇÃO	18
1.1	Caracterização do problema.....	18
1.2	Objetivos	21
1.2.1	Objetivo geral	21
1.2.2	Objetivos específicos	21
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	Espectroscopia no Infravermelho Próximo (NIR)	23
2.1.1	Fundamentação	23
2.1.2	Instrumentação.....	25
2.2	Imagens Digitais.....	27
2.2.1	Fundamentação	27
2.2.2	Instrumentação.....	30
2.3	Quimiometria	31
2.3.1	Técnicas de pré-processamento.....	31
2.3.2	Calibração multivariada.....	33
2.3.3	Reconhecimento de padrões	43
2.3.3.1	Análise por Componentes Principais (PCA)	43
2.3.3.2	Modelagem independente e flexível por analogia de classes direcionada pelos dados (DD-SIMCA)	44
2.3.4	Fusão de dados.....	46
2.4	Ovos de galinha em pó	48
2.4.1	Considerações gerais	48
2.4.2	Composição anatômica do ovo.....	49
2.4.3	Classificação dos ovos.....	54
2.4.4	Ovoprodutos.....	55
2.4.4.1	Ovo desidratado.....	56
3.1	DETERMINAÇÃO DE UMIDADE E TEORES DE PROTEÍNAS E FÓSFORO TOTAIS EM AMOSTRAS DE OVOS DE GALINHA EM PÓ	60
3.1	Experimental	60
3.1.1	Amostras e preparo.....	60
3.1.2	Reagentes e soluções.....	61

3.1.3	Análise das amostras de ovos de galinha em pó pelos métodos de referência	62
3.1.3.1	Umidade	62
3.1.3.2	Teor de proteínas totais	63
3.1.3.3	Teor de fósforo total	64
3.1.3.4	Faixa de calibração, média e desvio padrão	65
3.1.4	Aquisição das imagens digitais	66
3.1.5	Aquisição dos espectros NIR	66
3.1.6	Procedimento quimiométrico.....	67
3.2	Resultados e discussão	67
3.2.1	Calibração multivariada utilizando o espectrômetro NIR de bancada e portátil.	67
3.2.1.1	Seleção da região espectral de trabalho	67
3.2.1.2	Pré-processamentos	69
4.2.1.3	Estudo de <i>outlier</i>	72
4.2.1.4	Escolha do número de variáveis incluídas no modelo.....	73
4.2.1.5	Métodos de calibração multivariada	73
3.2.2	Calibração multivariada utilizando os histogramas de cor das imagens digitais	77
3.2.2.1	Histogramas de cor das imagens digitais	77
3.2.2.2	Estudo de <i>outlier</i>	78
4.2.2.3	Escolha do número de variáveis incluídas no modelo.....	78
3.2.1.4	Métodos de calibração multivariada	79
3.2.3	Calibração multivariada utilizando fusão de dados	83
3.2.4	Comparação dos métodos propostos com outros métodos da literatura.....	85
4	AUTENTICAÇÃO DE OVOS DE GALINHA EM PÓ EM TERMOS DE ADULTERAÇÃO COM PROTEÍNA DE SOJA E AMIDO DE MILHO	88
4.1	Experimental	88
4.2.1	Amostras e preparo.....	88
4.2.2	Aquisição das imagens digitais e espectros NIR.....	88
4.2.3	Procedimento quimiométrico.....	88
4.2	Resultados e discussão	89
4.2.1	Espectros NIR portátil e imagens digitais	89
4.2.2	Pré-processamentos nos espectros NIR portátil	90

4.2.3	Análise por componentes principais	92
4.2.4	Autenticação de ovos de galinha em pó por DD-SIMCA.....	94
5	CONCLUSÕES E PERSPECTIVAS	100
5.1	Conclusões.....	100
5.2	Perspectivas.....	101
	REFERÊNCIAS	103
	ANEXO I – ARTIGO DA TESE	120

Capítulo 1

Introdução e objetivos





1 INTRODUÇÃO

1.1 Caracterização do problema

De acordo com a Organização das Nações Unidas para Alimentação e Agricultura (do inglês *Food and Agriculture Organization*), a produção mundial de ovos atingiu, aproximadamente, 92 milhões de toneladas em 2022, representando um aumento de 1,02% em relação ao ano anterior (FAO, 2023). Dentro deste cenário, o Brasil figura entre os dez maiores produtores de ovos de galinha, superando 2,9 milhões de toneladas em 2021, com um consumo per capita anual de 257 unidades por habitante. A produção nacional ocorre majoritariamente na região sudeste do país, especialmente nos estados de São Paulo e Minas Gerais (ABPA, 2022).

Os ovos de galinha é um dos alimentos mais completos para consumo humano, devido ao seu alto valor nutricional, uma vez que é uma rica fonte de vitaminas (A, B, D, E e K), proteínas, lipídios e carboidratos (Ávila e Soares, 2010, Amaral *et al.*, 2016). Contudo, os ovos de galinha é um alimento perecível um alimento perecível, por isso está susceptível a ação do ambiente, fazendo com que a indústria alimentícia opte por utilizá-lo na sua forma desidratada, ou seja, em pó (Biohaz, 2014). Os ovos de galinha em pó são empregados em várias formulações de alimentos nas indústrias de panificação e confeitaria, sobretudo por suas propriedades gelificantes, espumantes e emulsificantes (Belyavin, 2016). Além disso, podem promover o consumo de ovos como ingrediente, uma vez que são um alimento muito seguro em termos microbiológicos, fáceis de transportar e armazenar e apresentam forte potencial nutricional, funcional e econômico (Liu *et al.*, 2021). O monitoramento de sua qualidade tornou-se um motivo de preocupação, uma vez que podem ocasionar perdas econômicas (Kraemer *et al.*, 2010) e risco à saúde humana, principalmente se for adulterado.

Os ovos de galinha em pó podem ser adulterados com proteína de soja e amido de milho, representando uma prática fraudulenta que visa não só reduzir os custos de produção, mas também compromete a qualidade e a segurança do produto. A adulteração com proteína de soja pode resultar em uma alteração dos valores nutricionais e das propriedades dos ovos em pó, bem como representar um risco para pessoas que possuem alergia à soja. Por outro lado, o amido de milho pode ser utilizado a fim de aumentar o volume do produto e diluir o valor proteico, mascarando



ainda mais a sua composição. Essa prática não só engana os consumidores, mas também pode acarretar problemas de saúde pública, destacando ainda mais a importância de um rigoroso controle de qualidade e fiscalização na indústria alimentícia de ovos de galinha em pó (Wiederstein *et al.*, 2023; Caballero-Agosto *et al.*, 2024).

No Brasil, os métodos de referência incluem a determinação de umidade, proteína e fósforo totais (Instituto Adolfo Lutz, 2008; AOAC, Association of Official Analytical Chemists, 1995). Entretanto, essas análises possuem algumas desvantagens, como lentidão, utilização de produtos químicos, gerando resíduos nocivos à saúde e ao meio ambiente, são laboriosas e dispendiosas, já que necessitam de pré-tratamento das amostras e de materiais com alto custo de operação e manutenção, bem como necessitam de pessoas capacitadas, uma vez que essas análises envolve o manuseio de reagentes perigosos.

Neste cenário, a fim de superar tais desvantagens e evitar adulterações, novos métodos analíticos, a partir do uso da espectroscopia de Infravermelho Próximo (NIR, do inglês *Near Infrared*), Imagens Digitais (ID) e Quimiometria, podem ser desenvolvidos. A espectrometria NIR é uma técnica espectroscópica que se baseia na absorção de radiação eletromagnética na região do infravermelho próximo (aproximadamente 700 a 2500 nm). Esses métodos seguros, baratos e rápidos, tornaram-se cada vez mais populares em diversos campos de estudo (Zhao *et al.*, 2018). Por outro lado, a utilização da cor como propriedade óptica da matéria, a qual contém informações químicas, também tem sido cada vez mais recorrente. Quando obtida a partir da utilização de dispositivos de captura de imagens digitais de baixo custo, proporciona vantagens intrínsecas como instrumento analítico, como simplicidade, versatilidade e portabilidade (Gaiao *et al.*, 2006, Lyra *et al.*, 2011).

Devido a sua natureza intrinsecamente multivariada, as análises químicas baseadas em métodos que utilizam a espectroscopia NIR e ID produzem uma elevada quantidade de informações analíticas das amostras, tornando difícil classificar/autenticar ou determinar as concentrações dos analitos de interesse sem a utilização de ferramentas quimiométricas de calibração classificação e/ou multivariada (Arslan *et al.*, 2023, Sáez-Hernández *et al.*, 2022, Duarte *et al.*, 2022).

Recentemente, os seguintes algoritmos de calibração multivariada associados a espectrometria NIR ou ID têm sido utilizados na análise de alimentos: Algoritmo de Projeções Sucessivas para seleção de variáveis em Regressão Linear Múltipla (SPA-

MLR, do inglês *Successive Projections Algorithm for variable selection in Multiple Linear Regression*) (Yang *et al.*, 2022), Regressão por Mínimos Quadrados Parciais (*full-PLS*, do inglês *full-Partial Least Squares regression*) (Arslan *et al.*, 2023; Puertas e Vázquez, 2019; He *et al.*, 2022), Seleção de intervalo (*i*) em Mínimos Quadrados Parciais (*iPLS*) (Pereira *et al.*, 2020, Hosseini *et al.*, 2021) e Algoritmo de Projeção Sucessiva para Seleção de Intervalos em Mínimos Quadrados Parciais (*iSPA-PLS*) (Krepper *et al.*, 2018, Pereira *et al.*, 2020). Além disso, algoritmos de classificação, como a Modelagem Independente e Flexível por Analogia de Classes direcionada pelos dados (DD-SIMCA, do inglês *Data-Driven Soft Independent Modeling of Class Analogy*) também têm sido empregados (Nóbrega *et al.*, 2023; Araújo *et al.*, 2021).

A espectrometria NIR e acoplada a calibração ou classificação multivariada têm sido amplamente empregadas na análise de alimentos, como na determinação do teor de gordura e/ou proteína em hambúrgueres de frango (Krepper *et al.*, 2018), arroz (Arslan *et al.*, 2023), além do teor de sólidos solúveis em peras (Yang *et al.*, 2022); colesterol na gema de ovo (Puertas e Vázquez, 2019; Puertas e Vázquez, 2020) e teor de gordura, umidade e proteína em ovos de galinha homogeneizados (Zhao *et al.*, 2018); amendoim em alho em pó (Rady e Watson, 2022); umidade em amostras de ovos de galinha em pó de uma indústria de processamento de ovos (Watanabe *et al.*, 2018); adulteração de leite de cabra com leite de vaca (Pereira *et al.*, 2021); trigo, farro, triticale, cevada, centeio e azevém em aveia (Mahgoub *et al.*, 2024) e água em ovo líquido (Uysal e Boyaci, 2019).

Vale ainda salientar que as imagens digitais associadas a técnicas de calibração ou classificação multivariada também ganharam destaque na literatura, levando a um aumento significativo no número de artigos publicados, principalmente na utilização em análise de alimentos, como a determinação de acrilamida em torradas (Sáez-Hernández, 2022), amido de mandioca na farinha de trigo (Duarte *et al.*, 2022), gordura em hambúrgueres de frango (Krepper *et al.*, 2018) e adulteração de queijo com farinha de trigo, sêmola de trigo e serragem (Visconti *et al.*, 2023) e leite com formaldeído (Silva *et al.*, 2021).

Os dados provenientes da espectrometria NIR e as imagens digitais podem ser combinados por técnicas de fusão de dados, a fim de obter uma visão mais abrangente e precisa de um determinado parâmetro estudado. Essas combinações normalmente melhoram a capacidade preditiva do modelo (Vohland *et al.*, 2024).



1.2 Objetivos

1.2.1 Objetivo geral

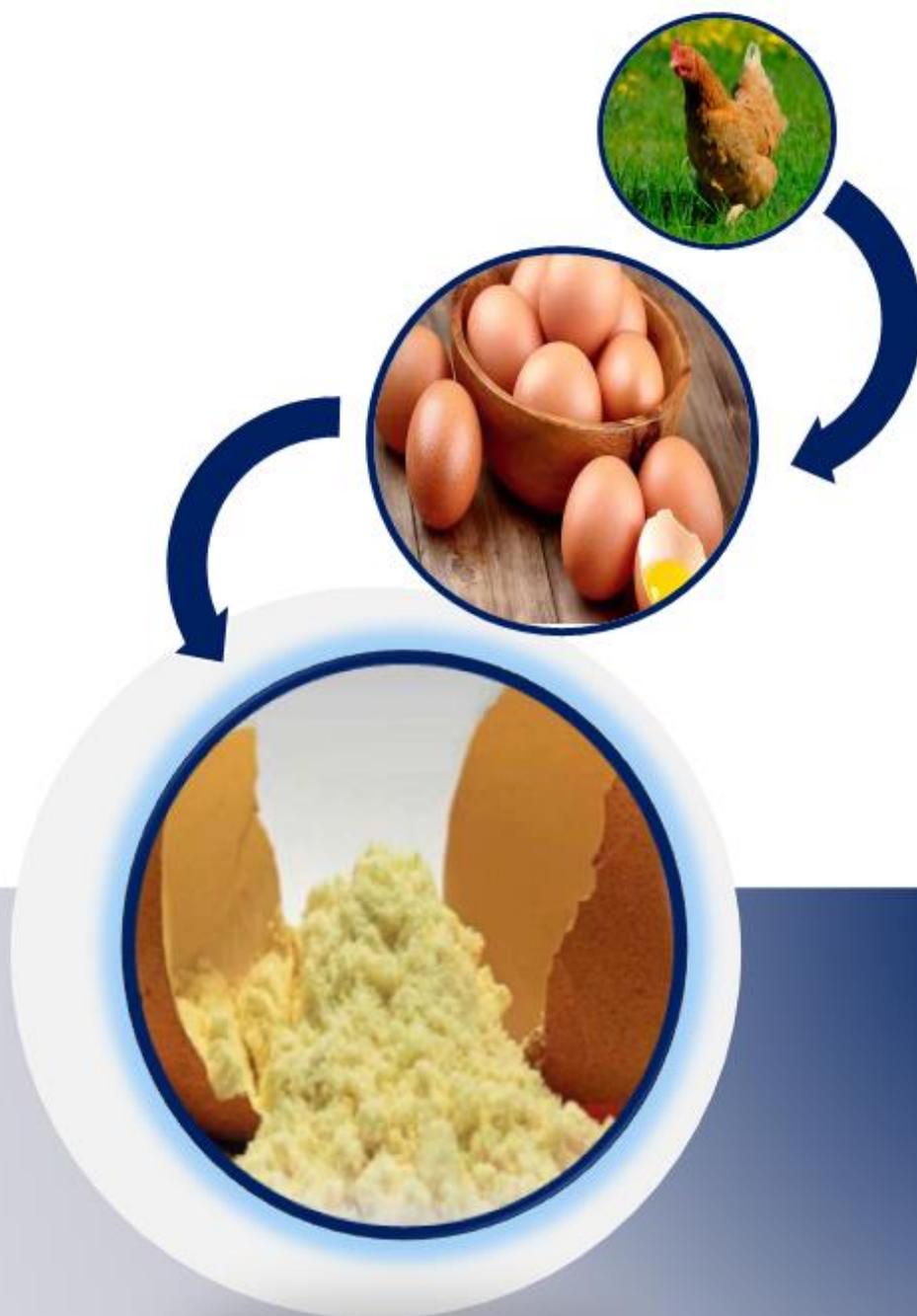
- ❖ Desenvolver métodos utilizando a espectroscopia NIR, imagens digitais e modelagens quimiométricas de classificação e calibração multivariada para avaliação da qualidade de ovos de galinha em pó.

1.2.2 Objetivos específicos

- ❖ Obter as amostras de ovos de galinha em pó por meio do processo de liofilização;
- ❖ Registrar os espectros NIR utilizando um equipamento comercial de bancada e portátil e as imagens digitais usando um *scanner* de amostras de ovo de galinha em pó;
- ❖ Utilizar técnicas de pré-processamento, a fim de aumentar a capacidade preditiva dos modelos quimiométricos construídos e validados para determinação de umidade e teores de proteínas e fósforos totais em amostras em ovos de galinha em pó;
- ❖ Desenvolver métodos quimiométricos para determinação de umidade e teores de proteínas e fósforos totais em amostras em ovos de galinha em pó utilizando espectros NIR, imagens digitais, fusão de dados e quimiometria;
- ❖ Validar os modelos SPA-MLR, *full*-PLS, *i*PLS, *i*SPA-PLS construídos;
- ❖ Desenvolver métodos quimiométricos utilizando DD-SIMCA para autenticar amostras de ovos de galinha em pó quanto a sua adulteração com proteína de soja e amido de milho;
- ❖ Comparar o desempenho dos algoritmos e dos instrumentos utilizados em relação a capacidade preditiva dos modelos construídos e validados.

Capítulo 2

Fundamentação teórica



2 FUNDAMENTAÇÃO TEÓRICA

2.1 Espectroscopia no Infravermelho Próximo (NIR)

2.1.1 Fundamentação

A espectroscopia no infravermelho é baseada no fenômeno de absorção da luz nos comprimentos de onda entre 800 nm e 1.000.000 nm ($12.500 - 10 \text{ cm}^{-1}$), podendo ser dividida, conforme sua faixa espectral, em três sub-regiões: Infravermelho Distante (FIR, do inglês *Far-Infrared*) de 25.000 – 1.000.000 nm ($400 - 10 \text{ cm}^{-1}$), Infravermelho Médio (MIR, do inglês *Mid-Infrared*) de 2.500 – 25.000 nm ($4.000 - 400 \text{ cm}^{-1}$) e Infravermelho Próximo (NIR, do inglês *Near-Infrared*) de 800 – 2.500 nm ($12.500 - 4.000 \text{ cm}^{-1}$) (Deidda *et al.*, 2019).

A espectroscopia NIR pode ser facilmente diferenciada da espectroscopia na região do MIR, uma vez que a primeira é proveniente de sobretons e combinações de modos vibracionais fundamentais, enquanto a segunda envolve, principalmente, as bandas de vibrações fundamentais (Ozaki; Huck; Béc, 2018).

De acordo com a literatura, os principais fatores que proporcionam as propriedades espectrais, ou seja, frequência e bandas de absorção na região do NIR são: ressonância de Fermi e a anarmonicidade. Para que isso ocorra, é necessário que a molécula possua uma variação no momento de dipolo, durante o movimento rotacional ou vibracional. Nesse caso, o campo elétrico formado pode interagir com a molécula, ocasionando variações na amplitude de um dos seus movimentos (Singh; Szostak; Czarnecki, 2021).

A ressonância de Fermi é um acoplamento não harmônico de dois modos vibracionais, o fundamental e harmônico, quando a sua diferença de energia é muito baixa (Han, *et al.*, 2020). Este tipo de interação promove uma maior separação entre as posições das duas bandas, bem como intensifica a banda harmônica (Pasquini, 2018), favorecendo o aparecimento de bandas de combinação na região de 1.900 nm e 2.500 nm. Isso ocorre devido ao acoplamento entre as vibrações fundamentais e os sobretons, ou seja, as frequências resultantes são somas dos múltiplos de cada frequência da interação. Desse modo, duas ligações que vibram em frequências similares produzem duas bandas de absorção na região do NIR com frequências próximas (Kafle, 2020).

De maneira geral, as bandas de absorção da espectroscopia NIR são mais fracas do que aquelas obtidas por MIR. Além disso, como sobretons e modos de combinação se sobrepõem devido à ressonância de Fermi, a análise de bandas nessa região não é trivial. A maioria das bandas que aparecem são relacionadas aos grupos funcionais OH, CH e NH. Isso pode ser explicado devido às vibrações do tipo alongamento, as quais aparecem em uma região de alta frequência. A ligação de hidrogênio e a interação entre as moléculas podem ainda causar o deslocamento de banda, sendo a mudança muito maior do que uma banda de MIR (Ozaki, Genkawa, Futami, 2017).

A espectroscopia NIR pode fornecer informações importantes acerca das transições *d-d*, sendo as bandas de alongamento do tipo OH e NH nas espécies poliméricas e monoméricas mais bem separadas, quando comparadas com a região MIR. Até mesmo aquelas atribuídas aos terminais livres das espécies poliméricas (OH e NH) podem ser facilmente identificadas nessa faixa espectral. Como a anarmonicidade é maior nessa região, as bandas dos primeiros sobretons dos modos de alongamento (OH e NH) de espécies monoméricas são maiores do que aquelas que surgem das espécies poliméricas. Isso facilita o monitoramento do processo de dissociação das espécies poliméricas em monoméricas, utilizando o primeiro tom do modo de alongamento (OH ou NH) de espécies monomérica estudadas, ao invés de usar as bandas fundamentais (Ozaki, Genkawa, Futami, 2017).

A sensibilidade dos métodos analíticos baseados em espectroscopia NIR permite a obtenção de informações espectrais qualitativas e quantitativas, as quais estão associadas com o comprimento de onda e a intensidade das absorções. Isso acontece porque o comprimento de onda de uma banda de absorção aumenta em função da massa dos átomos envolvidos em cada modo vibracional anarmônico (sobretons ou bandas de combinação) e diminui com o aumento da força da ligação química. Desse modo, as mudanças que podem ocorrer nessas duas quantidades fundamentais (massa e energia da ligação química) são detectadas deslocando o comprimento de onda da banda de absorção (Pasquini, 2018). O registro da informação analítica pode ser realizado utilizando espectrofotômetros NIR convencionais de bancada e dispositivos portáteis, cuja instrumentação é descrita a seguir.

2.1.2 Instrumentação

A utilização da espectroscopia NIR teve início na década de 1970, quando apareceram os primeiros espectrômetros NIR computadorizados. Essa técnica analítica foi inicializada por Philip Williams, que trabalhava na Comissão Canadense de Grãos, e atualmente é amplamente usada em laboratórios industriais para realização de análises de rotina. Este pesquisador substituiu a determinação convencional de proteína total em trigo empregando o método de Kjeldahl por análise espectrométrica na região do NIR. O método Kjeldahl é muito laborioso, poluente, de baixo desempenho, além de consumir muitos reagentes. A partir de tal estudo, ficou evidente as vantagens da utilização da espectroscopia NIR, como a rapidez de análise, ausência de pré-tratamento da amostra e exatidão. Sendo assim, o trabalho proposto por Williams foi extremamente importante para motivar as empresas a produzir e comercializar os instrumentos NIR (Pasquini, 2018).

Recentemente, os instrumentos NIR de bancada evoluíram com a aplicação da Transformada Fourier (FT, do inglês *Fourier Transform*), baseada em interferometria, em substituição àqueles que empregavam grades de difração. Dessa maneira, os espectrofotômetros NIR ganharam mais robustez e durabilidade (Pasquini, 2018), além de reduzir de tamanho e peso com os dispositivos portáteis. Isso possibilitou a realização de diversos estudos em campo, tendo como principal vantagem, a preservação da amostra, já que evita a contaminação, uma vez que a análise pode ser realizada *in situ* (Cherney, Digman, Cherney, 2021).

Existem vários instrumentos NIR convencionais de bancada que funcionam de maneira diferente. No entanto, considerando a ilustração da Texas Instruments® (Figura 2.1), durante a varredura, a amostra absorve uma determinada quantidade específica de radiação, refletindo a radiação difusa não absorvida no sistema, que é captada pela lente de coleta, direcionada para a fenda e colimada. Em seguida, a radiação atinge uma grade de difração que, com a lente de foco, dispersa a radiação nos comprimentos de onda constituintes da amostra, espalhando-as horizontalmente no detector de arranjo linear de InGaAs. O sistema ótico registra os comprimentos de onda de 12000 cm^{-1} a 4000 cm^{-1} (Texas Instruments, 2017).

A principal diferença entre o NIR convencional de bancada e o equipamento portátil, modelo DLP NIRscan Nano EVM (Figura 2.1), da Texas Instruments®, refere-se à substituição do arranjo linear de detectores tradicional por um Dispositivo de

Microespelhos Digitais (DMD, do inglês *Digital Micromirror Device*) para seleção de comprimento de onda e um detector pontual. Desse modo, enquanto no NIR portátil a varredura é realizada sequencialmente nas colunas específicas, direcionando (+17°) ou desviando (-17°) a luz do caminho ótico, no NIR convencional de bancada toda a luz difratada é direcionada para o detector (Texas Instruments, 2017).

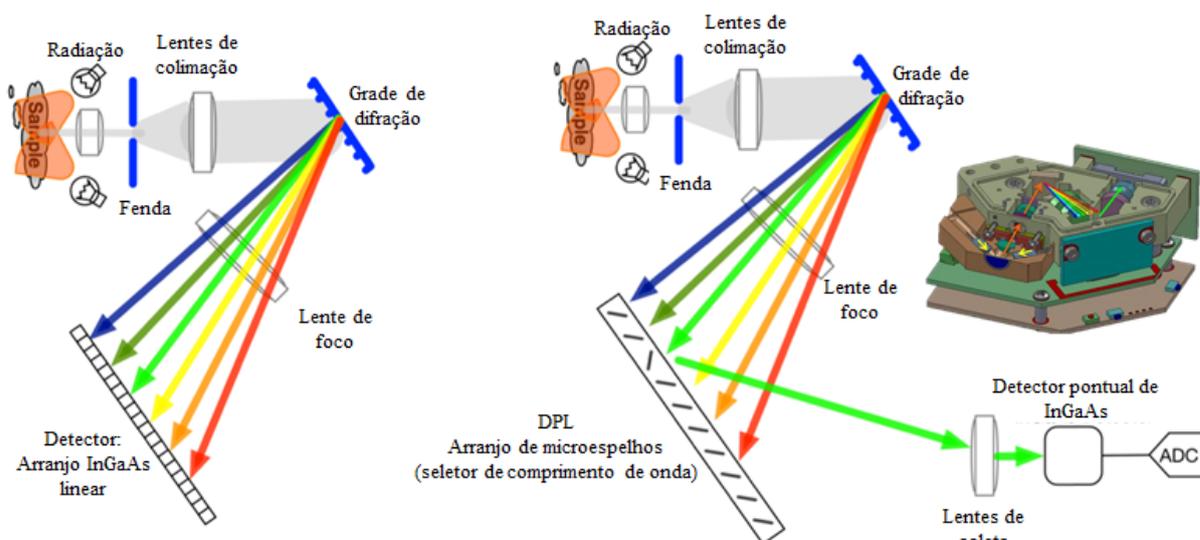


Figura 2.1 - Esquema do Espectrofotômetro NIR convencional e portátil. Adaptado de Texas Instruments (2017)

Os detectores mais comuns, empregados em instrumentos modernos, são aqueles baseados em semicondutores de InGaAs, os quais apresentam uma boa detecção na região de 1100-1750 nm, bem como permite a obtenção de uma resposta mais rápida, principalmente nos dispositivos portáteis (Pasquini, 2018).

Durante os últimos anos, a maior parte dos instrumentos, especialmente os NIR portáteis, foram avaliados frente aos espectrofotômetros convencionais de bancada. A maioria dos estudos em amostras de alimentos revelam resultados satisfatórios no desempenho dos instrumentos portáteis associado às técnicas quimiométricas de calibração e classificação (Zhou *et al.*, 2021; Yu, *et al.*, 2020; Patel; Toledo-Alvarado; Bittante, 2021). No entanto, não existem comparações na literatura entre dispositivos portáteis e convencionais para a matriz estudada nesse trabalho de tese. Uma outra maneira de se obter informações analíticas, é utilizando as imagens digitais, cujos conceitos são apresentados no tópico seguinte.

2.2 Imagens Digitais

2.2.1 Fundamentação

Uma imagem pode ser definida através de uma função bidimensional, $f(x, y)$, onde x e y são as coordenadas espaciais e sua amplitude denominada de intensidade. Quando os valores de $f(x, y)$ e intensidade são todos finitos e discretos, a imagem é chamada de imagem digital (Gonzalez; Woods, 2018). O menor elemento formado por uma imagem digital é denominado de pixel. Cada localização do pixel no espaço representa um valor numérico único, ou seja, o nível de sinal naquele ponto, que pode variar de 0-255, caso os sinais sejam de 8 bits (Rowlands, 2020).

As imagens digitais estão fundamentadas nos modelos de cor, que tem por objetivo especificar as cores de maneira padronizada. De modo geral, um modelo de cor é uma especificação de um sistema de coordenadas tridimensional (x, y, z), em que cada ponto representa uma cor (Ware, 2021; Diniz, 2020). A cor depende da intensidade da fonte de radiação. As propriedades ópticas concernentes às variações de cor que envolvem a emissão e absorção da radiação da faixa do visível podem ser utilizadas para identificar ou quantificar um determinado analito, já que há uma correlação entre a mudança, aparecimento ou desaparecimento da radiação eletromagnética, e a informação química da espécie de interesse. Assim, a luz interage com a amostra, que por sua vez é registrada por um dispositivo de captura de imagens (Diniz, 2020).

Conforme a literatura, os espaços de cores mais comumente utilizados são Vermelho-Verde-Azul (RGB, do inglês *Red-Green-Blue*), Ciano-Magenta-Amarelo-Preto (CMYK, do inglês *Cyan-Magenta-Yellow-Black*), Matiz-Saturação-Valor (HSV, do inglês *Hue-Saturation-Value*), Matiz-Saturação-Luminância (HSL, do inglês *Hue-Saturation-Luminance*), Comissão Internacional de Iluminação (CIE, do inglês *Commission Internationale de l'Éclairage*) e Escala de cinzas (GS, do inglês *Gray Scale*) (Gonzalez; Woods, 2018; Ware, 2021; Diniz, 2020).

O RGB é o sistema de cor mais utilizado pelos dispositivos de aquisição de imagens digitais. Refere-se à mistura de cores dominantes (primárias aditivas), em diferentes níveis de combinação e intensidade, podendo realizar a simulação de todas as cores da região visível. Assim, o espaço tridimensional é formado por três radiações primárias (vermelho, verde e azul), sendo que qualquer cor interna pode ser produzida

por cada uma das cores primárias (**Figura 2.2**). Nesse caso, o olho humano visualiza branco se a luz refletida contém máxima intensidade de vermelho, verde e azul, ao passo que não existe luz quando a cor preta é percebida (Ware, 2021; Diniz, 2020).

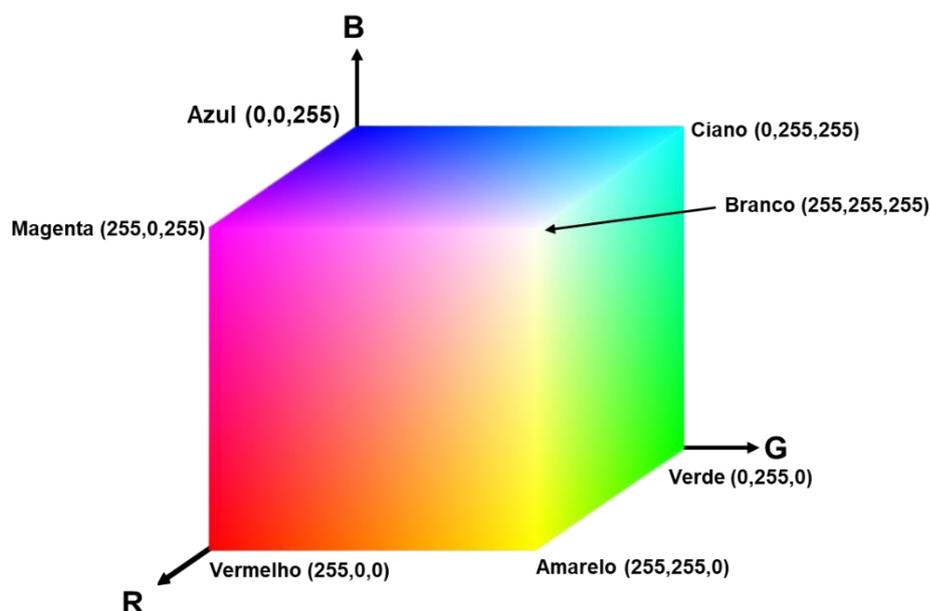


Figura 2.2 - Modelo de cor RGB. Adaptado de <https://stackoverflow.com/questions/22991809/how-to-produce-rgb-cube-matrix-in-python>

No cubo de cores RGB (**Figura 2.2**), cada eixo é representado pela contribuição de uma cor primária com a intensidade de brilho variando entre 0 e 255. Seu vértice é representado pelas cores secundárias em um plano, gerado por um plano por um par de eixo (Lyra, 2012)

A imagem digital do modelo de cor denominado de escala de cinzas possui valores iguais dos componentes do RGB, sendo constituída por uma imagem monocromática de tons de cinza, na qual o valor do pixel é composto por uma única amostra dentro do espaço de cores. Assim, a diagonal do cubo RGB (**Figura 2.2**) representa a escala de cinzas e sua intensidade varia entre o preto (baixa intensidade) e o branco (máxima intensidade) (Mu, *et al.*, 2021).

O modelo de cores denominado de HSV (**Figura 2.3**) representa as cores de maneira semelhante ao olho humano, já que se baseia nos parâmetros de matiz, saturação e intensidade definidos em um triângulo para definir uma determinada cor. A matiz (H) é o ângulo entre o vetor e o eixo vermelho, a saturação (S) é a distância do ponto P ao centro do triângulo e o valor (V) refere-se à distância perpendicular do ponto preto (Jain, *et al.*, 2021).

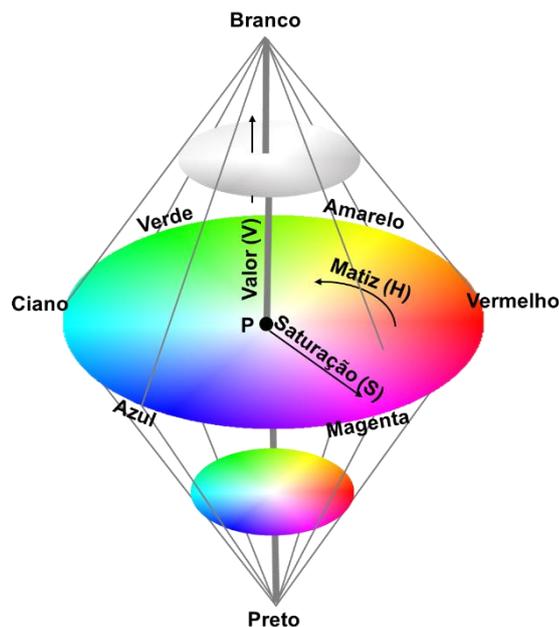


Figura 2.3 - Sistema de cor HSV. Adaptado de Blotta et al. (2011)

Em análises químicas, a quantificação de cores é convencionalmente realizada por colorimetria, através de métodos que possibilitam a decomposição, análise e a descrição da radiação visível dos pixels delimitados na região de interesse selecionada (ROI, do inglês *Region of Interest*) (Fan, et al., 2021). Neste tipo de procedimento, a estatística de primeira ordem, obtida a partir dos valores médios dos canais individuais dos pixels decompostos usando o espaço RGB, é utilizada para realizar regressão linear simples, correlacionando-a com a concentração do analito. Essa etapa de preparo da amostra é importante para garantir a seletividade e obter uma cor homogênea do sistema químico estudado, simplificando, assim, o procedimento de calibração matemática. No entanto, isso falha quando se tem que realizar análises não destrutivas ou em grande escala, principalmente, quando a amostra é sólida ou complexa (presença de interferentes). Desse modo, a calibração multivariada é necessária para extrair as informações químicas presentes em uma imagem digital (Diniz, 2020).

Neste sentido, a utilização de histogramas de cores como um sinal analítico é muito importante, uma vez que permitem acessar de maneira simples as informações analíticas sem a necessidade de qualquer outro processamento da imagem digital, exceto a extração do histograma, permitindo análises rápidas quando acoplada às técnicas quimiométricas. Os histogramas de cores revelam a ocorrência das

intensidades de cores ao longo da superfície da amostra, no canal que está sendo monitorado. (Diniz, 2020).

2.2.2 Instrumentação

As informações analíticas podem ser adquiridas através de simples dispositivos de aquisição de imagens, dentre os principais encontram-se os de projeção, como câmera digital, *webcam* e *smartphone*, além dos de varredura, como o *scanner*. A **Figura 2.4** resume o procedimento de aquisição das imagens digitais e extração do histograma.



Figura 2.4 - Esquema da aquisição das imagens digitais e extração do histograma, os quais guardam a informação analítica. Elaborado no biorender.com

Em linhas gerais, os métodos desenvolvidos a partir de imagens digitais, baseiam-se na captura da imagem da amostra, seleção da Região de Interesse (ROI), obtenção do histograma, arranjo dos dados em uma matriz de dados bidimensional e realização dos tratamentos quimiométricos, os quais buscam a partir das variações de cores intrínsecas de cada amostra uma correlação química com os analitos investigados (Diniz, 2020).

Do ponto de vista prático, para obter um bom desempenho dos modelos quimiométricos, cujos conceitos serão apresentados no tópico seguinte, deve-se garantir uma boa reprodutibilidade das medidas, ou seja, obtenção de imagens digitais de alta qualidade (alta resolução, bom contraste e foco, além da ausência de ruído ou distorção) com a menor interferência do meio externo possível. Para isso, deve-se manter fixos o posicionamento, a luminosidade e a distância entre o porta-amostra e a câmera. Isso evita etapas de pré-processamento/correção de imagens empregando

operadores matemáticos para melhoramento de imagem, tais como redução de ruído e correção do nível de cinza e desfoque. Uma vez que os histogramas são resistentes a mudanças de posição e orientação e guardam informação química útil da amostra, métodos de classificação e/ou calibração multivariada podem ser utilizados, dependendo da finalidade de aplicação (Diniz, 2020). Tais técnicas quimiométricas serão abordadas a seguir.

2.3 Quimiometria

A quimiometria é uma área da química que utiliza métodos estatísticos e matemáticos a fim de extrair a máxima quantidade de informações possíveis, a partir de dados químicos. Os primeiros registros da quimiometria surgiram no início dos anos de 1970, realizados pelos pesquisadores Wold, Kowalski e Massart, sendo o primeiro o responsável por incluí-la como uma disciplina e campo de estudo dentro da Química (Héberger, 2008; Alexandre-Tudo, 2022; Inobeme *et al.*, 2022).

O desenvolvimento de computadores mais poderosos e modernos contribuiu para um grande avanço da quimiometria, principalmente no que se refere à velocidade no processamento das informações, uma vez que anteriormente os algoritmos passavam horas para serem processados pelos *desktops* ou exigiam acesso à *mainframes*, tornando os cálculos mais demorados (Brereton *et al.*, 2017). Entretanto, os dados obtidos nos instrumentos possuem alguns artefatos prejudiciais às análises e, portanto, surge a necessidade de pré-processar os dados. Tais técnicas utilizadas nesse trabalho serão discutidas a seguir.

2.3.1 Técnicas de pré-processamento

Os pré-processamentos dos dados referem-se a qualquer manipulação dos dados realizada antes da construção dos modelos quimiométrico e tem por finalidade tratar as matrizes de dados, reduzindo as fontes de variação não informativas existentes. Em medidas espectroscópicas, especialmente em NIR, tais flutuações são ocasionadas por diferentes fatores, como desvio da linha de base, ruídos espectrais ou distribuição não uniforme das partículas. Como essas fontes de variação não estão relacionadas ao objetivo para o qual os dados foram coletados, uma vez que não há correlação com as características químicas da amostra, essas flutuações devem ser

minimizadas ou removidas. Sua permanência prejudica o desempenho preditivo dos modelos quimiométricos elaborados (Gerretzen *et al.*, 2015; Mishra *et al.*, 2021).

Os pré-processamentos podem ser realizados nas amostras, em que se opera uma amostra de cada vez sob todas as variáveis da matriz, e nas variáveis, o qual opera-se uma variável por vez sob todas as amostras. Existem diferentes técnicas de pré-processamento, dentre as quais se destacam: derivação *Savitzky-Golay* (SGD); Correção de Linha de Base Linear (LBC, do inglês *Linear Baseline Correction*); Correção de Linha de Base *Offset* (OBC, do inglês *Offset Baseline Correction*); Variação Normal Padrão (SNV, do inglês *Standard Normal Variation*) e Correção Multiplicativa de Sinal (MSC, do inglês *Multiplicative Scatter Correction*).

O método de derivação *Savitzky-Golay* é realizado em duas etapas: a suavização do espectro, seguido do cálculo da derivada em cada ponto, utilizando uma janela móvel simétrica. Antes da derivação, um polinômio de baixa ordem é ajustado, por mínimos quadrados, aos pontos da janela, reduzindo a relação sinal ruído. Esse tipo de técnica permite a melhor visualização de picos sobrepostos, bem como corrige o deslocamento da linha de base existente nos dados. Sendo assim, a primeira derivada é aplicada para eliminar o *offset*, enquanto a segunda derivada elimina a inclinação da linha de base (Robert; Grosselin, 2022).

Além do ruído aleatório, os dados podem conter variações sistemáticas as quais não estão correlacionadas com o objetivo da análise química. Tais variações são denominadas de linhas de base e se não removidas podem prejudicar o modelo. A correção de linha de base pode ser realizada por *offset* e/ou linear. No *offset*, o menor valor para uma determinada variável do espectro é subtraído de todas as outras, resultando como valor mínimo o zero e em demais valores positivos. Já a correção de linha de base linear transforma uma linha de base inclinada em uma linha de base horizontal. Essa técnica consiste em apontar duas variáveis as quais definiram a nova linha de base. Assim, ambas as variáveis são definidas como zero e as demais são transformadas conforme a interpolação / extrapolação linear (Robert; Grosselin, 2022; Bhise; Kulkarni, 2017).

A técnica de pré-processamento denominada de SNV corrige os dados a partir do espectro médio e seu desvio padrão, através da equação 2.1:

$$x_{i\text{SNV}} = \frac{x_i - \bar{x}_i}{s_i} \quad (2.1)$$

em que $\mathbf{x}_{i\text{SNV}}$ = Espectros pré-processados por SNV; \mathbf{x}_i = Valor do i -ésimo espectro; \bar{x}_i = Valor médio do i -ésimo espectro; s_i = Valor do desvio padrão do i -ésimo espectro. Esse tipo de método corrige os efeitos multiplicativos, bem como as interferências de espalhamento e não necessita de um espectro médio de referência (Gautam *et al.*, 2015; Robert; Grosselin, 2022; Rinnan; Berg; Engelsen, 2015; Ferreira, 2015).

O cálculo do MSC é realizado de maneira similar ao SNV, porém ao invés de utilizar o espectro médio e o desvio padrão, usa-se a regressão linear para estimar o coeficiente linear e angular entre o espectro médio de referência (X_{ref}) e o adquirido. O cálculo é realizado através da equação 2.2:

$$\mathbf{x}_{i\text{MSC}} = \frac{\mathbf{x}_i - b_i}{a_i} \quad (2.2)$$

em que: $\mathbf{x}_{i\text{MSC}}$ = Espectros pré-processados por MSC; $\mathbf{x}_i = X_{\text{ref}} \cdot a + b$; a_i = coeficiente angular da equação obtida entre os espectros; b_i = coeficiente linear da equação obtida entre os espectros. Nessa técnica, o espectro de referência é adquirido por meio de uma amostra padrão. No entanto, em sua falta pode-se utilizar o espectro médio como referência (Gautam *et al.*, 2015; Robert; Grosselin, 2022; Rinnan; Berg; Engelsen, 2015; Ferreira, 2015).

Após os dados serem processados, são submetidos a técnicas de classificação (reconhecimento de padrões) ou calibração multivariada, as quais são as principais áreas da Quimiometria, sendo seus conceitos discutidos nos tópicos seguintes.

2.3.2 Calibração multivariada

A calibração multivariada é uma subárea da quimiometria que tem por finalidade relacionar, correlacionar ou modelar a resposta medida com o parâmetro físico ou químico de interesse previamente determinado a partir de um método de referência. Sendo assim, seu principal objetivo é construir modelos para previsão de amostras desconhecidas (Lavine; Workman, 2010).

Várias ferramentas foram desenvolvidas, a fim de realizar o tratamento dos dados em diversas aplicações, principalmente daqueles originados da Espectroscopia NIR e os histogramas de cores provenientes das imagens digitais. Esses instrumentos geram uma grande quantidade de dados (variáveis) que são difíceis de serem

interpretados. Desse modo, torna-se necessária a utilização de métodos quimiométricos a fim de extrair informações químicas úteis (Medina, *et al.*, 2019). Tais ferramentas podem ser organizadas em duas principais categorias: reconhecimento de padrões e calibração multivariada (Figura 2.5).

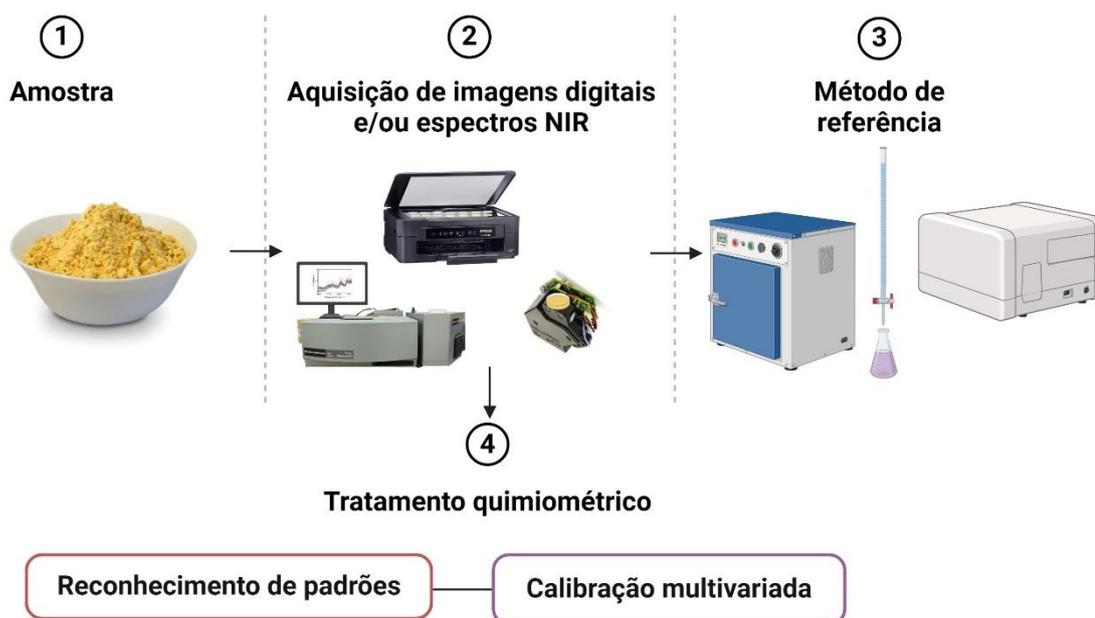


Figura 2.5 - Diagrama esquemático da aquisição dos dados utilizando NIR e imagens digitais. Elaborado no biorender.com

Em linhas gerais, a calibração multivariada é realizada em três etapas: calibração, validação e predição. A primeira é obtida correlacionando a matriz de dados das variáveis medidas (matriz X) com a coluna da propriedade físico ou química de interesse (matriz Y). Já na segunda, o modelo é validado com as amostras que não foram incluídas na calibração. Na última etapa é realizada a predição com amostras desconhecidas ou por validação cruzada (Westad; Marini, 2015).

Existem dois métodos para realizar a validação: a externa, quando um conjunto distinto daquele que o modelo foi elaborado é utilizado, e a interna ou validação cruzada (do inglês *cross-validation*), utilizada nesse trabalho. Nesse tipo de validação, o conjunto de calibração é dividido em dois subconjuntos: calibração e validação. Inicialmente, constrói-se o modelo com o subconjunto da calibração, o qual é utilizado na previsão do conjunto de validação. O processo é repetido várias vezes até que todas as amostras do conjunto de calibração tenham sido usadas no conjunto de

validação. Os resultados da previsão são utilizados para verificar a performance do modelo final (Berrar, 2019; Steidle Neto *et al.*, 2017).

No que se refere ao particionamento das amostras nos diferentes conjuntos de dados, normalmente utiliza-se diferentes algoritmos, dentre os quais destacam-se o Kennard-Stone (KS) e a partição do conjunto de amostras baseado nas distâncias conjuntas XY (SPXY, do inglês *Sample set Partitioning based on join XY distances*). O primeiro utiliza como critério de seleção a maior distância existente entre amostras que já selecionada pelo algoritmo (Kennard; Stone 1969). Já o segundo, é uma extensão do KS, porém no cálculo das distâncias entre as amostras leva em consideração as diferenças existentes entre as variáveis independentes X e a variável Y que é a propriedade avaliada, garantindo a uniformidade na distribuição das amostras nos conjuntos particionados (Galvão *et al.*, 2005).

Para realizar a calibração multivariada usando conjunto de dados originados da Espectroscopia NIR e os histogramas de cor provenientes das imagens, as ferramentas quimiométricas mais interessantes para ser utilizadas são: a Regressão Linear Múltipla (MLR, do inglês *Multiple Linear Regression*) e a Regressão por Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*). Tais métodos serão abordados no tópico seguinte.

2.3.2.1 Regressão Linear Múltipla

A Regressão Linear Múltipla é o método de calibração multivariada mais simples. Nela assume-se que a propriedade de interesse, como a concentração, é função linear das respostas instrumentais, podendo ser descrita como (Equação 2.3):

$$Z = Xb + e \quad (2.3)$$

em que X = resposta instrumental; Z = propriedade de interesse; b = vetor dos coeficientes lineares de regressão; e = vetor dos resíduos ou erro na modelagem.

O valor do vetor b das estimativas dos coeficientes lineares é encontrado a partir dos mínimos quadrados, os quais minimizam o vetor dos resíduos (Equação 2.4):

$$b = (X^T X)^{-1} X^T Y \quad (2.4)$$

em que: b = vetor dos coeficientes lineares; X = respostas instrumentais; Y = propriedade de interesse; X^T = Transposta da matriz X .

No entanto, dois problemas podem afetar os modelos MLR: a colinearidade, uma vez que pode haver uma alta correlação ao longo das colunas de X , tornando-se impossível de se encontrar o termo $(X^T X)^{-1}$ ou grandes erros nos valores de b ; e infinitas soluções, não devendo o número de variáveis ser maior que o número de amostras do conjunto de calibração (Beebe; Pell; Seasholtz, 1998; Ferreira, 2015). Tais problemas podem ser solucionados utilizando seleções de variáveis, as quais são discutidas no tópico 2.3.2.3.

2.3.2.2 Regressão em Mínimos Quadrados Parciais

A Regressão por Mínimos Quadrados Parciais é o método de calibração multivariada mais utilizado em análises de alimentos. PLS recupera uma equação que prevê os valores da matriz Y (parâmetro) com base nos valores da matriz X (amostras), sendo possível estabelecer relação entre duas matrizes. A análise PLS é extremamente útil para conjuntos de amostras nos quais o número de variáveis independentes é muito maior que as variáveis dependentes (Medina *et al.*, 2019).

Nesse método, as matrizes X e Y são modeladas em fatores gerando dois grupos de matrizes: *scores* (T) e *loadings* (P), conforme equações 2.5 e 2.6, consecutivamente:

$$X = T P^T + E \quad (2.5)$$

$$Y = U C^T + F \quad (2.6)$$

O modelo resultante é calculado relacionando linearmente os *scores* de X (matriz T) com os *scores* de Y (matriz U), de acordo com as equações 2.7 e 2.8:

$$U = B T + G \quad (2.7)$$

$$Y = B T C^T + H \quad (2.8)$$

O cálculo para estimar os vetores dos coeficientes de regressão b pode ser realizado utilizando os Mínimos Quadrados Parciais Iterativos Não Lineares (NIPALS, do inglês *Non Linear Iterative Partial Least Squares*), no qual os vetores dos *loadings* e *scores* são calculados de maneira iterativa, um de cada vez. Outro método muito reportado na literatura e empregado pelo algoritmo utilizado nesse trabalho é a Decomposição de Valor Singular (SVD, do inglês *Singular Value Decomposition*), que decompõe os vetores em eixos ortogonais, resultando em três matrizes (Beebe; Pell; Seasholtz, 1998; Ferreira, 2015).

Apesar da importância da técnica e de sua imensa aplicabilidade, o PLS pode apresentar algumas desvantagens, como natureza complexa dos cálculos e a tendência de superestimar (*overfitting*) os ruídos na matriz Y , fornecendo modelos quimiométricos que podem não traduzir a realidade (Beebe; Pell; Seasholtz, 1998). Somasse a isso, o MLR também é uma ferramenta que apresenta desvantagens, principalmente devido à colinearidade. Sendo assim, esses inconvenientes podem ser sanados utilizando técnicas de seleções de variáveis, que serão discutidas no tópico seguinte.

2.3.2.3 Seleções de variáveis em Quimiometria

Em linhas gerais, na seleção de variáveis atribui-se peso zero para as variáveis não informativas. Isso possui várias vantagens, como redução de custo e tempo, indo ao encontro do princípio da parcimônia; facilita a interpretação química, uma vez que menos variáveis são incluídas na construção dos modelos; minimiza a propagação de erros, reduzindo o impacto de variáveis irrelevantes, ruidosas ou redundantes e melhora a predição (Chowdhury; Turim, 2020). Dentre os métodos de seleção de variáveis destacam-se: a seleção de intervalos, o algoritmo das projeções sucessivas e o algoritmo das projeções sucessivas para seleção de intervalos.

O Algoritmo das Projeções Sucessivas (SPA, do inglês *Successive Projection Algorithm*) é um método de seleções de variáveis utilizado para melhorar o condicionamento da matriz X (respostas instrumentais), reduzindo os efeitos da colinearidade existentes no conjunto de dados de calibração e ajudando na diminuição

da propagação de ruídos. Para isso, o algoritmo utiliza simples operações de projeções de subconjuntos de vetores (variáveis) no espaço, começando com um vetor de partida, em seguida incorpora vetores minimamente colineares até que uma condição de parada seja satisfeita. A cadeia de variáveis resultante é avaliada a partir da previsão do modelo MLR resultante (Galvão *et al.*, 2008; Gomes *et al.*, 2013).

O SPA pode ser dividido em três fases. Na fase 1, os dados das respostas instrumentais são centrados na média e colocados em uma matriz (X_{cal}) com dimensões ($n \times k$) de maneira que k -ésima variável (X_k) esteja associada ao k -ésimo vetor coluna $X_k \in \mathfrak{R}^N$. Tais vetores são submetidos à sucessivas projeções, resultando em K cadeias de M variáveis, em que o número máximo de variáveis, que podem ser incluídas no modelo MLR, é dado por $M = \min(N - 1, k)$. A cadeia k é iniciada de forma que X_k , $k = 1, \dots, k$, sendo escolhidos os elementos que possuem menor colinearidade com os calculados anteriormente. Na fase 2, os subconjuntos produzidos anteriormente são avaliados. Para isso, um conjunto de índices $\{SEL(1, k), SEL(2, k), \dots, SEL(m, k)\}$ é definido para os subconjuntos de m variáveis, iniciando por X_k , em que k varia de 1 a w (colunas representativas) e m varia de 1 a $(w - 1)$. Desse modo, o melhor par (m, k) é escolhido quando se obtém um menor erro da Raiz Quadrada do Erro Médio Quadrático da Validação (RMSEV, do inglês *Root Mean Square Error of Validation*). Na fase 3, ocorre a eliminação das variáveis não informativas, com base em um teste F , visando melhorar a parcimônia do modelo (Galvão *et al.*, 2008; Gomes *et al.*, 2013; Soares *et al.*, 2013).

O algoritmo de seleção de Intervalos (i) é uma extensão interativa elaborado para o *full*-PLS, sendo denominado de Regressão de Mínimos Quadrados Parciais por Intervalos (*i*PLS, do inglês *Interval Partial Least-Squares Regression*). Trata-se de um método no qual são desenvolvidos modelos PLS em intervalos com distâncias iguais. Dessa maneira, o algoritmo avalia as subdivisões espectrais, a fim de manter as informações relevantes e remover as interferências de outras regiões. Então, um novo modelo PLS é construído somente com as variáveis que foram selecionadas (Nørgaard *et al.*, 2000).

Em linhas gerais, os submodelos construídos utilizando seleção de Intervalos são comparados frente ao *full*-PLS. Essa análise é baseada, principalmente, na RMSECV que é calculada para cada fator e mostrada através de um gráfico. Outro parâmetro utilizado é o coeficiente de correlação (r) obtido entre os valores reais versus previstos pelo modelo. O espectro é dividido em quantos intervalos se desejar.

Além disso, as amostras anômalas detectadas devem ser removidas antes da aplicação do *i*PLS. Os intervalos com os valores de RMSECV que sejam menores que o modelo *full*-PLS são selecionadas para a construção do modelo final (Nørgaard *et al.*, 2000).

Os dois métodos anteriormente citados foram combinados e modificados por Gomes *et al.*, (2013), originando o Algoritmo das Projeções Sucessivas para seleção de Intervalos em Regressão de Mínimos Quadrados Parciais (*i*SPA-PLS, do inglês *Successive Projection Algorithm for Interval selection in Partial Least Squares*). Trata-se de uma técnica que seleciona intervalos em PLS e utiliza o SPA para melhorar os resultados, uma vez que auxilia na ordenação dos intervalos. Assim como o algoritmo de Projeções Sucessivas para seleção de variáveis em Regressão Linear Múltipla (SPA-MLR, do inglês *Successive Projections Algorithm for variable selection in Multiple Linear Regression*), o *i*SPA-PLS pode ser dividido em duas fases. Na fase 1, as k variáveis x_1, x_2, \dots, x_k que correspondem a matriz X_{cal} são divididas em intervalos não sobrepostos, baseado no cálculo do número ótimo de fatores determinados utilizando um modelo *full* PLS. As colunas representativas (coluna com a maior norma dentro de cada um dos w intervalos) são armazenadas em uma matriz W_{cal} ($N_{cal} \times w$), a qual é submetida a projeções sucessivas, gerando a matriz SEL com dimensão $(w - 1) \times w$, diferentemente do SPA-MLR que utiliza a matriz X_{cal} , ou seja, todas as variáveis. A fase 2 é similar ao SPA-MLR, porém as colunas da matriz SEL correspondem aos índices da cadeia dos intervalos. Em seguida, o modelo PLS é construído, utilizando validação cruzada ou validação por série de teste. Por fim, ocorre a avaliação, sendo selecionado o modelo que apresentar o menor valor de RMSECV ou RMSE.

2.3.2.4 Parâmetros de desempenho em calibração multivariada

Para avaliar o desempenho dos modelos multivariados, é recorrente na literatura a utilização de algumas figuras de mérito. Uma delas é a Soma Quadrática dos Erros de Previsão (PRESS, do inglês *Prediction Error Sum of Squares*) (Equação 2.9). O menor valor de PRESS determina o número ideal de fatores ou variáveis latentes que representa a maior capacidade preditiva do modelo.

$$\text{PRESS} = \sum_{i=1}^n [(C_{\text{pred}})_i - (C_{\text{ref}})_i]^2 \quad (2.9)$$

onde $(C_{\text{ref}})_i$ é o valor da concentração obtida pelo método de referência para i -ésima amostra no conjunto de calibração ou predição; $(C_{\text{pred}})_i$ é o valor da concentração predita pelos modelos quimiométricos para i -ésima amostra no conjunto de calibração ou predição (Gomes *et al.*, 2022; Ferreira, 2015).

Outros parâmetros importantes referem-se à Raiz Quadrada do Erro Médio Quadrático da Validação Cruzada (RMSECV, do inglês *Root Mean Square Error of Cross-Validation*) e à Raiz Quadrada do Erro Médio Quadrático da Predição (RMSEP, do inglês *Root Mean Square Error of Prediction*) (Equação 2.10). Trata-se de ferramentas que são capazes de avaliar o erro da validação cruzada ou predição, por meio da diferença entre os valores observados e aqueles obtidos preditos pelo modelo. Esse parâmetro leva em consideração os erros sistemáticos e aleatórios. Sendo assim, o número de amostras deve ser grande o suficiente para refletir todas as fontes de variação do conjunto de calibração. Quanto menor o seu valor, melhor o modelo quimiométrico.

$$\text{RMSECV} = \text{RMSEP} = \sqrt{\frac{1}{N} \sum [(C_{\text{pred}})_i - (C_{\text{ref}})_i]^2} \quad (2.10)$$

onde N é o número de amostras no conjunto de calibração ou predição (Gomes *et al.*, 2022; Ferreira, 2015).

Uma parte do RMSECV e RMSEP refere-se ao Erro Sistemático (Bias) (Equação 2.11) que é a diferença entre o valor obtido e o predito, considerando as flutuações aleatórias nos dados e erros experimentais. O desejável é que seu valor seja mais próximo de zero, indicando a ausência ou a existência de pouco bias.

$$\text{Bias} = \frac{\sum ((C_{\text{ref}})_i - (C_{\text{pred}})_i)}{N} \quad (2.11)$$

Uma maneira de testar a significância do bias é através do teste t , normalmente considerando 95% de confiança. No entanto, para realizar esse teste, necessita-se do

Desvio Padrão dos Erros Validação (SDV, do inglês *Standard Deviation of Validation*) que pode ser obtido através da equação 2.12.

$$SVD = \sqrt{\frac{\sum [(C_{ref})_i - (C_{pred})_i - Bias]^2}{N - 1}} \quad (2.12)$$

A partir do SVD pode-se calcular o valor de t por meio da equação 2.13. O resultado é comparado com o t crítico tabelado, conforme os graus de liberdade. Caso o valor de t calculado for maior que o valor de t crítico, o modelo possui bias.

$$t = \frac{|Bias| \sqrt{N}}{SVD} \quad (2.13)$$

onde t = Teste t para Bias (Fernandes *et al.*, 2011; Almeida, 2017).

A partir do RMSEP pode-se calcular o Erro Relativo de Predição (REP, do inglês *Relative Error of Prediction*) (Equação 2.14) que apresenta o quanto o valor experimental se aproxima daquele estimado pelo modelo.

$$REP = \left(\frac{RMSEP}{\frac{1}{N} \sum (C_{pred})_i} \right) \times 100 \quad (2.14)$$

onde $(C_{pred})_i$ é o valor da concentração predita pelos modelos quimiométricos para i -ésima amostra no conjunto predição; e N é o número de amostras no conjunto de predição (Gomes *et al.*, 2022).

O coeficiente de correlação (r) (Equação 2.17) é outra figura de mérito usada na calibração multivariada, para mensurar a relação entre a matriz espectral e os vetores respostas. Seu valor varia de -1 a 1, de maneira que quanto mais próximo de 1, maior é a relação existente entre as matrizes X e Y , de maneira positiva, ou seja, quando uma aumenta a outra também aumenta, enquanto quando o valor de aproxima de -1 também há correlação, porém negativamente, isto é, quanto uma variável aumenta a outra diminui.

$$r = \frac{1}{N - 1} \sum \left[\frac{(C_{ref})_i - (\bar{C}_{ref})}{s_{C_{ref}}} \right] \left[\frac{(C_{pred})_i - (\bar{C}_{pred})}{s_{C_{pred}}} \right] \quad (2.15)$$

onde \bar{C}_{ref} e \bar{C}_{pred} são os valores médios de C_{ref} e C_{pred} ; $s_{C_{ref}}$ e $s_{C_{pred}}$ são os valores de desvio padrão de C_{ref} e C_{pred} (Devianti *et al.*, 2022).

A Relação entre Desempenho e Desvio (RPD, do inglês *Ratio Performance to Deviation*) também pode ser utilizado para avaliar os modelos quimiométricos, podendo ser calculado conforme a Equação 2.16 (Chin, 1998).

$$RPD = \frac{s_{C_{ref}}}{RMSEP} \quad (2.16)$$

De maneira geral, os modelos quimiométricos com RPD maiores que 2,5 indicam uma boa capacidade preditiva, enquanto aqueles que possuem RPD entre 2,0 e 2,5 possuem uma predição razoável e pode ser melhorada com outras técnicas de modelagem, já os que resultam em um RPD menor que 2,0 a capacidade preditiva do modelo é classificada como ruim (Włodarska *et al.*, 2021).

Uma outra forma de avaliar os modelos de regressão é por meio do teste da Região Elíptica de Confiança Conjunta (EJCR, do inglês *Elliptical Joint Confidence Region*) (Equação 2.17). Nesse método, realiza-se uma regressão linear dos resultados do gráfico da concentração real versus a concentração prevista, utilizando os Mínimos Quadrados Ordinários (OLS, do inglês *Ordinary Least Squares*), o qual fornece uma reta com intercepto próximo de 0 e a inclinação próxima de 1. O ponto teórico (0, 1) é testado nos limites da EJCR e se o contiver, conclui-se que não existe diferença significativa entre as concentrações reais e aquelas previstas pelo modelo.

$$EJCR = N(a - \hat{a})^2 + 2 \left[\sum (C_{ref,i}) \right] (a - \hat{a})(b - \hat{b}) + \left[\sum (C_{ref,i})^2 \right] (b - \hat{b})^2 = 2 S_{C_{pred}/C_{ref}}^2 F_{(\alpha, N-2)} \quad (2.17)$$

Onde $S_{C_{pred}/C_{ref}}$ é a variância residual; F é o valor crítico; α é o nível de confiança; e $(N - 2)$ são os graus de liberdade (Riu; Rius, 1996; Lyra, 2012; Gomes *et al.*, 2022).

Assim como a calibração multivariada, a classificação multivariada também vem sendo empregada na literatura, principalmente técnicas de reconhecimento de padrões, as quais serão discutidas no tópico seguinte.

2.3.3 Reconhecimento de padrões

Os métodos de reconhecimento de padrões visam identificar semelhanças e diferenças entre os conjuntos de amostras que passaram por algum tipo de análise. As técnicas de reconhecimento de padrões podem ser de dois tipos: os métodos supervisionados e não supervisionados.

A técnica de reconhecimento de padrões não supervisionado é utilizada para realizar uma análise exploratória, dividindo no espaço o conjunto de dados conforme as suas diferenças e similaridades, sem que haja necessidade de rotular as amostras (Brereton, 2015). Dentre os principais algoritmos que podem ser utilizados para esta finalidade pode-se citar a análise por agrupamento hierárquico (HCA, do inglês: *Hierarchical Cluster Analysis*) e análise de componente principal (PCA, do inglês: *Principal Component Analysis*), sendo esta última utilizada nesse trabalho de tese.

Por outro lado, nas abordagens supervisionadas as amostras do conjunto amostral são rotuladas em classes. Sendo assim, uma nova amostra desconhecida pode ser classificada pertencente a uma, várias ou nenhuma das classes conhecidas (Brereton, 2015). Três métodos se destacam com grande aplicabilidade em problemas de classificação: modelagem independente e flexível por analogia de classes direcionado pelos dados (DD-SIMCA, do inglês: *Data-Driven – Soft Independent Modelling of Class Analogy*), análise discriminante linear (LDA, do inglês: *Linear Discriminant Analysis*) e a análise discriminante pelos mínimos quadrados parciais (PLS-DA, do inglês: *Partial Least Squares-Discriminant Analysis*). Porém, apenas o primeiro foi empregado neste trabalho.

2.3.3.1 Análise por Componentes Principais (PCA)

A análise por componentes principais é considerada uma das ferramentas mais utilizadas para realizar a análise exploratória de dados. A PCA é uma técnica de reconhecimento de padrões não supervisionada usada para projetar os dados multivariados em um espaço menor, reduzindo sua dimensionalidade, porém sem afetar as relações entre as amostras, uma vez que a maior parte da variância dos dados originais são preservadas. Utilizando esse método é possível identificar as diferenças existentes entre as variáveis e as relações presentes entre as amostras ou objetos do conjunto de dados (Ferreira, 2015; Nóbrega, 2021).

Em linhas gerais, a PCA transforma as variáveis originais, as quais estão possivelmente correlacionadas, em variáveis ortogonais que são independentes entre si. Essas variáveis são denominadas de componentes principais (PCs, do inglês *Principal Components*). Cabe ressaltar que essa transformação nos dados leva em consideração que as primeiras componentes principais resultam em uma maior variância possível, de modo que a primeira componente principal (PC1) é direcionada em relação a maior variância dos dados originais, já a segunda componente (PC2) representa a segunda maior variância, sendo ortogonal a PC1 e assim sucessivamente (Jolliffe, Cadima, 2016; Ferreira, 2015; Nóbrega, 2021).

Em termos matemáticos, a PCA decompõe a matriz pré-processada ou original (X) em duas matrizes: scores (T) e *loadings* (P). A primeira refere-se as coordenadas das amostras relativas aos eixos de cada componente principal (PC), enquanto a segunda está relacionada com contribuição de cada variável original de cada eixo da componente principal (PC), ou seja, o cosseno do ângulo entre o eixo variável e o eixo PC e uma matriz de resíduo (E) não explicado pelo modelo (Equação 2.18) (Jolliffe, Cadima, 2016; Beebe *et al.*, 1998).

$$X = TP^T + E \quad (2.18)$$

2.3.3.2 Modelagem independente e flexível por analogia de classes direcionada pelos dados (DD-SIMCA)

A Modelagem independente e flexível por analogia de classes direcionada pelos dados (DD-SIMCA) é umas das técnicas de reconhecimento de padrões supervisionadas de uma única classe (*one-class*) empregada como uma ferramenta de detecção de padrões para a verificação de autenticidade de alimentos, medicamentos, combustíveis e outros. Nesse método, apenas a classe "alvo" é modelada, e qualquer amostra que não se enquadre nessa classe (alvo) é considerada extrema, adulterada ou falsificada, conforme a abordagem dos dados analisados (Nóbrega, 2021).

O DD-SIMCA (Zontov *et al.*, 2017) é uma adaptação do SIMCA tradicional. A diferença está relacionada na forma de construção da área de aceitação de cada classe. Nessa abordagem, é realizada uma análise estatística nas distâncias dos

scores e nas distâncias ortogonais. Esses valores de distância devem ser estimados como quaisquer outros dados adquiridos, sendo as suas distribuições estimadas por meio de um método de dados orientados.

De maneira geral, o DD-SIMCA começa decompondo a matriz de treinamento utilizando a abordagem PCA, já discutida na seção anterior. Posteriormente, as distâncias dos scores e a distância ortogonal são calculadas para cada amostra pertencente ao conjunto de treinamento levando em consideração os resultados obtidos na PCA. Por fim, uma região de aceitação para classe é definida com base nos cálculos das distâncias dos scores, levando em consideração as taxas de erro de falso negativo (erro tipo I) e verdadeiro negativo (erro tipo II). Com esta modelagem da classe alvo bem estabelecida, uma última etapa de classificação do conjunto de teste é realizada, a fim de avaliar o desempenho do modelo construído (Marcelino Netto *et al.*, 2023).

O desempenho dos modelos construídos utilizando o DD-SIMCA podem ser avaliados utilizando os valores de sensibilidade, que mede o quão bem o modelo classifica as amostras puras como pura; especificidade, a qual mede a precisão do modelo em lidar com amostras falso positivas; e acurácia que indica a capacidade do modelo construído de identificar amostras verdadeiras e falsas levando em consideração todas as amostras. Tais cálculos são realizados utilizando as equações abaixo (Marcelino Netto *et al.*, 2023; Nóbrega, 2021; Nóbrega *et al.*, 2023):

$$\text{Sensibilidade} = \left(\frac{TP}{TP + FN} \right) \times 100 \quad (2.19)$$

$$\text{Especificidade} = \left(\frac{TN}{TN + FP} \right) \times 100 \quad (2.20)$$

$$\text{Acurácia} = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \times 100 \quad (2.21)$$

Onde: TP é o número de verdadeiros positivos, TN é o número de verdadeiros negativos, FN é o número de falsos negativos e FP é o número de falsos positivos.

2.3.4 Fusão de dados

Em linhas gerais, a fusão de dados é uma estratégia que tem por objetivo fundir blocos de dados originados de diferentes técnicas analíticas, a fim de elaborar um único modelo. Este procedimento visa a obtenção de um panorama mais completo e preciso acerca do objetivo de estudo, melhorando a confiabilidade dos modelos. A fusão de dados teve seu início na década de 70 com aplicações no campo militar. Na época, múltiplas fontes de informações eram utilizadas de forma integrada, a fim de identificar alvos, tais como aeronaves, formações militares, mísseis, além de avaliar seu potencial de ameaça. Na química, essa abordagem tem sido cada vez mais utilizada, possibilitando a integração de dados de várias técnicas, como cromatografia, espectrometria de massas, ressonância magnética nuclear (RMN) e espectroscopia no ultravioleta visível (UV-Vis) ou no infravermelho próximo (NIR) para análises de alimentos (Li *et al.*, 2020; Zhu *et al.*, 2021; Khaleghi *et al.*, 2013).

A fusão de dados pode ser realizada em diferentes níveis: baixo, médio e alto (Figura 2.6). No nível baixo, o sinal analítico das diferentes fontes, após pré-processamento individuais, é concatenado em uma única matriz. Assim, essa informação é utilizada para calcular um único modelo que descreve o parâmetro a ser estudado. Apesar de ser o nível de fusão de dados mais simples, essa estratégia possui um alto número de variáveis que podem ser irrelevantes para o modelo (Li *et al.*, 2020; Zhu *et al.*, 2021).

Na fusão de dados de nível médio, inicialmente extrai-se, separadamente, algumas informações relevantes de cada fonte e logo após os dados são concatenados em uma única matriz. A abordagem mais comum é fundir as variáveis selecionadas, variáveis latentes ou componentes principais da matriz do sinal analítico de cada instrumento. Essa técnica é mais eficiente do que a fusão de dados de baixo nível, visto que apenas a variação relevante é utilizada para a construção do modelo final (Li *et al.*, 2020; Zhu *et al.*, 2021).

Na fusão de nível alto, também chamada de nível de decisão, os modelos são calculados separadamente e os resultados são integrados em uma única resposta, a fim de obter uma decisão de conjunto. O principal desafio é construir modelos que funcionem adequadamente para cada conjunto de dados de maneira que a sua combinação tenha um desempenho melhor do que os modelos individuais (Li *et al.*, 2020; Zhu *et al.*, 2021).

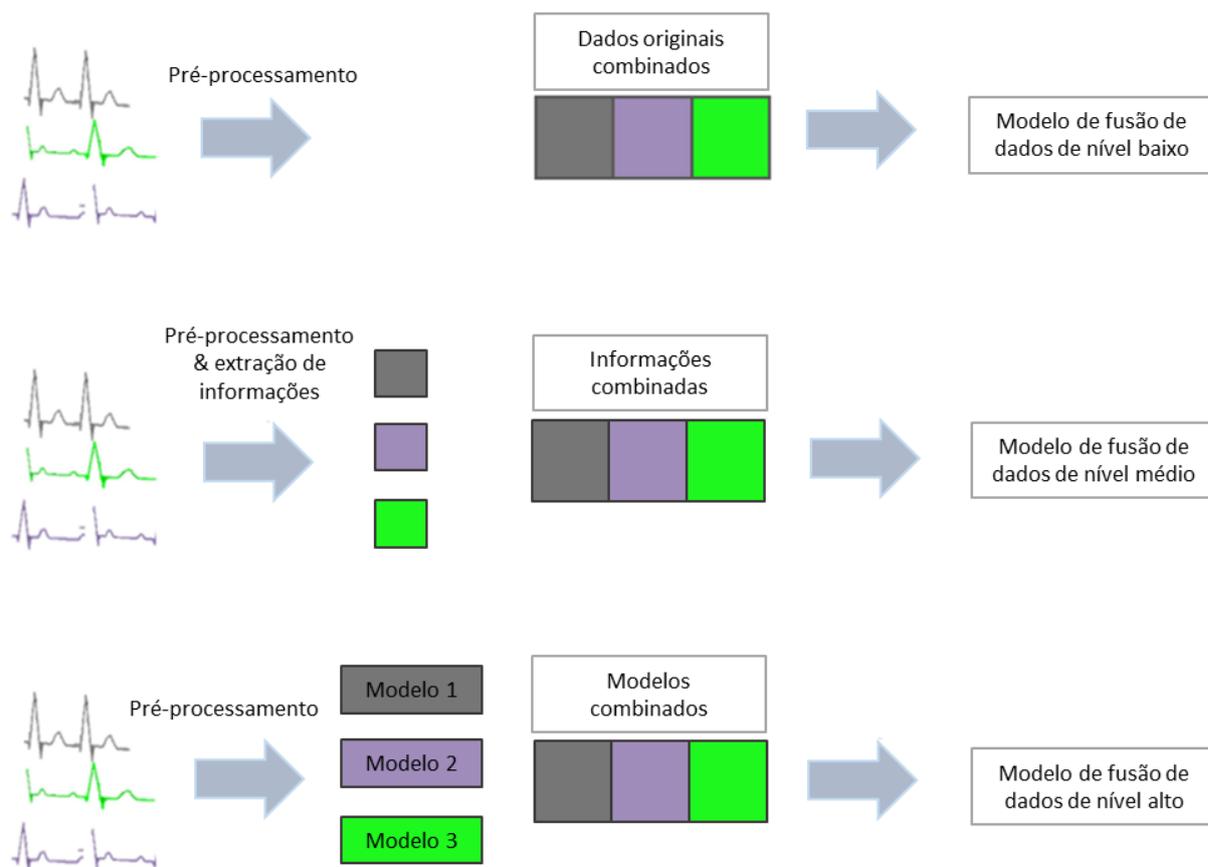


Figura 2.6 - Resumo gráfico dos três tipos de fusão de dados: nível baixo, nível médio e nível alto. Adaptado de [Hayes et al., 2023](#).

De maneira geral, o melhor nível deve ser escolhido após a obtenção do conjunto de dados das diferentes fontes. Essa escolha depende de vários fatores, como técnicas instrumentais selecionadas, natureza dos dados, volume de dados gerados, entre outros. As técnicas de fusão de dados podem melhorar os resultados dos modelos gerados. O principal objetivo da fusão de dados é aumentar a sinergia dos dados fundidos, isto é, utilizar as técnicas com informações complementares, a fim de obter melhores modelos quimiométricos ([Li et al., 2020](#); [Zhu et al., 2021](#)).

Outrossim, [Hayes et al., 2023](#) relatam que os principais desafios da fusão de dados estão relacionados em encontrar uma técnica apropriada, a fim de integrar os dados heterogêneos de múltiplos sistemas complexos. Para os autores, combinar dados de vários instrumentos, especialmente os originados do NIR, pode levar a inclusão de ruídos nos modelos, prejudicando sua capacidade preditiva. Assim, diferentes técnicas de fusão de dados são avaliadas empiricamente, sendo os resultados dos modelos fundidos comparados aos individuais, a fim de determinar a técnica ideal para o desenvolvimento do modelo.



2.4 Ovos de galinha em pó

2.4.1 Considerações gerais

O ovo é um organismo unicelular desenvolvido no ovário ou oviduto dos animais ovíparos, sendo constituído de protoplasma, envoltórios e vesículas germinativas (Philippi, 2014). O termo é empregado genericamente para identificar aqueles originados da galinha (*Gallus gallus domesticus*), quando oriundos de outros ovíparos, a espécie é identificada logo após o nome “ovo” (Brasil, 1990).

A avicultura é uma atividade muito antiga, tendo seus primeiros registros há cerca de 8 mil anos com a domesticação da espécie *Gallus gallus* em países da Ásia, como China e Índia. Foi a partir desse período que as galinhas passaram pela Mesopotâmia e Grécia, através das tribos nômades, e se disseminaram por todo continente europeu. No Brasil, os registros revelam sua chegada com os navios portugueses na época da chegada dos portugueses ao Brasil (Santos Filho *et al.*, 2011).

De acordo com Carvalho (2012), a domesticação das galinhas contribuiu para subsistência do homem, uma vez que, até então, alimentavam-se de ovos encontrados em aves silvestres. No entanto, seu consumo ocorreu de maneira muito insignificante até meados do V a.C., momento que a espécie chegou nas regiões da Grécia e Itália, tornando o alimento um ingrediente indispensável.

O ovo é um alimento muito consumido na alimentação mundial, especialmente no Brasil que vem ocupando um lugar de destaque na escala produtiva. De acordo com dados da Associação Brasileira de Proteína Animal (ABPA), o país produziu mais de 4,5 bilhões de dúzias de ovos, resultando em um valor bruto de US\$ 17,734 bilhões para o setor em 2021. Desse total, mais de 99% são destinados para o consumo interno. No entanto, os dados da ABPA relevam um aumento de, aproximadamente, 80% em exportações, considerando os ovos *in natura* e industrializados, totalizando US\$ 18,054 milhões em receitas. Dentre os estados que mais exportaram destacam-se o Mato Grosso, Minas Gerais, Rio Grande do Sul e São Paulo. Tais exportações foram destinadas, principalmente, para os países Emirados Árabes Unidos, Japão, Catar e Estados Unidos (ABPA, 2022).

No Brasil, as galinhas podem ser classificadas em quatro diferentes tipos de manejo: *free range*, orgânico, caipira (colonial) e granja (Amaral *et al.*, 2016). Cada

manejo origina ovos com características e nutrientes específicos, tornando um dos alimentos mais completos para alimentação humana, uma vez que é uma rica fonte de proteína de alto valor biológico, já que é formado por todos os aminoácidos essenciais; vitaminas A, B, D, E, K, gorduras; minerais, principalmente, ferro, cálcio, fósforo, selênio e zinco; carotenoides, como zeaxantina e luteína. Sendo assim, é considerado como um ingrediente multifuncional, podendo ser incrementado em vários alimentos, a fim de conferir cor, aroma e sabor, além de possuir propriedades funcionais, como viscosidade, formação de espuma e gelificação (clara) e emulsificação (gema). Por possuir um baixo valor calórico, seus nutrientes previnem doenças, uma vez que agem em atividades antivirais, antibacterianas e na modulação do sistema imunológico. Devido à sua qualidade, alto poder nutritivo e custo-benefício quando comparado com demais proteínas de origem animal, torna-o um relevante alimento no combate à fome (Amaral *et al.*, 2016; Walsh, 2014; Abreu; Spinelli, 2014).

2.4.2 Composição anatômica do ovo

Em linhas gerais, os ovos são compostos anatomicamente por três partes principais: casca, clara e gema (**Figura 2.7**)

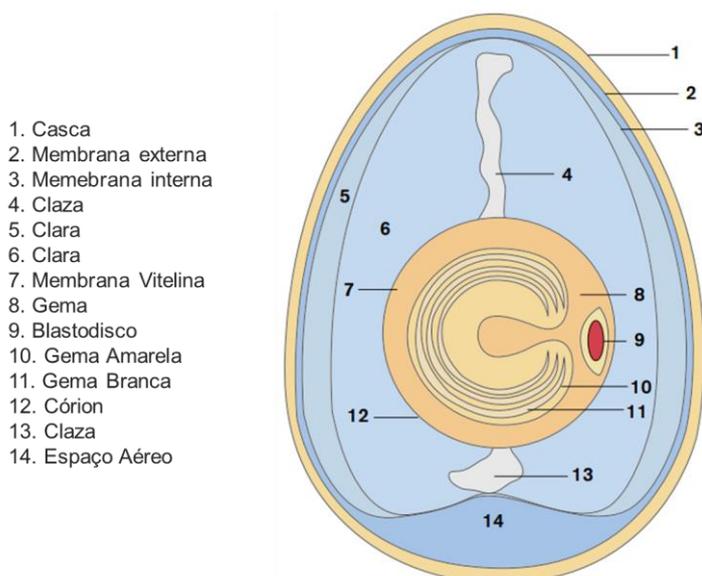


Figura 2.7 - Estrutura anatômica do ovo. Fonte: Moreira (2016).

A casca do ovo é constituída por uma grande quantidade de carbonato de cálcio e possui uma estrutura porosa que permite a troca de gases. Esse componente serve para protegê-lo contra danos físicos e possíveis contaminantes, uma vez que é



revestida internamente por duas membranas (interna e externa) que funciona como uma barreira, a fim de evitar a perda de água e que bactérias ou outros micro-organismos penetrem para seu interior (Amaral *et al.*, 2016).

A coloração da casca possui caráter genético e depende da raça e linhagem da espécie, alimentação e idade. Os genes controlam sua pigmentação por meio da deposição de porfirinas, resultando a casca na cor branca ou de cor avermelhada, que apresenta uma maior resistência. No entanto, a cor da casca não afeta a qualidade do alimento, bem como suas propriedades de cocção (calor) e os nutrientes existentes (Bertechini, 2003). A casca equivale de 8,0 a 11,0% da massa do ovo, sendo composta, principalmente, por carbonato de cálcio (94,0%), carbonato de magnésio (1,4%), glicoproteínas (3,0%), mucoproteínas, mucopolissacarídeos e colágeno (Ornellas, 2001).

Já a clara é constituída por um material semissólido gelatinoso e representa cerca de 57% da massa total do ovo, sendo composta, principalmente, por água, sais minerais, glicose, lipídeos, além da mistura de 10% de proteínas, como ovalbumina, conalbumina ou ovotransferrina, avidina, ovomucoide, ovomucina, flavoproteína, ovoidinibidor, cistatina e lisozima (Quadro 2.1). As três primeiras correspondem cerca de 77% do total de proteínas presentes na clara e são responsáveis por seu aspecto gelatinoso. No seu interior, encontram-se dois cordões proteicos que formam uma estrutura denominada de calazas, a qual mantém a gema centralizada, além de protegê-la contra impactos (Ramos, 2008; Carvalho, 2012).

Quadro 2.1 - Proteínas da clara e suas características.

PROTEÍNAS DA CLARA	CARACTERÍSTICAS
Ovoalbumina	É uma glicofosfoproteína constituída de 385 aminoácidos, principalmente lisina e triptofano. Possui tiol e um grupo dissulfeto em sua estrutura e corresponde a 54% do total de proteína da clara. É relativamente fácil de ser desnaturada, por meio da agitação ou batimento da solução, devido ao desdobraimento e agregação de moléculas de proteínas.
Conalbumina ou ovotransferrina	Consiste em uma cadeia peptídica que contém um oligossacarídeo, composta por quatro manose e oito N-acetilglucosamina. Corresponde a 12% das proteínas da clara, sendo responsável por inibir crescimento de micro-organismo. Forma complexos, principalmente com o ferro, originando a coloração avermelhada nos ovoprodutos.

Ovomucoide	Corresponde a 11% do total de proteínas existentes na clara, sendo constituída por 3 oligossacarídeos ligados através da asparagina. Possui 9 dissulfetos que são responsáveis por sua resistência a coagulação e inibe a atividade da enzima proteolítica (tripsina).
Lisozima	Composta por uma cadeia peptídica com 129 aminoácidos e 4 ligações de dissulfetos, representa cerca de 3,4% do total de proteínas da clara. Atua no rompimento da parede de celular de bactérias Gram-positivas, a partir da mureína, N-acetil-glucosamina e ácido N-acetilmurâmico, que impede a entrada de bactérias.
Ovomucina	É uma glicoproteína que pode apresentar estruturas fibrilares, responsáveis pelo aumento da viscosidade da gema. Corresponde a 3,5% da proteína da clara, sendo estável ao calor. Além disso, pode formar um complexo insolúvel com água e lisozima.
Flavoproteína	Realiza uma forte ligação com a riboflavina, facilitando a transferência de uma coenzima do soro sangue para o ovo, além de representar um total de 0,8% das proteínas existentes na clara.
Ovoinibidor	Representa 1,4% das proteínas da clara, sendo responsável por inibir a proteinase, como tripsina, quimotripsina e outras de origem microbiana.
Avidina	É uma glicoproteína que pode realizar ligação com a biotina (vitamina), formando um complexo avidina-biotina que impede a absorção de proteína. No entanto, representa baixo teor (0,05%) das proteínas existentes na clara.
Cistatina	Consistem em uma cadeia peptídica com 120 aminoácidos e inibe a endopeptidases, como ficina e papaína, possuindo um baixo teor no ovo (0,05%).

Adaptado de Belitz; Grosch; Schieberle (2009).

Em sua forma natural, a clara é translúcida, ficando opaca quando exposta a cocção, bem como devido à formação de espuma em decorrência da entrada de ar na estrutura proteica. A clara pode apresentar uma cor amarela ou esverdeada que está relacionada à quantidade de riboflavina presente no alimento (Araújo *et. al.*, 2014).

A gema do ovo representa cerca de 32% da massa total do ovo, sendo uma emulsão de gordura em água. É composta por 65% de lipídios, 31% de proteínas e 4% de carboidratos, vitaminas, minerais e fosfolipídios (lecitinas e esteróis). É constituída de carotenos e riboflavina, além de fosvitina (4%), livitina (10%), Lipoproteína de Baixa Densidade (LDL, do inglês *Low Density Lipoprotein*) que totaliza



68% e Lipoproteína de Alta Densidade (HDL, do inglês *High Density Lipoprotein*), que representa 16% do total da gema (**Quadro 2.2**) (Xiao *et al.*, 2020; Chang *et al.*, 2018; Belitz; Grosch; Schieberle, 2009).

Quadro 2.2 - Proteínas da gema e suas características.

PROTEÍNAS DA GEMA	CARACTERÍSTICAS
Lipovitelinina	É um fosfolípideo que apresenta duas estruturas: α -lipovitelinina que apresenta duas cadeias polipeptídicas e β -lipovitelinina que possui apenas uma cadeia. Sua estrutura é composta por vitelinas que são ligadas covalentemente a oligossacarídeos, constituídos por manose, galactose, glucosamina e ácido siálico. É estável ao calor e forma monômeros com o aumento do pH ácido.
Fosvitina	É uma glicofosfoproteína que possui uma alta quantidade de ácido fosfórico ligado a serina. É estável ao calor e forma complexos com cátions multivalentes, como o ferro, diminuindo a sua biodisponibilidade.
Livitina	É solúvel em água, sendo constituída por α -, β - e γ -livetinas.

Adaptado de Belitz; Grosch; Schieberle (2009).

No que se refere à lecitina, sua ação é emulsificante, em decorrência da sua capacidade de combinar compostos hidrofóbicos e hidrofílicos, oferecendo ao ovo a propriedade estabilizadora de emulsões (Domene, 2018). A lecitina é um fosfolípido composto principalmente por glicerofosfolípideo: fosfatidilcolina, fosfatidiletanolamina e o fosfatidilinositol, sendo a primeira a mais importante, uma vez que contribui para os processos de metabolismo do organismo, por meio da digestão e absorção de lipídeos (Dukes; Reece, 2006).

Os fosfolípídios são uma classe de lipídios que compõem as membranas das células. Em sua composição encontra-se o fósforo cujo elemento representa cerca de 1% da massa corpórea de uma pessoa, sendo primordial para vida humana, uma vez que colabora com o desenvolvimento da estrutura óssea. Esse mineral contribui com a digestão de carboidratos e gorduras, além de ajudar o corpo na produção de ATP, responsável por armazenar energia no corpo. Além disso, é um componente importante para o RNA, DNA e atua juntamente com as vitaminas do complexo B, colaborando com a função renal, contrações musculares, batimentos cardíacos e sinalização do sistema nervoso (Bridges, 2021; Cooper; Gosnell, 2018; Mancini *et al.*, 2018).

De maneira geral, os fosfolípidios são produzidos a partir de uma via metabólica não linear e compartilham da mesma rota sintética dos carotenoides, conforme a **Figura 2.8**.

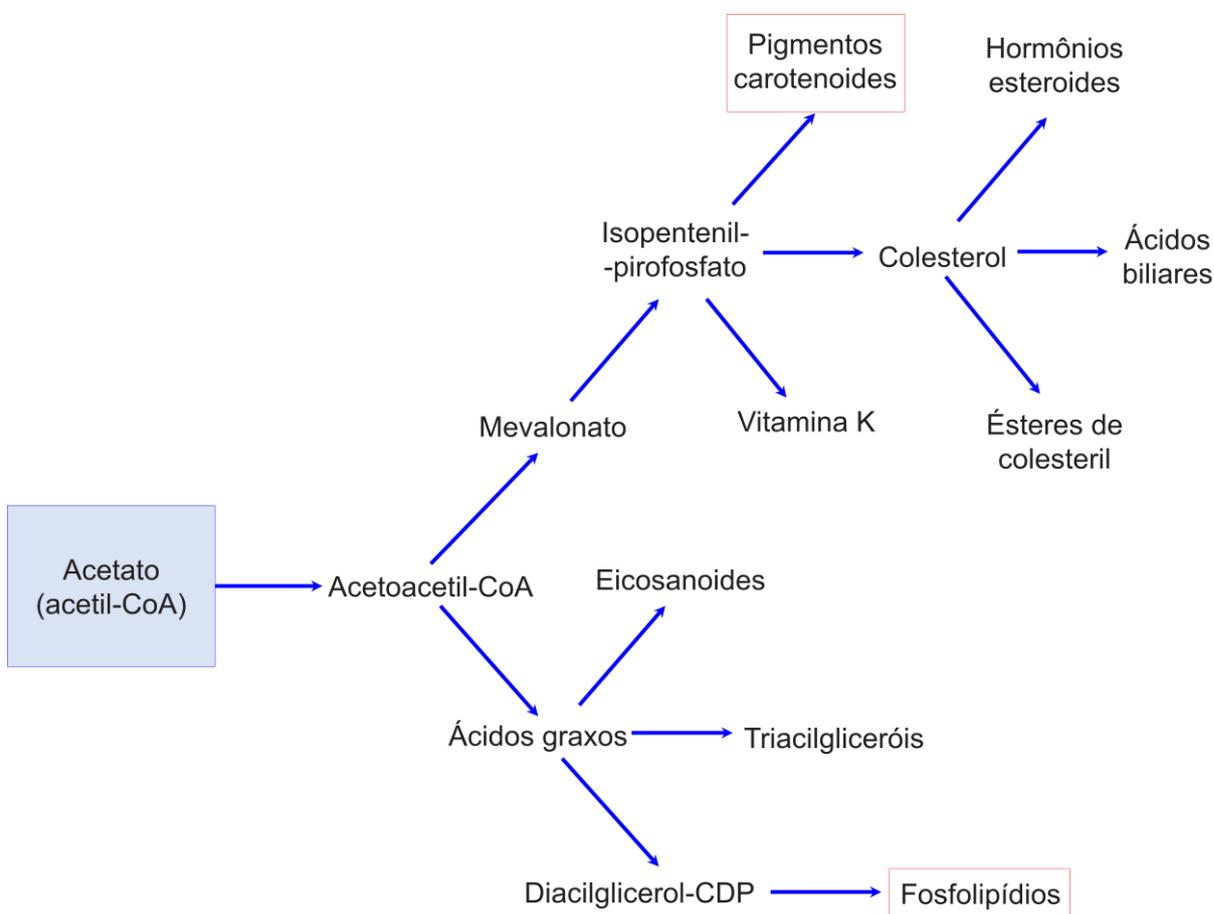


Figura 2.8 - Produção dos fosfolípidios, colesterol e pigmentos carotenoides via metabólica não linear com anabolismo divergente. Fonte: Nelson Lehninger; Cox, 2022.

Os carotenoides são compostos antioxidantes responsáveis por proteger as células de estresse oxidativo, colaborando com a melhora do sistema imunológico. As galinhas não são capazes de produzi-los, podendo apresentar a deficiência de vitamina A. Sendo assim, esses compostos são adicionados na alimentação, seja de maneira natural ou sintética. Dentre os aditivos naturais que podem ser utilizados, encontram-se os extratos vegetais, como flor de calêndula ou pimenta vermelha seca. Essa adição é responsável pela coloração amarelo-alaranjada do ovo. Os principais carotenoides presentes nesse tipo de alimento são as xantofilas e β carotenos (Moreno *et al.*, 2020).

Na **Figura 2.9**, visualiza-se a estrutura química das proteínas, lecitina e colesterol, principais constituintes presentes nos ovos de galinha em pó estudados nesta tese.

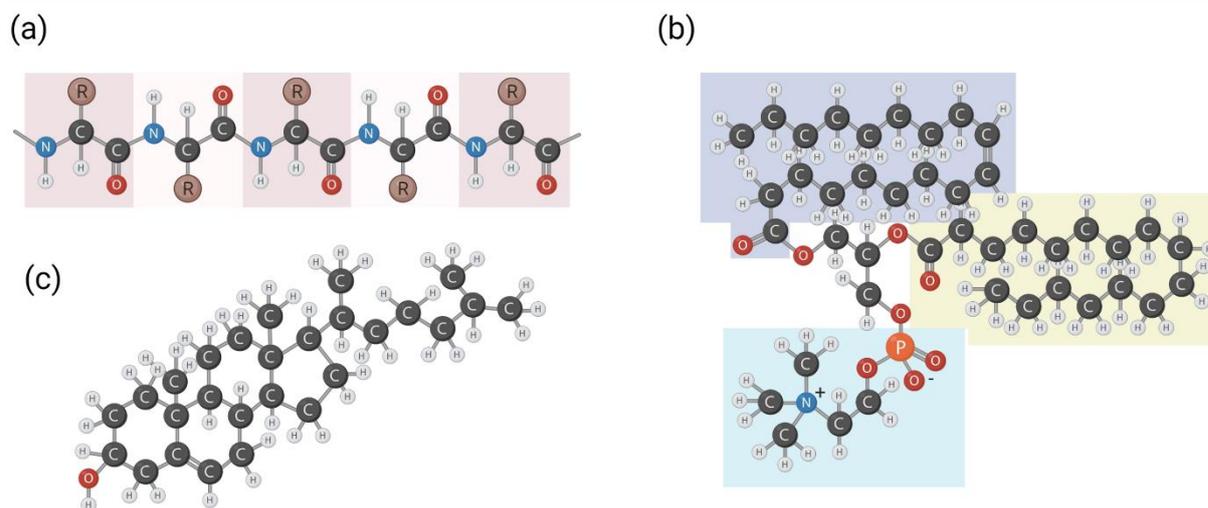


Figura 2.9 - Estrutura química (a) das proteínas, (b) lecitina e (c) colesterol. Elaborado no biorender.com

2.4.3 Classificação dos ovos

No Brasil, os ovos só podem ser disponibilizados para o consumo humano quando submetidos à inspeção e classificação, conforme o decreto Nº 9.013 de 29 de março de 2017. Sendo assim, quanto às características qualitativas apresentadas, os ovos podem ser classificados nas categorias A e B. Essa etapa de classificação pode ocorrer de duas maneiras: manual ou mecânica ([Brasil, 2017; 2022](#)).

Os ovos da categoria A devem apresentar casca e cutícula de forma normal, lisa, limpa e intacta; a câmara de ar deve possuir altura não superior a 6 mm (seis milímetros) e ser imóvel; a gema deve, a partir da ovoscopia, constituir-se em forma de sombra e quando mover-se em caso de rotação, deve voltar para sua posição central; a clara deve ser límpida e translúcida, não apresentar manchas ou turvação e suas calazas deve estar intactas; e a cicatrícula deve possuir um desenvolvimento imperceptível ([Brasil, 2017](#)).

Os ovos são classificados da categoria B quando forem considerados inócuos, sem que sejam classificados na categoria A; apresentar pequenas manchas de sangue na clara e na gema; e devem ser provenientes de estabelecimentos avícolas,



os quais em seu processo de reprodução não foram submetidos à incubação. Essa categoria é destinada exclusivamente à industrialização (Brasil, 2017). Quanto a massa, podem ser classificados em jumbo, extra, grande, médio e pequeno (Quadro 2.3).

Quadro 2.3 - Classificação do ovo quanto ao seu massa.

Classificação do ovo	Massa unitário (em gramas)
Jumbo	Acima de 66
Extra	Entre 60 e 65,99
Grande	Entre 55 e 59,99
Médio	Entre 50 e 54,99
Pequeno	Abaixo de 50

Fonte: Brasil (2022).

De acordo com a legislação, para separar e avaliar os ovos por categoria e massa deve-se utilizar inicialmente a ovoscopia seguida da classificação. A primeira deve garantir a avaliação das características internas e externas do ovo, a fim de classificá-los conforme as suas características qualitativas (Brasil, 2022).

As características dos ovos naturais são importantes para a fabricação dos ovoprodutos. Os derivados dos ovos são interessantes, uma vez que possuem preço acessível; menor massa; facilita o transporte; proporciona a segurança microbiológica e um menor espaço para armazenamento, bem como aumenta a vida útil desse tipo de alimento. (Philippi, 2014; Abreu; Spinelli, 2014; Amaral *et al.*, 2016).

2.4.4 Ovoprodutos

De acordo com a legislação brasileira, os derivados de ovos são aqueles obtidos a partir dos diferentes componentes ou misturas de ovos, depois da eliminação da casca e membranas. Esses alimentos podem ser líquidos, concentrados, pasteurizados, desidratados, cristalizados, resfriados, congelados, ultracongelados, coagulados ou que apresentem outras formas utilizadas como alimento (Brasil, 2017; 2022).

A indústria dos ovoprodutos surgiu nos Estados Unidos por volta de 1900, sendo impulsionada pela Segunda Guerra Mundial (Stadelman; Cotterill, 2017). Inicialmente planejada para reduzir as perdas na produção, tornou-se uma alternativa

para os produtores, devido ao maior prazo de validade e a estabilidade do preço quando comparado com o ovo *in natura* (Amaral *et al.*, 2016).

Dentre os alimentos industrializados mais interessante, encontram-se o ovo desidratado, o qual devido às suas altas funcionalidades, como formação de espuma, coagulação e emulsificação, é um ingrediente essencial para preparação de vários alimentos, como massas, biscoitos, pães, tortas, empanados, confeitaria, sorvete, panetone e maionese (Gésan-Guiziu, 2013; Philippi, 2014).

2.4.4.1 Ovo desidratado

Para o Ministério da Agricultura, Pecuária e Abastecimento (MAPA), o ovo desidratado é aquele resultante do processo de desidratação do ovo integral pasteurizado, obtido por processo tecnológico específico e que resulte em um produto com umidade igual ou menor que 5% (m/m) e que mantém as características nutricionais do ovo natural (Brasil, 2022).

De maneira geral, a secagem é um dos métodos mais antigos utilizados para a conservação dos alimentos. Essa técnica pode ser natural ou artificial, sendo usado para remover a umidade de um produto por meio do fornecimento de energia que favorece a transferência simultânea de calor e de massa (Onwude *et al.*, 2017).

A secagem natural ocorre utilizando a energia do sol e do vento, a fim de remover a água do produto. É uma técnica econômica, já que não há gastos de utilização de equipamentos sofisticados. No entanto, além de necessitar grandes áreas para exposição dos alimentos, é demorada e depende do clima, afetando sua homogeneidade (Nunes, 2016; Leonardi; Azevedo, 2018). Já a secagem artificial baseia-se na utilização de artifícios para aumentar a velocidade de secagem utilizando equipamentos chamados de secadores. Para isso, realiza-se o controle de temperatura, umidade e corrente do ar (Di Domenico; Conrad, 2015). Dentre os secadores destacam-se: secador de túnel, secador de esteira, secadores de bandeja, secador de leite fluidizado, leite de jorro, secador por atomização (*spray dryer*), secador de tambor rotativo e liofilizador (Celestino, 2010).

O liofilizador, também conhecido como "*freezedrier*", é o equipamento de secagem utilizado para realizar o processo denominado de liofilização. Essa técnica é baseada na sublimação da água contida no alimento, utilizando baixas temperaturas e pressão inferior ao ponto triplice da água. Inicialmente, congela-se o alimento e, no

lioofilizador, sob vácuo (1 mmHg) ocorre a sua desidratação (Di Domenico; Conrad, 2015; Brasil, 2022).

A liofilização foi utilizada pela primeira vez nas indústrias alimentícias por volta de 1940, constituindo-se como uma importante técnica para desidratação de alimentos, uma vez que esse não é exposto a altas temperaturas, o que permite a manutenção das características nutricionais do alimento natural, diferentemente das secagens convencionais. Existem diversos tipos de liofilizadores, os quais se diferenciam na forma como fornecem o calor para a superfície do alimento. O equipamento é constituído por cinco principais componentes: câmara de vácuo, sistema de aquecimento, sistema de condensação, sistema de vácuo e controles (Figura 2.10) (Tonon *et al.*, 2016) e três etapas são empregadas: congelamento, secagem primária ou sublimação e secagem secundária ou dessorção.

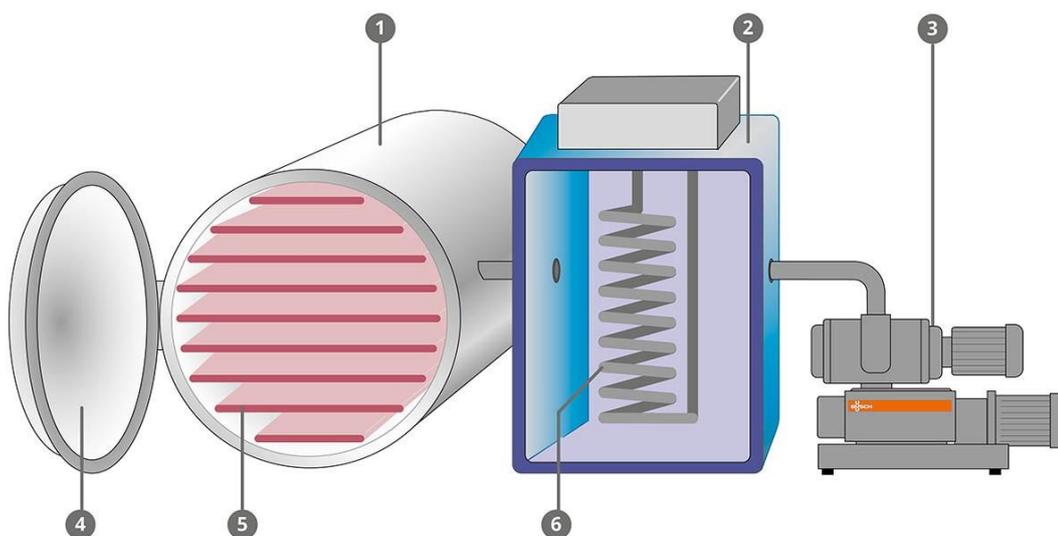


Figura 2.10 - Esquema de um liofilizador: (1) Câmara de secagem; (2) Condensador; (3) Sistema de vácuo; (4) tampa da câmara de secagem; (5) prateleira; e (6) Serpentina de resfriamento Fonte: <https://www.buschvacuum.com/in/en/news-media/the-role-of-vacuum-technology-in-freeze-drying.html>

O congelamento é o primeiro estágio do processo de liofilização, sendo responsável por converter de 65 a 90% de água aquosa em sólida. Nele o alimento é congelado a temperaturas entre $-23\text{ }^{\circ}\text{C}$ a $-40\text{ }^{\circ}\text{C}$. Esse procedimento é importante, uma vez que define o tamanho dos cristais e poros, além da distribuição da superfície liofilizada, sendo responsável pelo sucesso das demais etapas (Anandharamakrishnan, 2017; Rey, 2016). Em escala industrial, essa fase é realizada



dentro do liofilizador, porém em equipamentos laboratoriais, as amostras são congeladas em outros equipamentos, como o *freezer* (Dziki *et al.*, 2018).

Na segunda etapa, denominada de secagem primária ou sublimação, a pressão da câmara é reduzida a 611,73 Pa (0,006 bar), ponto que ocorre a sublimação da água sólida e simultaneamente aumento gradual da temperatura da prateleira, permitindo a transferência de calor para o alimento. Nesse estágio, a temperatura é muito importante para o sucesso da liofilização do produto. O controle é realizado através da temperatura fornecida para a prateleira e a pressão da câmara (Anandharamakrishnan, 2017).

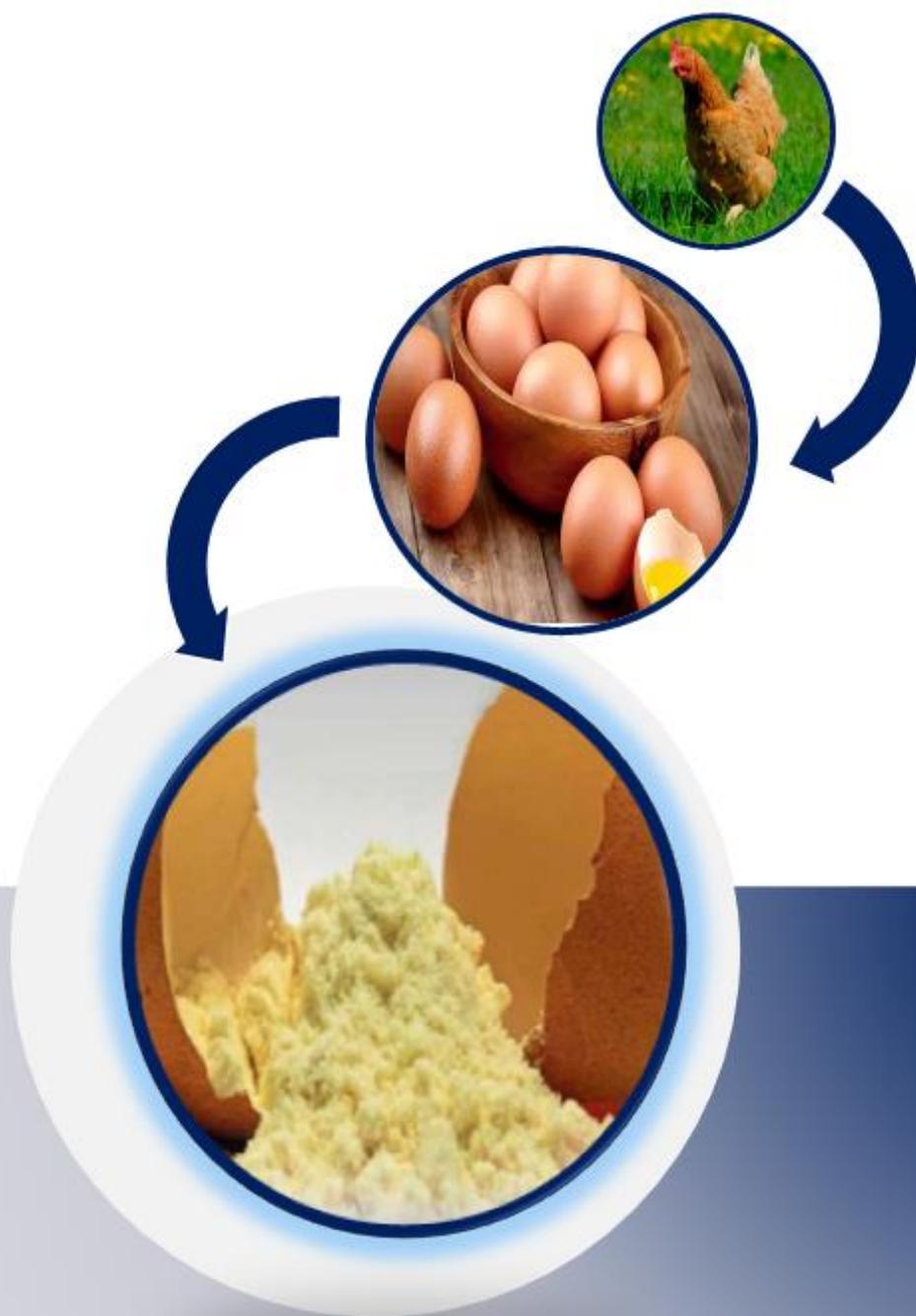
A última etapa, secagem secundária ou dessecção, ocorre somente quando toda água sólida for removida na etapa anterior. Tal procedimento é necessário, uma vez que o produto pode continuar retendo uma quantidade de água residual, como aquelas ligadas à estrutura do alimento. Esse estágio remove cerca de 5 a 10% do total de água do material e acontece mantendo-se o alimento sob vácuo em temperatura de 20 a 60 °C (König, 2016).

Quando comparada com a secagem convencional, a liofilização se destaca como um importante método de desidratação, principalmente devido à manutenção das características estruturais do alimento, remoção da umidade realizada em baixas temperaturas e aumento da estabilidade do alimento durante sua estocagem, contribuindo para diminuição das reações de degradação (Liu *et al.*, 2021).

Dentre as vantagens da liofilização de alimentos, pode-se citar: não degrada os nutrientes constituintes do ovo, uma vez que é utilizada baixa temperatura; pouca perda de substâncias voláteis; formação de uma estrutura esponjosa, facilitando sua redissolução; inibição do desenvolvimento de substâncias patogênicas, além do retardamento dos processos enzimáticos que aceleram a degradação do alimento, ajudando na conservação (Ribeiro, 2012).

Capítulo 3

Determinação de umidade e teores de proteínas e fósforo totais em ovos de galinha em pó



3.1 DETERMINAÇÃO DE UMIDADE E TEORES DE PROTEÍNAS E FÓSFORO TOTAIS EM AMOSTRAS DE OVOS DE GALINHA EM PÓ

3.1 Experimental

3.1.1 Amostras e preparo

Foram adquiridas 56 amostras de ovos de galinha em estabelecimentos comerciais da cidade de João Pessoa-PB, das quais eram compostas de 4 granjas e 4 lotes diferentes, sendo 36 de capoeira, 10 de granja e 10 de *free range*. Os ovos inteiros foram quebrados, sendo a clara e gema submetidas a uma etapa de processamento utilizando um miniprocessador da marca *Black & Decker*, modelo HC31T, os quais posteriormente foram colocados em placas de Petri e congelados. Em seguida, as amostras foram liofilizadas, usando um liofilizador da marca Liobras, modelo L101, em uma temperatura de 38 °C e pressão de 211 mmHg. Um diagrama esquemático do procedimento é apresentado na **Figura 3.1**.

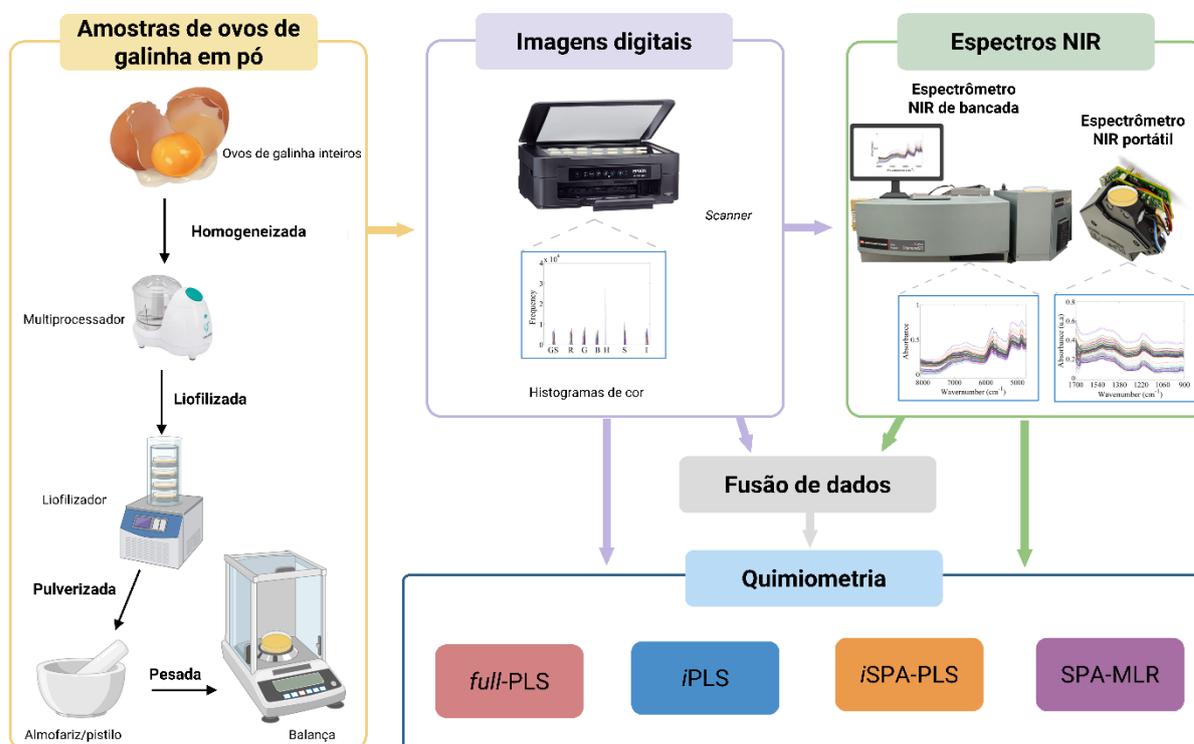
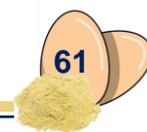


Figura 3.1 – Diagrama esquemático do processo de tratamento da amostra, obtenção das imagens, espectros NIR e ferramentas quimiométricas utilizadas. Elaborado no biorender.com



Após o processo de liofilização, o ovo em pó foi pulverizado. No intuito de aumentar o número de amostras e assegurar variabilidade e representatividade aos modelos quimiométricos, foram realizadas misturas secundárias e terciárias, resultando em 132 amostras de ovos de galinha em pó. Posteriormente, foram submetidas à vácuo e acondicionadas em sacos plásticos zip com fechamento hermético, sendo devidamente etiquetadas e guardadas em um dessecador, a fim garantir a preservação de suas características físico-químicas. Para ampliar a faixa de concentração dos parâmetros de qualidade analisados, politetrafluoretileno (PTFE) em pó foi adicionado a 15 ovos de galinha em pó para diluir as amostras e a outras 15 amostras foram adicionadas albumina em pó para aumentar o teor dos parâmetros analisados. Essas adições foram realizadas nas proporções 1:1, 1:4 e 1:9 m/m.

3.1.2 Reagentes e soluções

Para realização dos métodos de referência, os seguintes reagentes de grau analítico foram utilizados: ácido sulfúrico (Neon, São Paulo, Brasil), ácido bórico (Vetec, São Paulo, Brasil), sulfato de cobre (Fmaia, Minas Gerais, Brasil), sulfato de sódio (QAC, São Paulo, Brasil), dióxido de titânio (Riedel-de Haën, Seelze, Alemanha), fenolftaleína (Dinâmica, São Paulo, Brasil), vermelho de metila (Vetec, São Paulo, Brasil), hidróxido de sódio (Dinâmica, São Paulo, Brasil), ácido clorídrico (Microquímica produtos para laboratórios, São Paulo, Brasil), óxido de zinco (Química e representações, Minas Gerais, Brasil) hidróxido de potássio (Fmaia, Minas Gerais, Brasil), molibdato de sódio (Vetec, São Paulo, Brasil), ácido ascórbico (Vetec, São Paulo, Brasil), ácido nítrico (Proquimios, Rio de Janeiro, Brasil) e albumina (Natura ovos, Rio Grande do Sul, Brasil).

A água utilizada no preparo de todas as soluções foi sempre recém destilada. Para o preparo da solução de ácido sulfúrico 0,05 mol/L diluiu-se 2,8 mL em 1 L de água destilada. Já o ácido bórico (H_3BO_3) foi preparado pesando-se 2,0506 g do reagente e diluindo-se para 1 L de água destilada. Preparou-se a solução de hidróxido de sódio (40% m/v), pesando-se 800,62 g para 2 L de água destilada. Para o preparo das soluções de indicadores, pesou-se 0,5000 g de fenolftaleína e 0,5000 g de vermelho de metila e diluiu-se ambos em 50 mL de álcool etílico e água destilada, respectivamente.

A solução de hidróxido de potássio (50% m/v) foi preparada dissolvendo 50 g do reagente em 100 mL de água. Para o preparo da solução de molibdato de sódio, adicionou-se 140 mL de ácido sulfúrico, 300 mL de água, 12,50 g de molibdato de sódio dihidratado ($\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$) e completou-se o volume para 500 mL. A solução de ácido ascórbico foi preparada no dia de uso a partir de 5 g de ácido ascórbico em 100 mL de água. No preparo da solução de molibdato-ácido ascórbico (25:10 v/v), adicionou-se 250 mL da solução de $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ em 100 mL da solução de ácido ascórbico, sendo aferido o volume para 1 L com água. Para o preparo da solução estoque padrão (1,0000 mg/mL), secou-se o dihidrogenofosfato de potássio (KH_2PO_4) em estufa por 2 horas a 101 °C. Em seguida, foram pesados 1,0967 g e dissolvidos em 250 mL de água. A solução padrão de trabalho (0,01 mg/mL) foi preparada por meio da diluição de 5 mL da solução estoque em 500 mL de água.

3.1.3 Análise das amostras de ovos de galinha em pó pelos métodos de referência

As amostras de ovos de galinha em pó foram submetidas ao método de referência em triplicata, sendo os parâmetros de umidade e proteína total determinados utilizando as normas do Instituto Adolfo Lutz. Já para o fósforo total usou-se o método internacional 995.11 da Associação Oficial dos Químicos Analíticos (AOAC, do inglês *Association of Official Analytical Chemists*). Tais metodologias serão descritas nos subtópicos seguinte.

3.1.3.1 Umidade

A umidade foi determinada pelo método gravimétrico, que corresponde à perda de massa quando a amostra é aquecida à 105 °C em estufa com circulação de ar, condição na qual a água é removida. Nesse processo, não somente a água é removida, mas também outras substâncias voláteis presentes. O resíduo obtido é denominado de resíduo seco ([Instituto Adolfo Lutz, 2008](#)).

Sendo assim, procedeu-se com a pesagem de 0,5 g da amostra em uma cápsula de porcelana, utilizando uma balança Shimadzu, modelo ATY224, previamente tarada. Levou-se a estufa (Quimis®) a 105 °C por 3 horas e resfriou-se em um dessecador de sílica em gel até temperatura ambiente. Em seguida, as amostras foram pesadas, sendo o teor de umidade calculado, conforme equação 3.1.



$$U = \frac{100 \times N}{P} \quad (3.1)$$

onde U = teor de umidade (% m/m); N = massa de umidade (perda de massa em g) e P = massa da amostra.

3.1.3.2 Teor de proteínas totais

A determinação de proteídeos é baseada no cálculo de nitrogênio total presente na amostra, sendo realizado pelo processo de *Kjeldahl* modificado. De maneira geral, a metodologia se baseia em três etapas: digestão, destilação e titulação. Na primeira, a matéria orgânica é decomposta com ácido sulfúrico, sendo o nitrogênio convertido em sal amoniacal. Na segunda, a amônia é liberada do sal amoniacal, por meio de uma reação com hidróxido e depositada em uma solução ácida de volume e concentração conhecida. A última etapa refere-se à determinação da quantidade de nitrogênio, a partir da titulação do excesso do ácido com hidróxido ([Instituto Adolfo Lutz, 2008](#)).

Desse modo, procedeu-se com a pesagem de 0,2000 g da amostra em papel seda, o qual foi transferido para o balão de *Kjeldahl*. Em seguida, foram adicionados 5 mL de ácido sulfúrico concentrado e 1,5000 g de mistura catalítica, composta de dióxido de titânio (TiO₂), sulfato de cobre (CuSO₄) e sulfato de sódio (Na₂SO₄) (0,3:0,3:6,0). O TiO₂ ataca a água, produzindo radicais livres, os quais destroem a matéria orgânica. O CuSO₄ funciona como catalisador acelerando a reação. Já o Na₂SO₄ aumenta o ponto de ebulição da água, funcionando como agente ebulioscópico. A amostra foi levada para o bloco digestor, na capela, a 350 °C com rampa de aquecimento de 50 °C a cada 20 minutos, a fim de não ocorrer perda do material. A digestão permaneceu até que a solução se tornasse azul esverdeada e livre de material não digerido. Depois ficou por mais uma hora em aquecimento e ao final deixou-se esfriar.

Após essa etapa, adicionou-se 10 gotas de fenolftaleína no balão de *Kjeldahl*, o qual foi transferido para o sistema de destilação, da marca Tecnal, modelo TE-036/1. A amônia foi recolhida em uma solução de ácido bórico 0,033 mol/L com 3 gotas do indicador vermelho de metila. Ao frasco que continha a amostra digerida, adicionou-

se, através de um funil com torneira do sistema, uma solução de hidróxido de sódio a 40% até garantir o excesso de base. A amostra foi levada ao aquecimento e destilada até a obtenção de 75 mL do destilado. Por último, a solução de hidróxido de amônio foi titulada com ácido sulfúrico 0,05 mol/L padronizado, utilizando como indicador o vermelho de metila. A quantidade de proteínas foi calculada por meio da equação 3.3.

$$P = \frac{(V_A - V_B) \times f_c \times F \times 0,14}{m} \times 100 \quad (3.3)$$

onde P = proteína total (g/100g); V_A = volume de ácido sulfúrico 0,05 mol/L padronizado gasto na titulação da amostra; V_B = volume de ácido sulfúrico 0,05 mol/L padronizado gasto na titulação do branco; f_c = fator de correção da solução de ácido sulfúrico 0,05 mol/L; F = fator de conversão (6,25); m = massa da amostra.

3.1.3.3 Teor de fósforo total

O teor de fósforo total foi determinado por espectrometria UV-Visível. A intensidade da cor azul foi medida em 823 ± 1 nm em um espectrofotômetro da marca Hewlett Packard, modelo 8453. A coloração é originada da formação de um complexo $[(\text{MoO}_2 \cdot 4\text{MoO}_3)_2 \cdot \text{H}_3\text{PO}_4]$, por meio da reação do resíduo de fósforo, que é solúvel em ácido, com Na_2MoO_4 , na presença do ácido ascórbico, o qual funciona como agente redutor (AOAC, 1995).

Para o preparo da curva de calibração, transferiu-se 0; 1,00; 2,00; 3,00; 4,00; 5,00 e 6,00 mL da solução padrão de trabalho e completou-se para 15 mL de água destilada. Adicionou-se 20 mL da solução de molibdato-ácido ascórbico e agitou-se. As soluções foram levadas para o banho-maria por 15 minutos e aguardou-se resfriar a 20-30 °C. Em seguida, aferiu-se para 50mL e mediu-se a absorbância. A equação da reta obtida foi: $y = 831,9745x + 0,0016$ com $R^2 = 0,9945$.

Para o preparo das amostras, foi pesado 0,5000 g de ovo em pó e 0,5000 g de óxido de zinco (ZnO) em um cadinho, homogeneizou-se. Em seguida, foram levadas para a estufa a 110 °C por 2 horas e posteriormente incinerou-se em forno mufla a 550 °C por 4 horas. Os resíduos de cinzas foram dissolvidos em 5 mL de ácido clorídrico (HCl) e 5 mL de água destilada, sendo cobertos com um vidro de relógio e

fervidos por 5 minutos em uma chapa aquecedora. Em seguida, esperou-se esfriar e a solução foi filtrada. Então, neutralizou-se com KOH até que a solução ficasse levemente opalescente $[Zn(OH)_2]$. Em seguida, adicionou-se HCl até o desaparecimento da opalescência. Duas gotas a mais foram adicionadas e as soluções foram colocadas para esfriar. Posteriormente, diluiu-se para 100 mL com água destilada, pipetou-se 1 mL sendo o volume completado para 15 mL e procedeu-se conforme descrito na construção da curva de calibração. A medição foi realizada dentro de 1 hora após a reação de cor ocorrer. O teor de fósforo presente nas amostras expressos em termos de P_2O_5 (g/100g), calculados conforme as equações 3.4 e 3.5.

$$P = \frac{(V_2 - V_1) \times T_P}{W} \times 100 \quad (3.4)$$

$$P_2O_5 = 2,29 \times P \quad (3.5)$$

onde V_1 = volume da solução da reação de cor (mL) utilizado; V_2 = volume do balão volumétrico contendo as cinzas (100 mL); T_P = quantidade de P obtido da curva de calibração correspondente a absorbância do analito (mg); W = massa da amostra (mg).

3.1.3.4 Faixa de calibração, média e desvio padrão

A faixa de calibração, média e desvio padrão após processamento estatístico de cada parâmetro são apresentados de forma resumida na **Tabela 3.1**.

Tabela 3.1 - Faixa de calibração, média e desvio padrão obtidos na determinação de umidade e teor de proteína e fósforo totais em amostras de ovos de galinha em pó utilizando métodos de referência.

Parâmetro	Amostras	Faixa de calibração (% m/m)	Média (% m/m)	Desvio padrão
Umidade	132	1,02-5,87%	4,62	0,05
Teor de proteínas totais	132	23,5-72,2%	46,4	0,12
Teor de fósforo total	132	2,8-33%	23	0,15



3.1.4 Aquisição das imagens digitais

As amostras foram pesadas (6,0000 g) em placa de petri de vidro lisas (60 x 15 mm). A aquisição das imagens digitais foi realizada utilizando o *scanner* de uma impressora multifuncional *expression* da marca Epson, modelo XP-214 com resolução de 300 dpi que minimiza a variação de luminosidade, foco e distância. A fim de garantir a reprodutibilidade no processo de obtenção das imagens, utilizou-se um suporte de Policloreto de Vinila (PVC) de tal forma o posicionamento das amostras foi mantido fixo em todos os ensaios. As imagens foram adquiridas no formato “BMP” (Bitmap ou Mapeamento de Bits), conforme recomendações de [Godoi \(2013\)](#). Todas as análises foram realizadas em triplicata, as quais permitiram a obtenção dos histogramas médios para cada amostra.

Os histogramas de cor (RGB, HSV e escala de cinzas) foram obtidos a partir de cada imagem digital usando a Interface gráfica Matlab gratuita “Imagens_gui”, baixada em www.ccen.ufpb.br/laqa/index.php/downloads. Para o tratamento dos dados foi delimitada uma região de trabalho circular (250 x 250 pixels) no centro de cada imagem representando aproximadamente 40% de sua área total.

Os histogramas foram construídos utilizando o vermelho, verde, azul, matiz, saturação, intensidade e escala de cinza. Cada componente dos modelos de cor é constituído por 256 tons, os quais foram utilizados como informação analítica da amostra que foi utilizada na construção dos modelos quimiométricos. Para verificar a existência de influência relativa de cada cor foram usados três diferentes modelos de cor e suas combinações: (a) escala de cinzas, (b) RGB, (c) HSV (d) cinzas + RGB, (e) cinzas + HSV, (f) RGB + HSV e (h) cinzas + RGB + HSV. Os modelos de cor foram compostos por (a) 3 x 768, (b) 3 x 768, (c) 3 x 256 (d) 3 x 1024, (e) 3 x 1024, (f) 3 x 1536 e (h) 3 x 1792 variáveis, respectivamente. Os histogramas médios foram obtidos e a informação analítica extraída foi tratada, conforme seção 3.1.6.

3.1.5 Aquisição dos espectros NIR

As amostras foram pesadas (6,0000 g) em placas de petri de vidro lisas (60 x 15 mm), sendo todos os ensaios realizados em triplicata e temperatura ambiente (23 ± 1 °C) em dois equipamentos: convencional e portátil. No primeiro experimento utilizou-se um espectrofotômetro de bancada FT-NIR da marca Applied Instrument Technologies e modelo Analyzer Diamond 20. A faixa utilizada foi de 12000 a 4000

cm⁻¹ (833 a 2500 nm), com uma resolução de 4 cm⁻¹, 16 varreduras, modo reflectância difusa e um detector de InGaAs. O segundo experimento foi realizado utilizando um NIR portátil da marca Texas Instruments, modelo DLP NIRscan Nano Evaluation Module. A faixa utilizada foi de 11111 a 5882 cm⁻¹ (900 a 1700 nm), com resolução digital de 228 pontos e 32 varreduras, utilizando a transformação de Hadamard. Utilizou-se como o branco das amostras uma pastilha de politetrafluoretileno (PTFE).

3.1.6 Procedimento quimiométrico

Os dados obtidos a partir dos espectros NIR foram pré-processados utilizando: a LBC; OBC; OBC + LBC; SNV; MSC; SGD, 1^a derivada e polinômio de 2^a ordem com tamanhos de janela: 17, 19 e 21 pontos para bancada e 5, 7 e 9 para portátil. Para as imagens digitais, os histogramas de cores foram utilizados sem nenhuma etapa de pré-processamento. A matriz foi separada em conjunto de treinamento (92) e teste (40) usando o algoritmo SPXY (Galvão *et al.*, 2005). Estudos de *outlier* foram realizados, usando o gráfico de influência versus resíduo Studentizado, para detectar a existência de possíveis amostras anômalas. Elaborou-se modelos *full*-PLS, SPA-MLR e *i*PLS, *i*SPA-PLS com 5, 10, 15, 20, 25, 30 intervalos com a interface disponibilizada pelo laboratório.

Todos os modelos de calibração multivariada e a elaboração dos gráficos utilizados nesse trabalho de tese foram desenvolvidos por meio do software Matlab R2010a da Mathworks®.

3.2 Resultados e discussão

3.2.1 Calibração multivariada utilizando o espectrômetro NIR de bancada e portátil

3.2.1.1 Seleção da região espectral de trabalho

Os espectros médios originais das 132 amostras de ovos de galinha em pó, adquiridos com o NIR de bancada (833 a 2500 nm) e portátil (900 a 1700 nm) são apresentados na **Figura 3.2 (a) e (b)**. Verifica-se na região espectral do NIR de bancada (2065 a 2500 nm) a existência de um sinal de alta absorbância e ruidoso, devido a saturação do detector. Outrossim, ocorre na região de 833 a 1213 nm do equipamento de bancada e na faixa de 1652 a 1700 nm do NIR portátil, as quais

apresentam uma região não informativa e com ruído espectral. Desse modo, foi necessário que tais variáveis fossem excluídas, utilizando seleção de variáveis a priori, a fim de obter uma melhor robustez dos modelos quimiométricos. Com isso, os dados foram analisados na faixa de 1214 a 2064 nm para o NIR de bancada (Figura 3.2 (c)) e 900 a 1651 nm para o portátil (Figura 3.2 (d)).

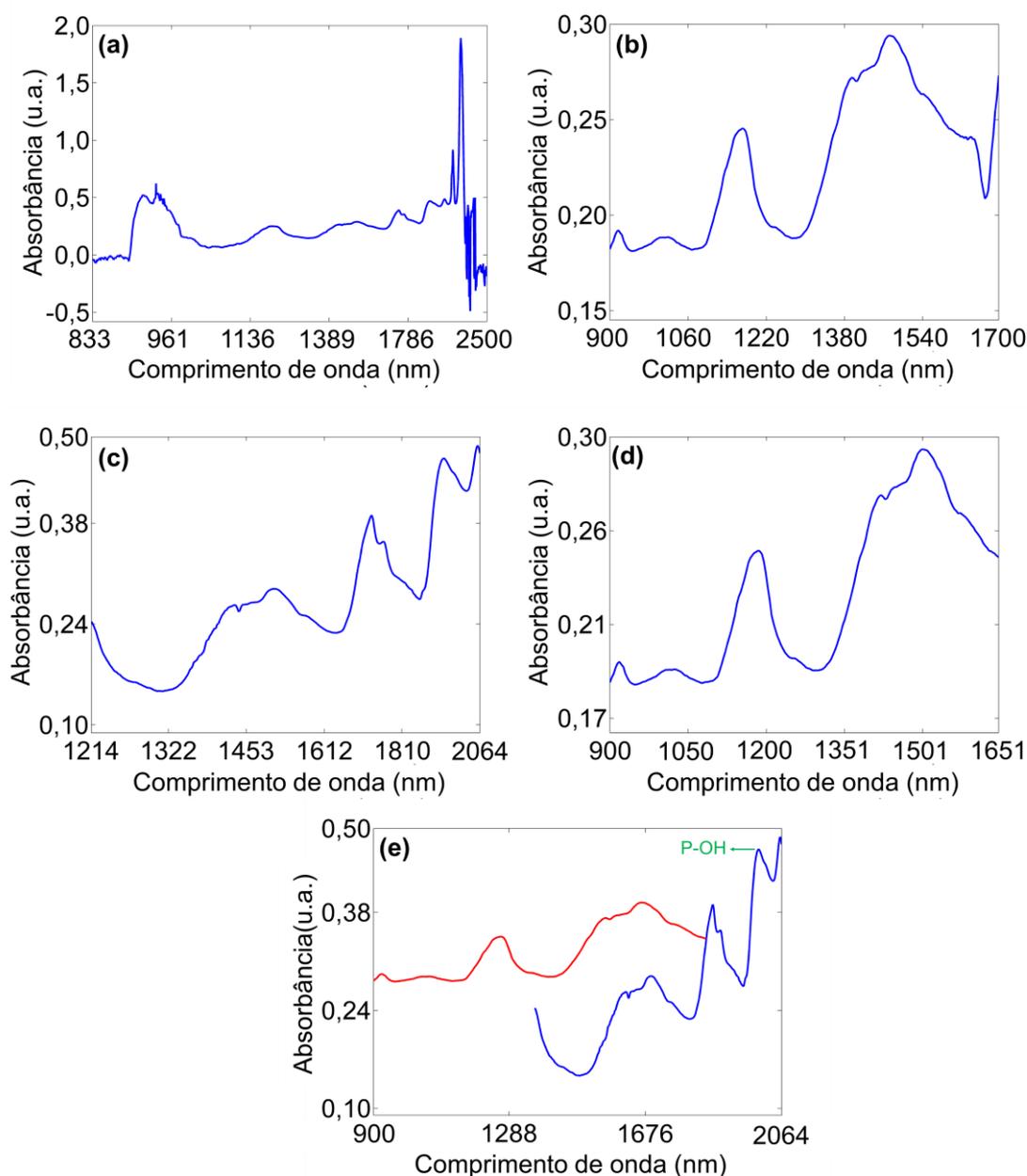


Figura 3.2 - Espectros médio NIR originais (a, b) e cortados (c, d) para as 132 amostras de ovos de galinha em pó, obtidos com os instrumentos de bancada e portátil, respectivamente e (e) espectros médios do NIR de bancada (—) e portátil (—) com *offset* que apresenta a localização da banda de P-OH presentes nas amostras de ovos de galinha em pó analisada nesta tese.

Ao investigar os espectros do equipamento de bancada (**Figura 3.2 (c)**) e portátil (**Figura 3.2 (d)**), verifica-se a presença de amplas bandas de absorção das amostras de ovos de galinha em pó entre 1333 e 1639 nm nos espectros do NIR de bancada e entre 1651 e 1280 nm nos espectros do NIR portátil, as quais podem ser atribuídas ao primeiro sobretom de bandas de combinação do alongamento CH e deformação CH, o primeiro sobretom do estiramento NH e OH das proteínas, lipídios, carboidratos e água. Além disso, nota-se a presença de picos em torno da região de 1724, 1942 e 2058 nm nos espectros do NIR de bancada (**Figura 3.2 (c)**), os quais podem ser atribuídos ao primeiro sobretom do alongamento CH, bandas de combinação do alongamento OH e deformação OH e bandas de combinação do alongamento NH + amida II, respectivamente. Somasse a isso, os picos em torno de 915, 1000 e 1170 nm podem ser atribuídos ao 3º harmônico do alongamento CH e ao 2º harmônico do alongamento OH e NH, respectivamente, da água e dos principais componentes encontrados nas amostras de ovos de galinha em pó, como carboidratos, lipídios e proteínas ([Bono, 2014](#); [Liu et al., 2022](#); [Pereira et al., 2020](#) e [Neves et al., 2022](#)).

Vale ainda salientar que a banda de absorção entre 1866 e 2020 nm (**Figura 3.2 (c)**) pode ser atribuída ao primeiro sobretom do P-OH, originada dos fosfolipídios nas amostras de ovo de galinha em pó ([Bono, 2014](#); [Workman](#); [Weyer, 2007](#)). Sendo assim, como o espectrômetro NIR portátil não cobre esta faixa (**Figura 3.2 (e)**), o teor de fósforo total foi determinado utilizando apenas imagens digitais e espectros NIR de bancada.

3.2.1.2 Pré-processamentos

Ao analisar os espectros obtidos por ambos os equipamentos, notou-se a presença de ruídos e variações sistemáticas na linha de base, ocasionada pelo espalhamento da luz devido à característica da amostra, uma vez que é majoritariamente sólida e possui partículas de diferentes tamanhos. Sendo assim, a fim de diminuir ou eliminar esses inconvenientes, aplicou-se diversas técnicas de pré-processamentos: LBC, OBC, LBC + OBC, SNV, MSC e SGD com primeira derivada, polinômio de segunda ordem e janelas de 17, 19 e 21 pontos (NIR de bancada) e 5, 7 e 9 pontos (NIR portátil) (**Figura 3.3** e **3.4**).

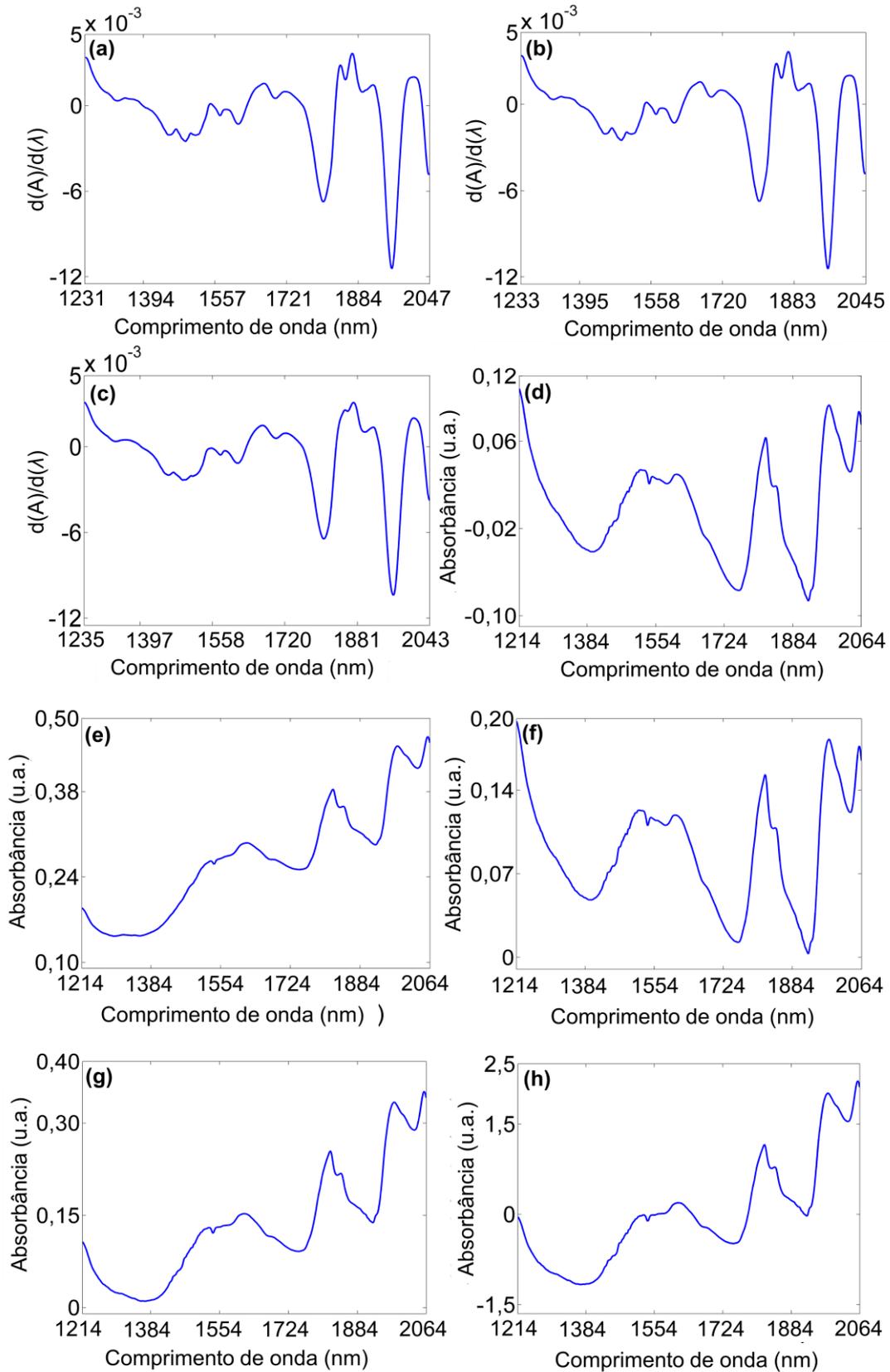


Figura 3.3 – Espectro NIR médio pré-processados com SGD e primeira derivada, polinômio de segunda ordem e janela de (a) 17 (b) 19 (c) 21 pontos; (d) LSC; (e) MSC; (f) OBC + LSC; (g) OBC; e (h) SNV para o equipamento de bancada.

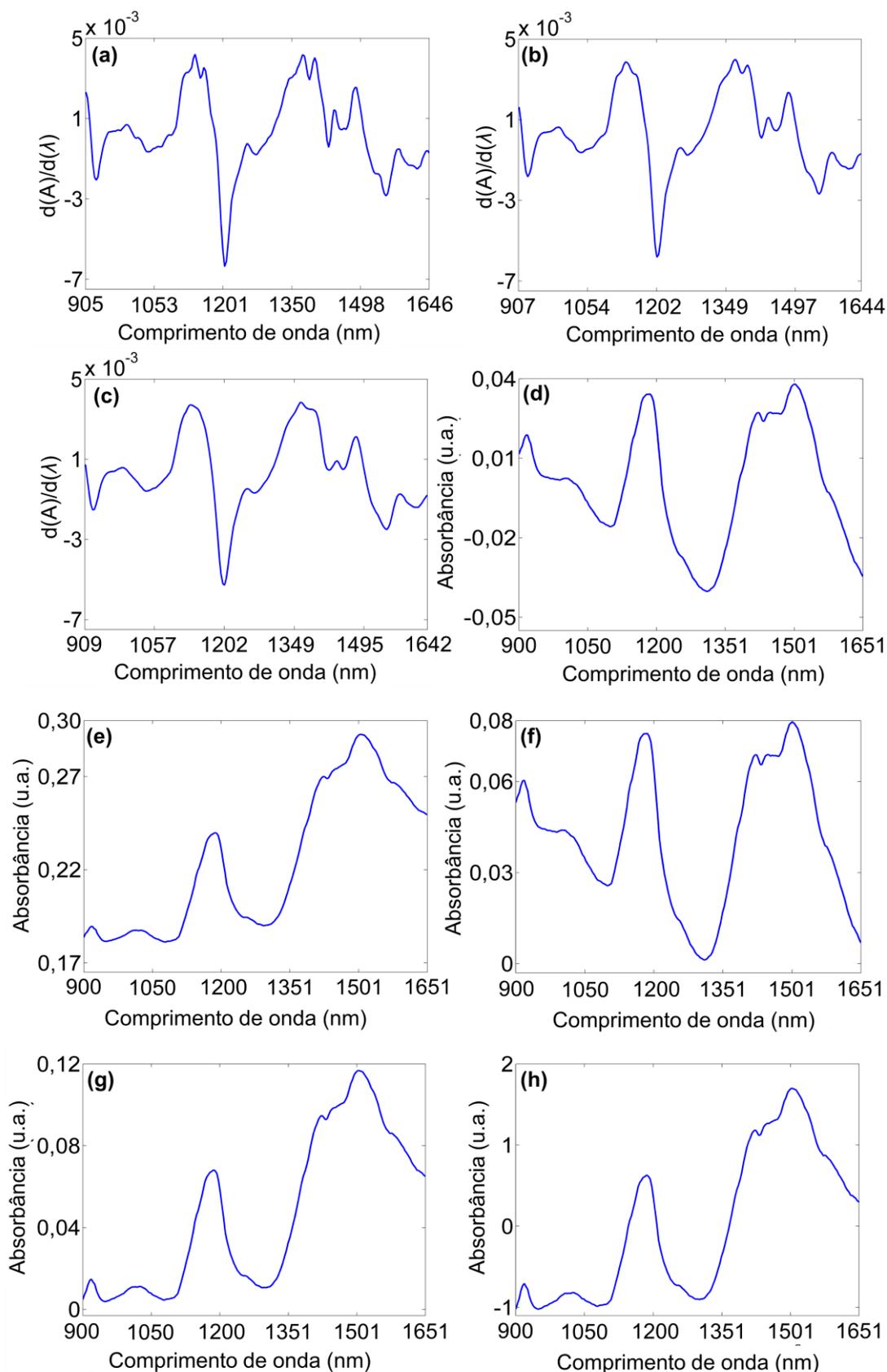


Figura 3.4 – Espectro NIR médio pré-processados com SGD com primeira derivada, polinômio de segunda ordem e janela de (a) 5 (b) 7 (c) 9 pontos; (d) LBC; (e) MSC; (f) OBC + LBC; (g) OBC; e (h) SNV para o equipamento portátil.

Nota-se nos gráficos das **Figuras 3.3 e 3.4**, que após a aplicação dos pré-processamentos houve uma redução do ruído, espalhamento e correção da linha de base dos espectros obtidos.

4.2.1.3 Estudo de *outlier*

Antes da construção dos modelos *full*-PLS, SPA-MLR, *i*PLS, *i*SPA-PLS foi realizado um estudo para verificar a presença de *outlier* utilizando o gráfico de Resíduo Studentizado *versus* Influência, conforme a **Figura 3.5**. A linha vertical indica um valor crítico de influência, definido como $3k/n$, no qual k corresponde ao número de fatores e n ao número de amostras utilizadas no modelo. Já a linha horizontal representa o limite dos resíduos Studentizados com um nível de confiança de 95% (Duarte *et al.*, 2022).

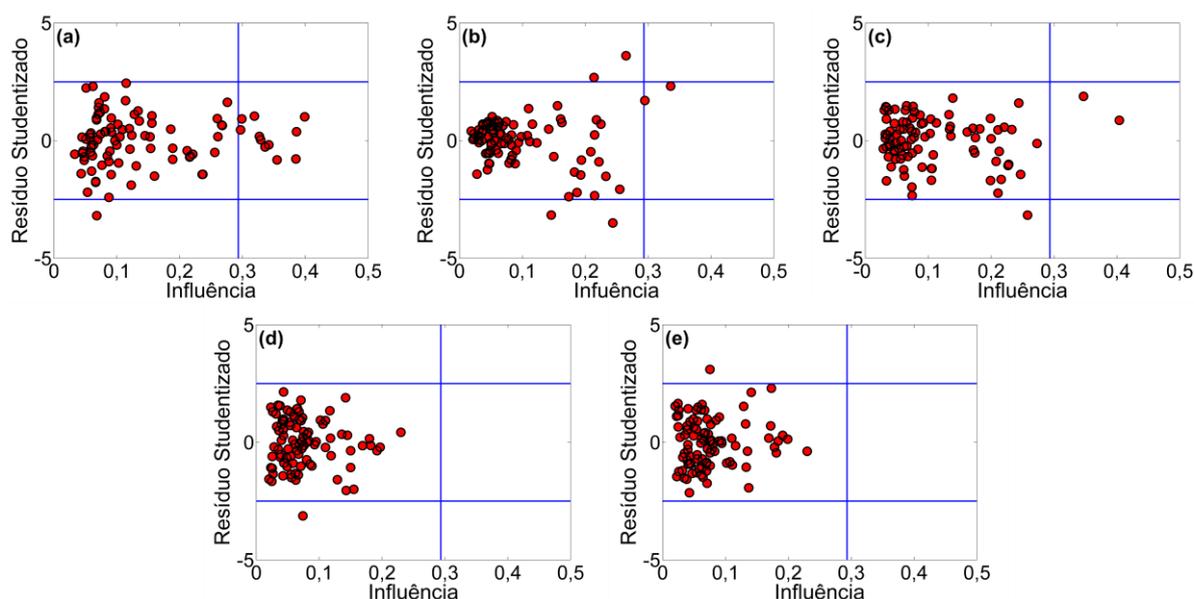


Figura 3.5 - Gráfico de Resíduo Studentizado *versus* Influência com o pré-processamento (a) OBC + LBC para umidade, SNV para (b) teor de proteínas e (c) fósforo totais do equipamento de bancada, derivada Savitzky-Golay com segunda derivada e janela de 17 pontos para (d) umidade e (e) teor de proteína total.

A partir do gráfico de Resíduo Studentizado *versus* Influência, construído a partir de um modelo *full*-PLS, verificou-se que algumas amostras apresentaram baixa influência e alto resíduo (**Figura 3.5 (b), (c), (d), (e) e (f)**), indicando que possivelmente seus vetores de resposta são equivalentes as amostras do conjunto de calibração, podendo suas concentrações ser mal previstas pelo modelo. No entanto, as amostras não apresentam uma alta influência e um alto resíduo, ou seja, acima dos pontos

críticos simultaneamente não há amostras, indicando que o modelo não força o ajuste dos pontos influentes. Sendo assim, nenhuma amostra do conjunto de dados foi considerada como *outlier* e todas foram utilizadas para construção dos modelos quimiométricos.

4.2.1.4 Escolha do número de variáveis incluídas no modelo

Nesse sentido, utilizou-se os valores de RMSECV na escolha do número de variáveis para os parâmetros de umidade e teores de proteínas e fósforo totais para o instrumento NIR de bancada e umidade e teor de proteínas totais para o NIR portátil (**Figura 3.6 (a) a (g)**). A partir dos gráficos, percebe-se que a partir de 12 e 7 variáveis espectrais e 9 variáveis latentes (NIR de Bancada) e 12 variáveis espectrais e 10 variáveis latentes (NIR Portátil), respectivamente, o valor de RMSECV obtido no conjunto de calibração não é alterado significativamente. Sendo assim, esses valores foram utilizados para a construção dos modelos *full*-PLS, *i*PLS, *i*SPA-PLS e SPA-MLR.

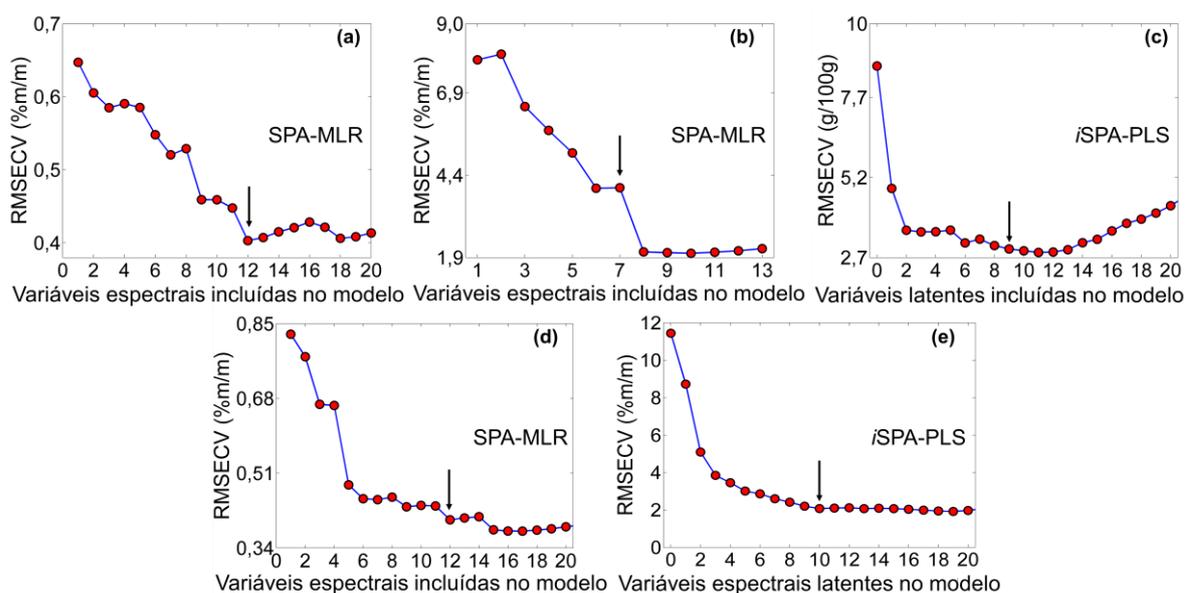


Figura 3.6 - Gráfico de RMSECV versus variáveis incluídas no modelo para (a) umidade, teor de (b) proteínas e (c) fósforo totais utilizando o equipamento NIR de bancada, (d) umidade e (e) teor de proteínas totais usando o equipamento NIR portátil.

4.2.1.5 Métodos de calibração multivariada

Os desempenhos de calibração e predição dos melhores métodos (SPA-MLR, *full*-PLS, *i*PLS e *i*SPA-PLS) construídos e validados em termos de RMSECV, r no conjunto de calibração e RMSEP, r , RPD e REP no conjunto de predição são apresentados nas **Tabela 3.2**.

Tabela 3.2 - Parâmetros de desempenho dos modelos nas etapas de calibração e predição para determinação de umidade e teor de proteína e fósforo totais em amostras de ovos de galinha em pó usando espectros NIR de bancada e algoritmos de calibração multivariada.

Parâmetro/ faixa de concentração da calibração	Pré-processamento	Algoritmo	NV	Conjunto de calibração				Conjunto de predição			
				RMSEC (% m m ⁻¹)	r	RMSECV (% m m ⁻¹)	r	RMSEP (% m m ⁻¹)	r	RPD	REP (%)
NIR de bancada											
Umidade (1,02-5,87% m m ⁻¹)	OBC + LBC	<i>full</i> -PLS	8	0,395	0,921	0,441	0,889	0,317	0,945	3,04	6,93
	OBC + LBC	SPA-MLR	12	0,414	0,905	0,403	0,909	0,222	0,974	4,35	4,84
	OBC + LBC	<i>i</i> PLS (15) ^a	6	0,390	0,921	0,445	0,887	0,262	0,963	3,69	5,71
	OBC + LBC	<i>i</i> SPA-PLS (20/5) ^a	6	0,406	0,914	0,424	0,898	0,291	0,954	3,31	6,35
Teor de proteínas totais (23,5-72,2% m m ⁻¹)	SNV	<i>full</i> -PLS	8	2,00	0,992	2,58	0,984	1,43	0,993	7,97	3,02
	SNV	SPA-MLR	7	2,07	0,999	2,07	0,999	1,29	0,994	8,83	2,72
	SNV	<i>i</i> PLS (5) ^a	8	1,83	0,993	2,57	0,984	1,29	0,994	8,83	2,72
	SNV	<i>i</i> SPA-PLS (5/2) ^a	8	1,90	0,992	2,28	0,988	1,37	0,993	8,33	2,88
Teor de fósforo total (2,8-33% m m ⁻¹)	SNV	<i>full</i> -PLS	9	2,6	0,96	2,8	0,94	1,7	0,98	4,2	6,8
	SNV	SPA-MLR	8	2,8	0,95	2,8	0,95	1,9	0,96	3,8	7,7
	SNV	<i>i</i> PLS (25) ^a	8	2,8	0,95	3,0	0,94	2,0	0,96	3,6	8,0
	SNV	<i>i</i>SPA-PLS (30/11)^a	9	2,4	0,97	2,7	0,95	1,7	0,98	4,3	6,8
NIR portátil											
Umidade (1,02-5,87% m m ⁻¹)	SGD-1 ^a D-2 ^a O-7P	<i>full</i> -PLS	8	0,375	0,931	0,456	0,885	0,373	0,926	2,50	8,03
	SGD-1 ^a D-2 ^a O-7P	SPA-MLR	12	0,436	0,923	0,404	0,910	0,330	0,944	2,83	7,11
	SGD-1 ^a D-2 ^a O-7P	<i>i</i> PLS (20) ^a	4	0,440	0,898	0,454	0,885	0,398	0,910	2,35	8,56
	SGD-1 ^a D-2 ^a O-7P	<i>i</i> SPA-PLS (10/5) ^a	6	0,375	0,929	0,420	0,903	0,348	0,936	2,69	7,48
Teor de proteínas totais (23,5-72,2% m m ⁻¹)	SGD-1 ^a D-2 ^a O-7P	<i>full</i> -PLS	10	1,58	0,992	2,07	0,983	2,12	0,993	8,35	4,77
	SGD-1 ^a D-2 ^a O-7P	SPA-MLR	8	1,92	0,985	1,93	0,986	2,18	0,994	8,12	4,90
	SGD-1 ^a D-2 ^a O-7P	<i>i</i> PLS (20) ^a	5	3,56	0,954	3,69	0,947	3,24	0,985	5,46	7,29
	SGD-1^aD-2^aO-7P	<i>i</i>SPA-PLS (15/11)^a	10	1,53	0,992	1,96	0,985	2,03	0,994	8,71	2,03

RMSECV: Raiz quadrada do erro médio quadrático da validação cruzada; r_{cv} : Coeficiente de correlação da validação cruzada; **RMSEP**: Raiz quadrada do erro médio quadrático da Predição; r_{pred} : Coeficiente de correlação da predição; **RPD_{pred}**: Razão do desempenho do desvio da predição; **REP**: Erro relativo de predição; **NV**: Número de variáveis espectrais ou variáveis latentes; ^a: (número total de intervalos em que a matriz de resposta instrumental foi dividida / número de intervalos selecionados pelo *i*PLS e *i*SPA-PLS); **LBC**: Correção linha de base linear; **OBC**: Correção de linha de base com offset; **SNV**: Variação normal padrão e **SGD-1^aD-2^aO-7P**: Derivada Savitzky-Golay (SGD) usando primeira derivada (1^aD), polinômio de segunda ordem (2^a), e janela de sete pontos (17P).



Conforme pode ser visualizado na **Tabela 3.2**, o melhor desempenho analítico para determinação de umidade nas amostras de ovos de galinha em pó foi obtido utilizando os espectros NIR de bancada pré-processados com algoritmos LBC + OBC e o algoritmo SPA-MLR, o qual resultou em um r de 0,974 e $RPD = 4,35$. Como verifica-se na **Figura 3.7 (a)**, este modelo selecionou doze variáveis espectrais em regiões espectrais associadas ao primeiro sobretom do alongamento OH e bandas de combinação do alongamento OH e deformação OH da água.

Nota-se ainda que o melhor método para determinação do teor de fósforo total nas amostras de ovos de galinha em pó foi alcançado utilizando o modelo *i*SPA-PLS e os espectros NIR de bancada pré-processados com o algoritmo SNV, o qual alcançou valores de r e RPD de 0,98 e 4,3, respectivamente. Nessa performance, o *i*SPA-PLS selecionou nove variáveis latentes e onze intervalos de variáveis espectrais (**Figura 3.7 (b)**) localizados em todas as regiões espectrais, exceto a região da ampla banda de absorção composta pelo 1º harmônico de bandas combinadas do estiramento CH e deformação CH, o primeiro sobretom do alongamento NH e OH. Vale ressaltar que o modelo selecionou intervalos na banda de absorção associados ao 1º sobretom do estiramento POH dos fosfolípidios.

Por outro lado, os espectros registrados no NIR portátil, pré-processados com o algoritmo SGD-1ªD-2ªO-7P, o modelo obtido com *i*SPA-PLS resultou em uma melhor performance para quantificação do teor de proteínas totais nas amostras de ovo de galinha em pó, alcançando um $r = 0,994$ e $RPD = 8,71$. Como pode ser visualizado na **Figura 3.7 (c)**, este modelo selecionou dez variáveis latentes e onze intervalos de variáveis espectrais localizados em todas as regiões espectrais, exceto as regiões do 2º harmônico do alongamento CH e o 1º harmônico de bandas de combinação do alongamento CH e deformação CH.

Notou-se ainda que resultados satisfatórios também foram obtidos, para quantificação de umidade com o algoritmo SPA-MLR e espectros NIR portáteis pré-processados com SGD-1ªD-2ªO-7P, resultando em um $r = 0,944$ e $RPD = 2,83$. Outrossim, o algoritmo SNV e a modelagem SPA-MLR resultaram em bons resultados para determinação do conteúdo total de proteínas nas amostras de ovo de galinha em pó ($r = 0,994$, $RPD = 8,83$).

Além disso, conforme pode-se visualizar nas **Tabela 3.2**, o *i*SPA-PLS (Gomes *et al.*, 2013) e SPA-MLR (Galvão *et al.*, 2008) forneceram os melhores resultados. Isso aconteceu porque foi utilizado o Algoritmo de Projeções Sucessivas (SPA) para

selecionar intervalos ou variáveis espectrais individuais. A seleção utilizando SPA minimiza a multicolinearidade e melhora a capacidade preditiva dos modelos (Galvão *et al.*, 2020).

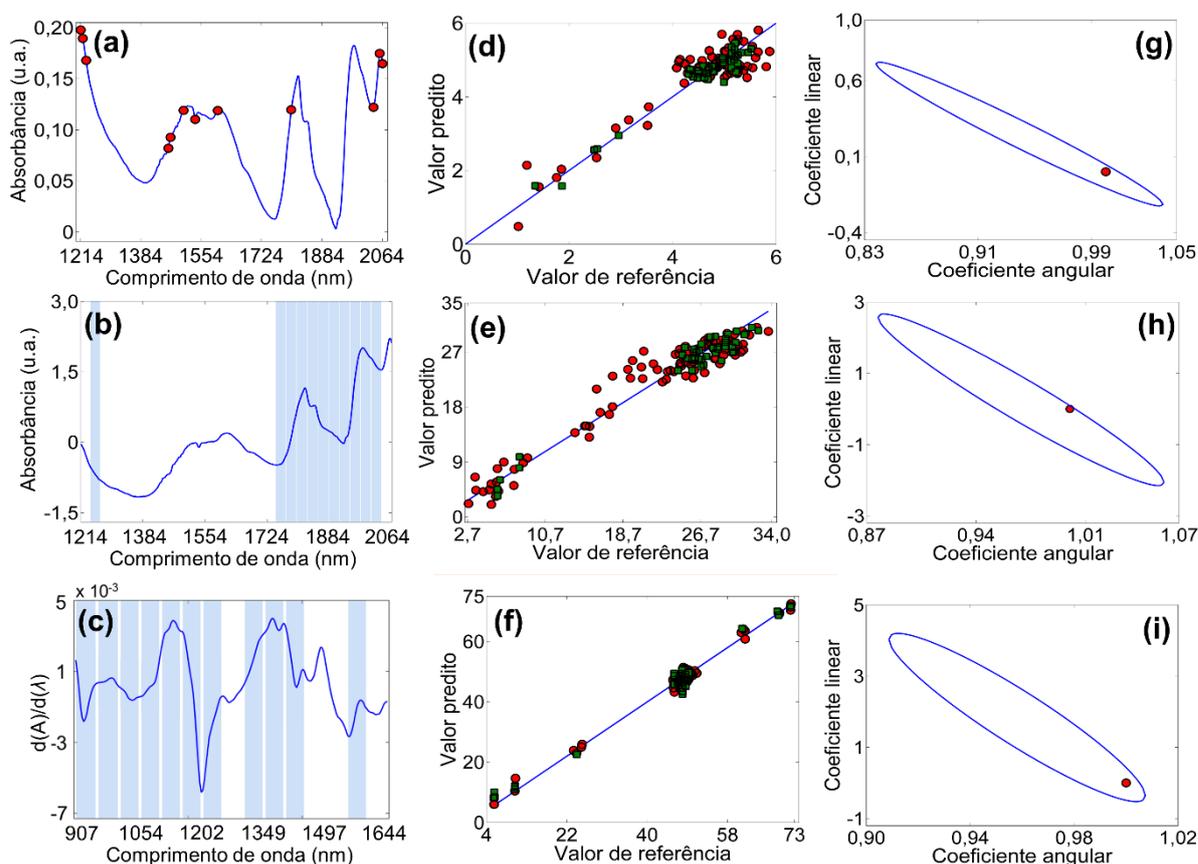


Figura 3.7 – Variáveis espectrais selecionadas pelo modelo SPA-MLR nos espectros NIR de bancada pré-processados com LBC + OBC (a). Curva de valor predito *versus* valor de referência (d) e curva da região de confiança elíptica conjunta (EJCR) para determinação da umidade (g). Intervalos de variáveis espectrais selecionadas pelo modelo *i*SPA-PLS em espectros NIR de bancada pré-processados com SNV (b), curva de valor previsto *versus* valor de referência (e) e curvas EJCR para determinação do teor de fósforo total (h). Intervalos de variáveis espectrais selecionadas pelo modelo *i*SPA-PLS em espectros NIR portáteis pré-processados com SGD (c), curva de valor predito *versus* valor de referência (f) e curva EJCR para determinação do conteúdo de proteína total (i).

A fim de ilustrar o bom ajuste dos melhores modelos (SPA-MLR e *i*SPA-PLS) para os parâmetros de umidade, proteínas e fósforo totais, elaborou-se um gráfico de valores preditos *versus* referência para as amostras de calibração e predição (**Figura 3.7 (d-f)**). Ademais, o teste EJCR também foi realizado, demonstrando que os modelos de calibração não contêm bias significativo, uma vez que a elipse de confiança contém o ponto teórico ideal (**Figura 3.7 (g-i)**).

3.2.2 Calibração multivariada utilizando os histogramas de cor das imagens digitais

3.2.2.1 Histogramas de cor das imagens digitais

Os histogramas de cor originais e com a remoção das variáveis não informativas das 132 amostras de ovos de galinha em pó são apresentados na **Figura 3.8 (a) e (b)**. Esses gráficos representam a distribuição de frequência dos níveis de cores constituintes das imagens digitais adquiridas pelo *scanner*. Em linhas gerais, as componentes podem conter 256 variáveis por canal (para 8 *bits*), devendo ser removida antes da construção do modelo as que possuem frequências iguais a zero, uma vez que são não informativas (Huang *et al.*, 2023). Sendo assim, nota-se que os histogramas originais são formados por 1792 variáveis e que 604 delas foram removidas por não colaborar com nenhuma informação. Portanto, os histogramas de cor de cada amostra de ovos de galinha em pó utilizados na modelagem de calibração multivariada tiveram um total de 1188 variáveis de cor.

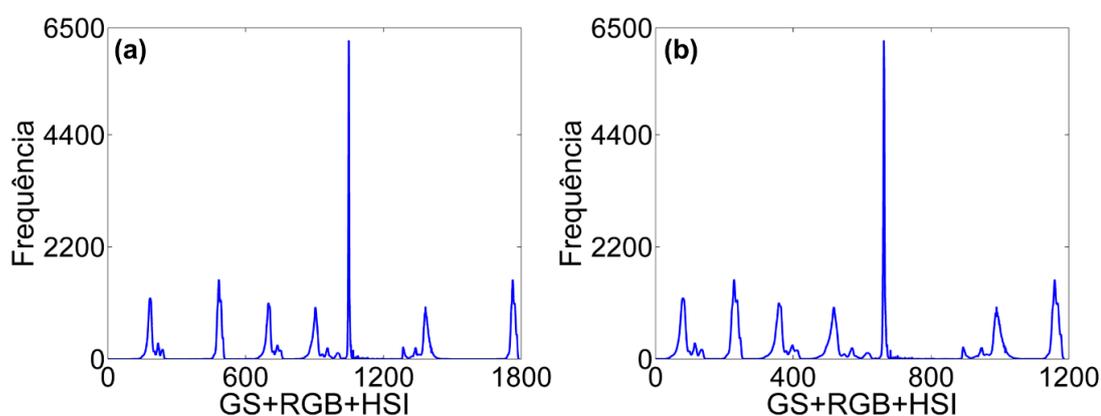


Figura 3.8 - Histogramas de cor (a) originais e (b) sem as variáveis não informativas.

Ao analisar os dados provenientes de umidade, notou-se que os modelos construídos não apresentaram resultados satisfatórios. Isso pode ter acontecido devido ao baixo teor de umidade, os quais não permitiram uma variação na coloração das amostras suficientes para que fosse detectada por imagens digitais. Sendo assim, apresenta-se os dados concernentes aos teores de proteínas e fósforo totais.

3.2.2.2 Estudo de *outlier*

Antes da construção dos modelos foi realizado um estudo de *outlier* a partir do gráfico de Resíduo Studentizado *versus* Influência, construído a partir de um modelo *full*-PLS, a fim detectar possíveis amostras anômalas, conforme a **Figura 3.9 (a) e (b)**.

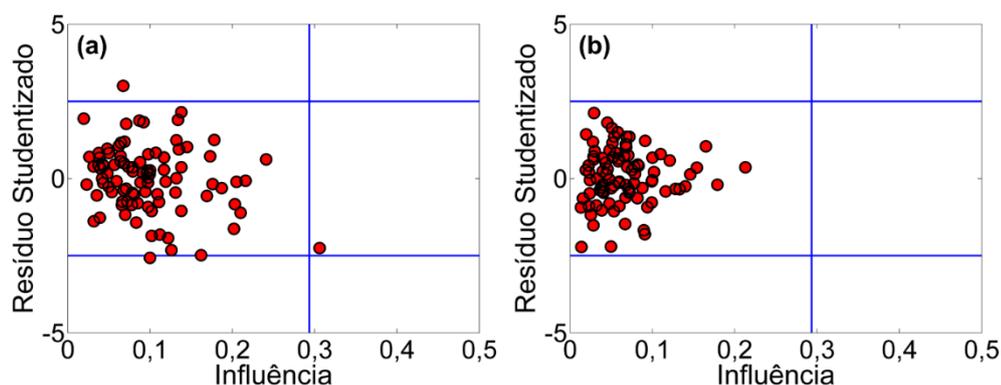


Figura 3.9 - Gráfico de Resíduo Studentizado *versus* Influência para o teor de (a) proteínas e (b) fósforo totais com os modelos de cor RGB+HSV e RGB, respectivamente.

Conforme visualiza-se nas **Figuras 3.9 (a) e (b)** nenhuma das amostras possuem alto resíduo e alta influência simultaneamente, indicando que no modelo de ambos os parâmetros, não possui amostras *outlier*, sendo todas utilizadas para construção dos modelos quimiométricos.

4.2.2.3 Escolha do número de variáveis incluídas no modelo

Os valores de RMSECV foram utilizados para a escolha do número de variáveis na calibração dos parâmetros de proteínas e fósforo totais para os dados adquiridos a partir dos histogramas de cor provenientes das imagens digitais (**Figura 3.10 (a) e (b)**). Nota-se nos gráficos que a partir de 10 variáveis latentes para o *i*SPA-PLS e 18 variáveis de cor para o modelo SPA-MLR, o valor de RMSECV obtido no conjunto de calibração não é alterado de maneira significativa. Sendo assim, esses valores foram utilizados para a construção dos modelos.

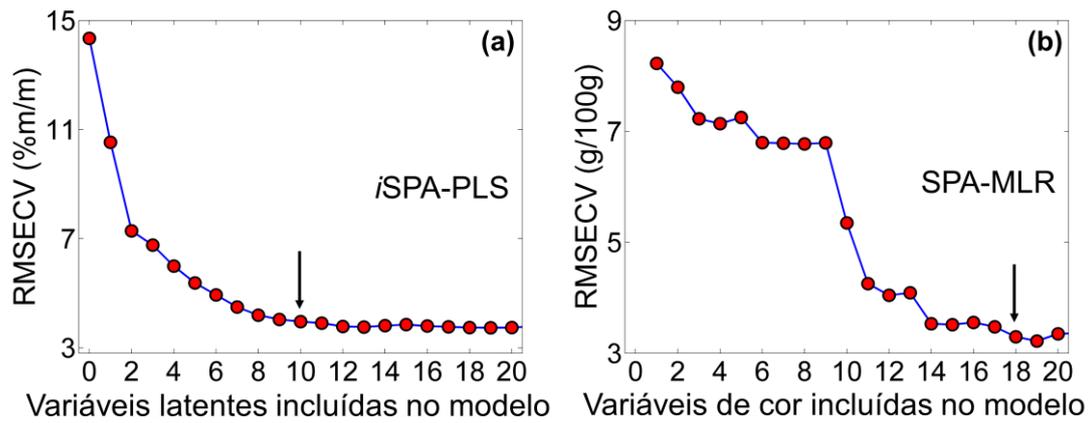


Figura 3.10 - Gráfico de RMSECV versus variáveis incluídas no modelo para o teor de (a) proteínas e (b) fósforo totais com os modelos de cor RGB + HSV e RGB, respectivamente.

3.2.1.4 Métodos de calibração multivariada

O desempenho dos modelos na etapa de calibração e predição dos melhores métodos construídos e validados em termos de RMSECV, RMSEP, r , RPD e REP são apresentados na **Tabela 3.3**.

Tabela 3.3 - Parâmetros de desempenho dos modelos nas etapas de calibração e predição para determinação do teor de proteínas e fósforo totais em amostras de ovos de galinha em pó usando imagens digitais e algoritmos de calibração multivariada.

Parâmetro/ faixa de concentração da calibração	Modelo de cor	Algoritmo	NV	Conjunto de calibração				Conjunto de predição			
				RMSEC (% m m ⁻¹)	r	RMSECV (% m m ⁻¹)	r	RMSEP (% m m ⁻¹)	r	RPD	REP (%)
Teor de proteínas totais (23,5-72,2% m m ⁻¹)	RGB + HSV	<i>full</i> -PLS	10	1,41	0,996	3,96	0,962	1,32	0,994	8,82	3,03
		SPA-MLR	13	4,86	0,942	5,04	0,937	5,06	0,942	2,31	11,60
		<i>i</i> PLS (10) ^a	10	1,03	0,998	1,55	0,994	1,59	0,992	7,37	3,63
		<i>i</i>SPA-PLS (30/21)^a	10	1,24	0,997	3,43	0,972	1,22	0,995	9,55	2,80
Teor de fósforo total (2,8-33% m m ⁻¹)	RGB	<i>full</i> -PLS	4	3,0	0,94	3,7	0,82	1,7	0,87	2,7	10
		SPA-MLR	18	3,2	0,94	3,3	0,93	2,5	0,94	2,8	9,9
		<i>i</i> PLS (15) ^a	4	4,1	0,90	4,2	0,88	2,5	0,94	2,8	10
		<i>i</i> SPA-PLS (20/17) ^a	4	3,1	0,94	3,5	0,92	2,7	0,94	2,7	10

RMSECV: Raiz quadrada do erro médio quadrático da validação cruzada; **r_{cv}**: Coeficiente de correlação da validação cruzada; **RMSEP**: Raiz quadrada do erro médio quadrático da Predição; **r_{pred}**: Coeficiente de correlação da predição; **RPD_{pred}**: Razão do desempenho do desvio da predição; **REP**: Erro relativo de predição; **NV**: Número de variáveis espectrais ou variáveis latentes; ^a: (número total de intervalos em que a matriz de resposta instrumental foi dividida / número de intervalos selecionados pelo *i*PLS e *i*SPA-PLS); **RGB**: Vermelho-Verde-Azul e **HSV**: Matiz-Saturação-Valor.

Conforme pode-se visualizar na **Tabela 3.3**, o melhor desempenho analítico dos modelos para determinação do teor proteínas totais nas amostras de ovos de galinha em pó foi obtido utilizando histogramas do modelo de cores RGB + HSV, e o *i*SPA construído e validado. Este modelo selecionou dez variáveis latentes e vinte e um intervalos de variáveis de cor localizadas nos canais R, G, B, H e S (**Figura 3.11 (a)**), alcançando valores de $r = 0,995$ e um RPD de 9,55. Por outro lado, para determinação do teor de fósforo total, o modelo com melhor performance foi aquele utilizando histogramas do modelo de cor RGB e o algoritmo SPA-MLR. Nessa modelagem os valores de r e RPD foram 0,94 e 2,8, respectivamente, selecionando dezoito variáveis de cor (**Figura 3.11 (b)**) localizadas nos canais G e B. Os melhores modelos de predição para todos os parâmetros calibrados neste trabalho apresentaram RPD > 2,80. Esse valor é aceitável pela literatura e indica uma boa capacidade preditiva (Li *et al.*, 2022).

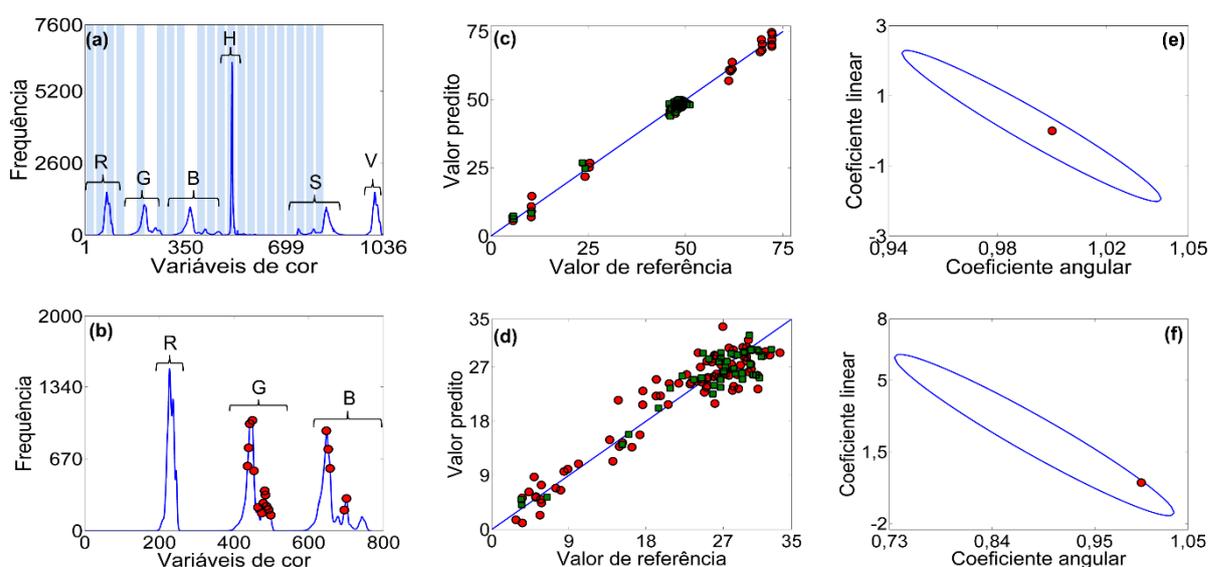


Figura 3.11 - Intervalos de variáveis de cor selecionadas pelo modelo *i*SPA-PLS nos histogramas de cor dos canais RGB + HSV (a), curva do valor predito *versus* valor de referência (c) e curva elíptica de confiança conjunta (EJCR) (e) para determinação do teor de proteínas totais. Variáveis de cor selecionadas pelo modelo SPA-MLR nos histogramas de cores dos canais RGB (b), curva do valor predito *versus* valor de referência (d) e curva EJCR (f) para o teor de fósforo total.

As curvas de valores predito *versus* valores de referência apresentaram um excelente ajuste dos métodos construídos e validados (**Figuras 3.11 (c)** e **(d)**), indicando a distribuição aleatória de amostras em ambos os lados da bissetriz (Gomes *et al.*, 2013). Além disso, também foram traçadas curvas EJCR para avaliar a tendência de bias nos métodos de calibração multivariada construídos. Como pode

ser visto nas **Figuras 3.11 (e) e (f)**, as elipses contêm o ponto teórico ideal (inclinação = 1 e intercepto = 0), indicando que os métodos propostos não apresentaram bias ao nível de confiança de 95% (Mandel e Linning, 1957).

A fim de ilustrar a variabilidade de proteínas e fósforo totais nas amostras, quatro imagens foram selecionadas, conforme as **Figuras 3.12 (a) e (b)**. Os histogramas de cores médios do canal GS+RGB+HSV, pode ser visualizado nas **Figuras 3.12 (c) e (d)**.

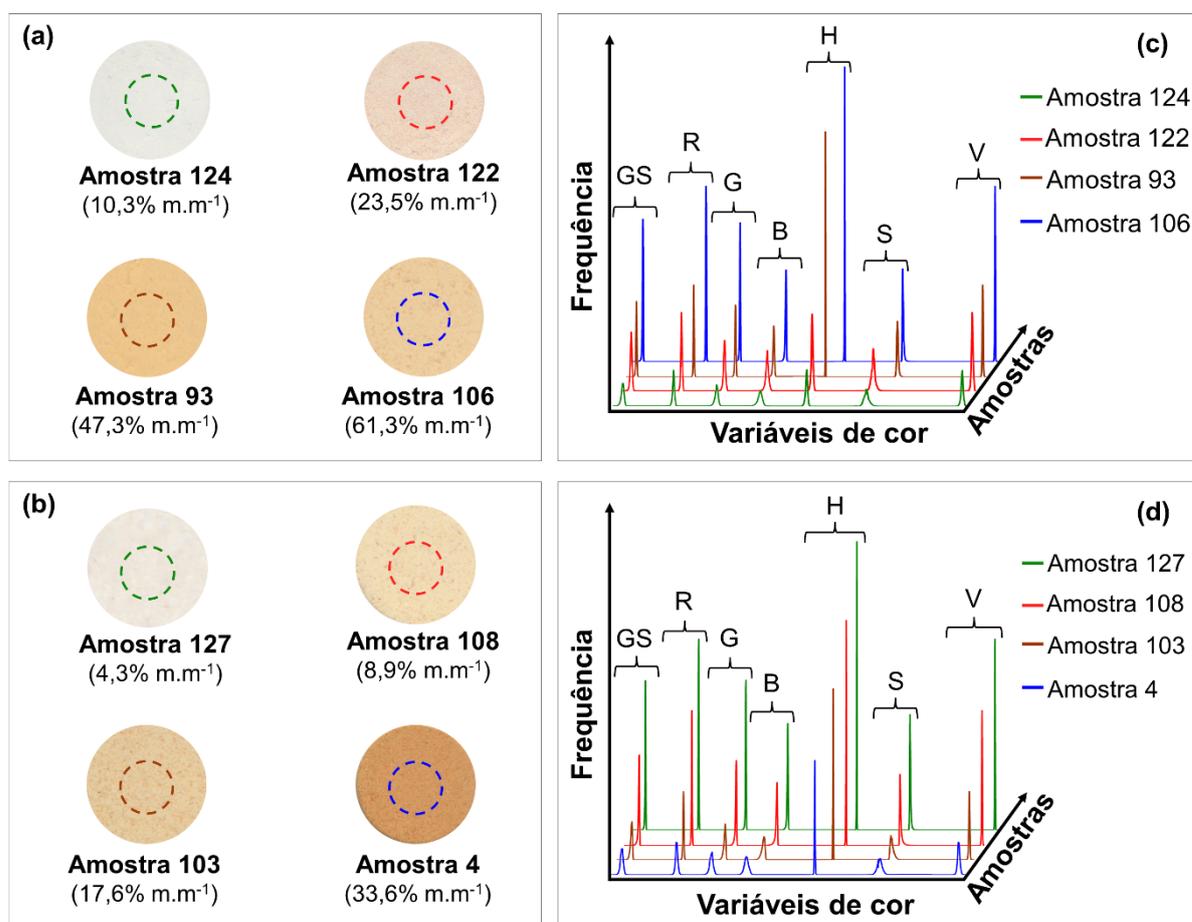


Figura 3.12 - Imagens digitais e histogramas de cor das amostras de ovos de galinha em pó com quatro diferentes teores de proteínas totais (a e c) e fósforo (b e d). O círculo tracejado nas imagens digitais é a região circular de interesse selecionada e usada pelo algoritmo IMAGENS_GUI para converter cada imagem em histogramas com os modelos de cores: GS, RGB e HSV.

As amostras de ovos de galinha em pó possuem uma coloração que ao olho humano varia entre amarelo e laranja, cuja cor complementar é verde e azul, respectivamente. Ao analisar os histogramas médios de cor (**Figura 3.12 (c)**), nota-se que com o aumento da concentração de proteína nas amostras, as distribuições de frequência aumentaram de intensidade, principalmente o vetor saturação e

intensidade do modelo de cor HSV, cujo canal se assemelha com o vetor vermelho. A mesma tendência é observada para a saturação que apresenta intensidade semelhante ao canal azul, enquanto a forma da escala de cinzas e o canal verde do modelo RGB são similares. Isso acontece devido ao teor de proteína na amostra, o qual é proporcional ao grau de pureza da cor, isto é, cromaticidade, o que leva a matiz, saturação e intensidade diferentes, proporcionando a obtenção de resultados satisfatórios para esse modelo de cor (Reile *et al.*, 2020).

Por outro lado, pode-se verificar nos histogramas de cor da **Figura 3.12 (d)** que quanto maior o teor de fósforo na amostra, menor foi a intensidade dos canais. Além disso, verificou-se que apenas os vetores verde e azul foram preferidos pelo algoritmo SPA-MLR (**Figura 3.11 (b)**). Isso pode ser explicado pela presença da lecitina, principal fonte de fósforo, cuja estrutura química é formada a partir de dois ácidos graxos principais: esteárico e palmítico. Na rota de síntese dos lipídios, o equilíbrio é deslocado para formação desses dois ácidos graxos, a fim de formar a lecitina, levando a produção de menos carotenoides no ovo, uma vez que compartilham da mesma rota de biossíntese. Os carotenoides são responsáveis pela coloração laranja avermelhado da gema e sua variação, influenciada pela presença de lecitina e detectada pelo modelo SPA-MLR, selecionado as variáveis nos vetores G e B do sistema de cor (Nelson Lehninger; Cox, 2022).

3.2.3 Calibração multivariada utilizando fusão de dados

Os modelos SPA-MLR, *full*-PLS, *i*PLS e *i*SPA-PLS foram construídos e validados usando fusão de dados em nível baixo, médio e alto, utilizando os espectros NIR pré-processados e as imagens digitais. O desempenho dos modelos nas etapas de calibração e predição dos métodos utilizando essa técnica quimiométrica em termos de RMSECV, RMSEP, *r*, RPD e REP são apresentados na **Tabela 3.4**.

Tabela 3.4 - Parâmetros de desempenho dos modelos nas etapas de calibração e predição para determinação de umidade e teor de proteínas e fósforo totais em amostras de ovos de galinha em pó usando fusão de dados.

Parâmetro/ faixa de concentração da calibração	Fusão de dados	Algoritmos	NV	Conjunto de calibração				Conjunto de predição			
				RMSEC (% m m ⁻¹)	r	RMSECV (% m m ⁻¹)	r	RMSEP (% m m ⁻¹)	r	RPD	REP (%)
Umidade (1,02-5,87% m m ⁻¹)	Nível baixo	full-PLS	8	0,394	0,921	0,440	0,889	0,316	0,945	3,05	6,90
	Nível médio	SPA-MLR	21	0,292	0,953	0,408	0,907	0,273	0,960	3,53	5,96
		iPLS	4	0,419	0,909	0,440	0,894	0,364	0,933	2,53	7,84
		iSPA-PLS	7	0,368	0,929	0,436	0,890	0,291	0,959	3,39	6,40
	Nível alto	-	-	0,341	0,935	0,421	0,887	0,357	0,931	2,73	7,68
Teor de proteínas totais (23,5-72,2% m m ⁻¹)	Nível baixo	full-PLS	6	1,12	0,998	1,95	0,993	0,98	0,984	5,50	2,04
	Nível médio	SPA-MLR	39	1,26	0,998	1,76	0,994	0,94	0,985	5,75	1,95
		iPLS	7	1,35	0,997	2,03	0,992	1,05	0,989	6,30	2,25
		iSPA-PLS	12	1,03	0,998	3,35	0,973	1,40	0,994	8,35	3,19
	Nível alto	-	-	4,61	0,954	2,98	0,975	1,91	0,987	6,27	4,37
Teor de fósforo total (2,8-33% m m ⁻¹)	Nível baixo	full-PLS	2	3,4	0,93	3,5	0,93	2,3	0,93	2,5	8,9
	Nível médio	SPA-MLR	20	3,2	0,94	3,4	0,93	2,1	0,93	2,7	8,1
		iPLS	9	3,3	0,94	3,4	0,93	2,5	0,64	1,2	9,2
		iSPA-PLS	2	3,4	0,93	3,6	0,92	2,2	0,94	2,9	8,7
	Nível alto	-	-	2,3	0,97	2,4	0,95	2,3	0,95	3,2	8,9

RMSEC: Raiz quadrada do erro médio quadrático da calibração; **RMSECV**: Raiz quadrada do erro médio quadrático da validação cruzada; **r_{cv}**: Coeficiente de correlação da validação cruzada; **RMSEP**: Raiz quadrada do erro médio quadrático da Predição; **r_{pred}**: Coeficiente de correlação da predição; **RPD_{pred}**: Razão do desempenho do desvio da predição; **REP**: Erro relativo de predição; **NV**: Número de variáveis espectrais ou variáveis latentes.

Conforme podemos visualizar na **Tabela 3.4**, todos os métodos desenvolvidos com fusão de dados em nível baixo, médio e alto, utilizando espectros NIR pré-processados e imagens digitais apresentaram performance significativa, em termos de r e RPD. No entanto, o desempenho destes métodos foi ligeiramente inferior ao dos melhores métodos utilizando espectros NIR pré-processados ou imagens digitais separadamente, como já apresentado anteriormente (**Tabelas 3.2 e 3.3**).

3.2.4 Comparação dos métodos propostos com outros métodos da literatura

Como pode-se notar na **Tabela 3.5**, os melhores métodos desenvolvidos nesta tese (**Tabelas 3.2 e 3.3**), apresentaram uma melhor performance em termos de r e RPD do que aqueles obtidos por [Zhao *et al.*, 2018](#); [Puertas *et al.*, 2023](#); [Watanabe *et al.*, 2018](#). Cabe ainda ressaltar que os melhores métodos para determinação de umidade e teor de proteínas e fósforo totais nas amostras de ovos de galinha em pó foram comparados com métodos de referência. Para tanto, esses métodos foram aplicados para analisar amostras de um conjunto de predição externo. Não foram observadas diferenças estatisticamente significativas entre os resultados ao aplicar o teste t pareado ao nível de confiança de 95%, comprovando a boa capacidade preditiva e confiabilidade dos melhores métodos desenvolvidos.

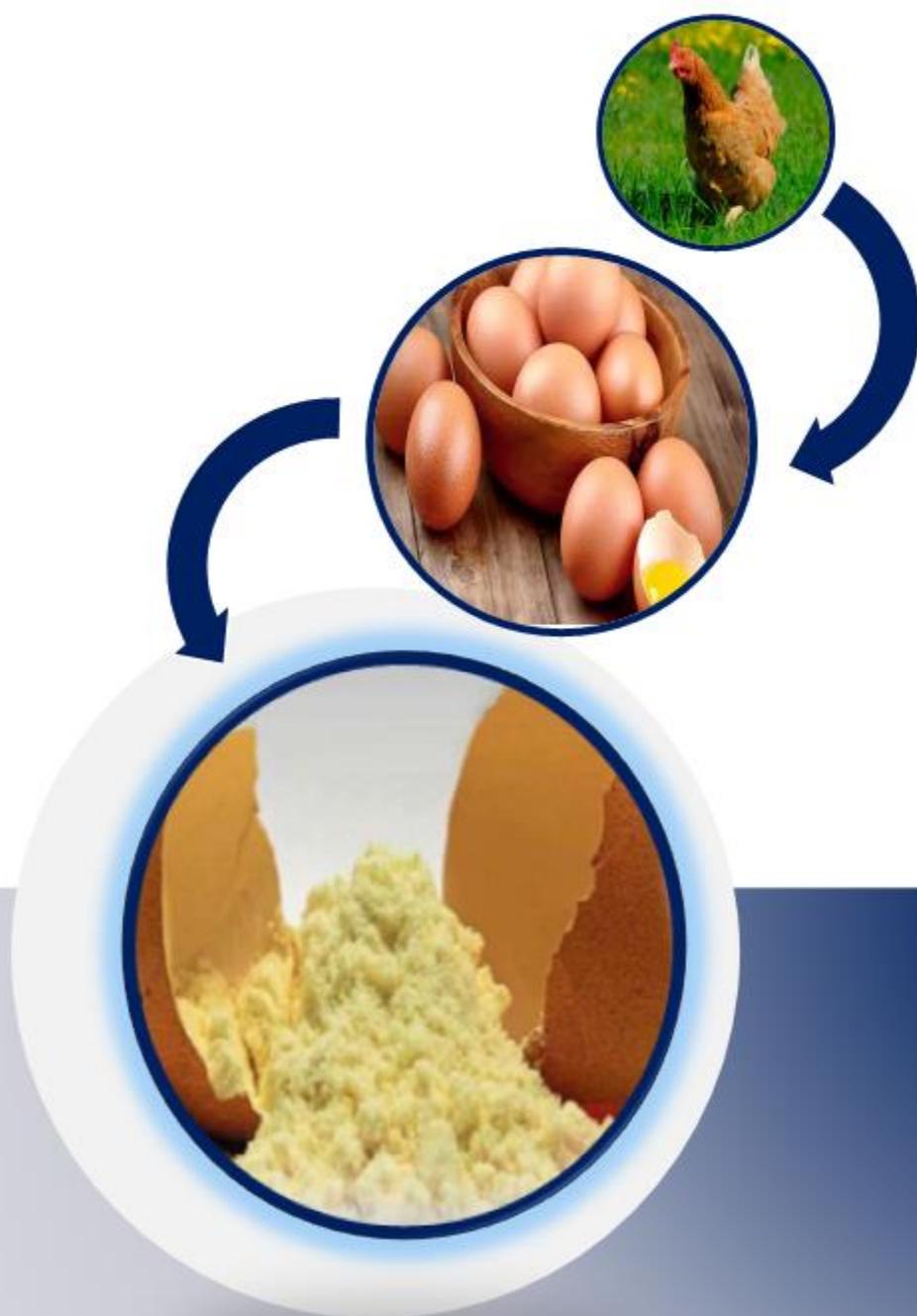
Além disso, métodos de calibração univariada foram previamente desenvolvidos para determinar o teor de proteínas totais e fósforo em ovos de galinha pré-processados usando, por exemplo, cromatografia líquida de alta eficiência ([Awadé e Efstathiou, 1999](#)), espectrofotometria ultravioleta e visível ([Birghila *et al.*, 2015](#)) e espectrofotometria de chama ([Attia *et al.*, 2014](#)). Entretanto, quando comparados com os métodos de calibração multivariada propostos nas **Tabela 3.2 a 3.4**, apresentam como desvantagens: lentidão, trabalhosos, requerem produtos químicos e geram resíduos prejudiciais à saúde e ao meio ambiente.

Tabela 3.5 - Características analíticas dos métodos quimiométricos propostos e os demais, utilizados para determinação de umidade e teores de proteínas e fósforo totais em amostras de ovos.

Parâmetro	Amostra	Processamento de ovo inteiro	Espectrômetro	Algoritmo	r	RPD	Referência
Umidade	claras de ovo líquidas	Quebra do ovo inteiro, separação da clara do ovo e homogeneização	NIR portátil	CARS-PLS	0,927	2,31	Zhao <i>et al.</i> (2018)
Teor de proteínas totais	claras de ovo líquidas	Quebra do ovo inteiro, separação da clara do ovo e homogeneização	NIR portátil	CARS-PLS	0,959	3,03	Zhao <i>et al.</i> (2018)
Teor de proteínas totais	ovos líquidos inteiros	Quebra de ovo inteiro, homogeneização, centrifugação e separação de membrana	UV-VIS-NIR de bancada	PLS	0,957	3,55	Puertas <i>et al.</i> (2023)
Umidade	Ovos de galinha em pó	fornecido por uma indústria de processamento de ovos	NIR de bancada	PLS	0,930	3,47	Watanabe <i>et al.</i> (2018)
Umidade	Ovos de galinha em pó	Quebra de ovo inteiro, homogeneização, liofilização e pulverização	NIR de bancada	SPA-MLR	0,974	4,35	Neste trabalho
Teor de proteínas totais	Ovos de galinha em pó	Quebra de ovo inteiro, homogeneização, liofilização e pulverização	Imagens digitais	iSPA-PLS	0,995	9,55	Neste trabalho
Teor de fósforo total	Ovos de galinha em pó	Quebra de ovo inteiro, homogeneização, liofilização e pulverização	NIR de bancada	iSPA-PLS	0,98	4,3	Neste trabalho

Capítulo 4

Autenticação de ovos de galinha em pó em termos de adulteração com proteína de soja e amido de milho





4 AUTENTICAÇÃO DE OVOS DE GALINHA EM PÓ EM TERMOS DE ADULTERAÇÃO COM PROTEÍNA DE SOJA E AMIDO DE MILHO

4.1 Experimental

4.2.1 Amostras e preparo

Este estudo compreende um total de 213 amostras de ovo de galinha em pó, sendo 76 puras e 137 adulteradas. As amostras adulteradas foram preparadas aleatoriamente pela adição de proteína de soja e amido de milho em diferentes níveis: 5, 10, 15, 20, 25 e 30 g/100g. As amostras puras foram adquiridas de fornecedores, enquanto os adulterantes foram comprados em supermercados locais da cidade de João Pessoa-PB.

4.2.2 Aquisição das imagens digitais e espectros NIR

A aquisição das imagens digitais usando um scanner e dos espectros utilizando um NIR portátil foi realizada, conforme as seções 3.5.4 e 3.5.5, respectivamente.

4.2.3 Procedimento quimiométrico

Os espectros NIR foram inicialmente pré-processados com diferentes técnicas, incluindo a LBC; OBC; OBC + LBC; MSC; SNV; SGD-1^aD-2^aO com janelas de 3, 5 e 7. Os histogramas de cores provenientes das imagens digitais sem nenhuma etapa de pré-processamento foram usados como informação analítica. A Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*) foi empregada, a fim de verificar a tendência natural de agrupamento das amostras. Em seguida, o algoritmo Kennard-Stone (KS) foi utilizado para dividir as amostras em conjuntos de treinamento (70%) e teste (30%).

Para construção do modelo DD-SIMCA foi adotado o nível de significância de outliers 0,05 para definição da área de aceitação utilizando a distribuição qui-quadrado no modo robusto. A escolha do número ótimo de Componentes Principais (PCs) foi obtido analisando os valores de sensibilidade do conjunto de treinamento com a validação cruzada completa, em função do número de PCs adicionada ao modelo.

O desempenho dos modelos DD-SIMCA foi avaliado em termos de sensibilidade, especificidade e acurácia. As rotinas para execução do KS e PCA foram disponibilizadas pelo laboratório, enquanto para calcular os modelos DD-SMICA, utilizou-se o pacote disponível em: <https://github.com/yzontov/dd-simca>. Todos os procedimentos quimiométricos foram realizados utilizando o software Matlab R2010a da Mathworks®.

4.2 Resultados e discussão

4.2.1 Espectros NIR portátil e imagens digitais

O espectro e o histograma de cor médios das 213 amostras de ovos de galinha em pó, obtidos na região do NIR de portátil (900 a 1700 nm) são apresentados nas **Figuras 4.1 (a)** e **(b)**, respectivamente. Nota-se na região de próxima de 910 e 1680 nm um sinal de saturação que foi descartado por seleção de variáveis a priori. Desse modo, a faixa de trabalho foi de 930 a 1650 nm, resultando em 204 variáveis (**Figura 4.1 (c)**). Além disso, a **Figura 4.1 (d)** apresenta o histograma de cor médio após a remoção dos zeros, o que resultou em 600 variáveis de cor.

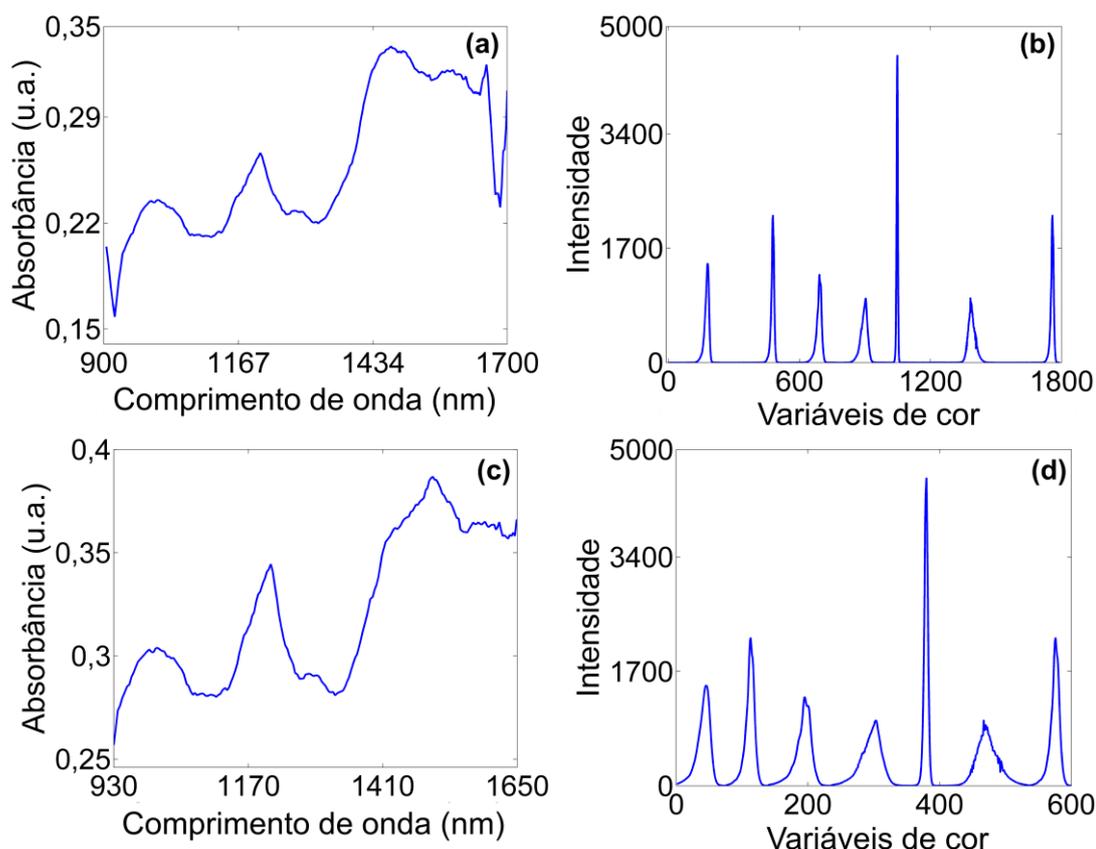


Figura 4.1 – Espectro médio completo do NIR portátil NIR (a) e (c) cortado e histograma de cor médio completo (b) e após a remoção dos zeros (d).

Os espectros provenientes do NIR portátil foram pré-processados, conforme seção a seguir.

4.2.2 Pré-processamentos nos espectros NIR portátil

Ao examinar os espectros obtidos a partir do NIR portátil, constatou-se a presença de ruídos e flutuações sistemáticas na linha de base, resultantes do espalhamento da luz. Isso se deve às propriedades da amostra, que é majoritariamente sólida e contém partículas de tamanhos variados. Sendo assim, aplicou-se técnicas de pré-processamentos, a fim de atenuar estes problemas. Inicialmente, utilizou-se o filtro derivativo Savitzky-Golay com primeira derivada, polinômio de segunda ordem e janela de 7 pontos. Somasse a isso, LBC; OBC; OBC + LBC; MSC e SNV também foram utilizados (**Figura 4.2 (a-f)**).

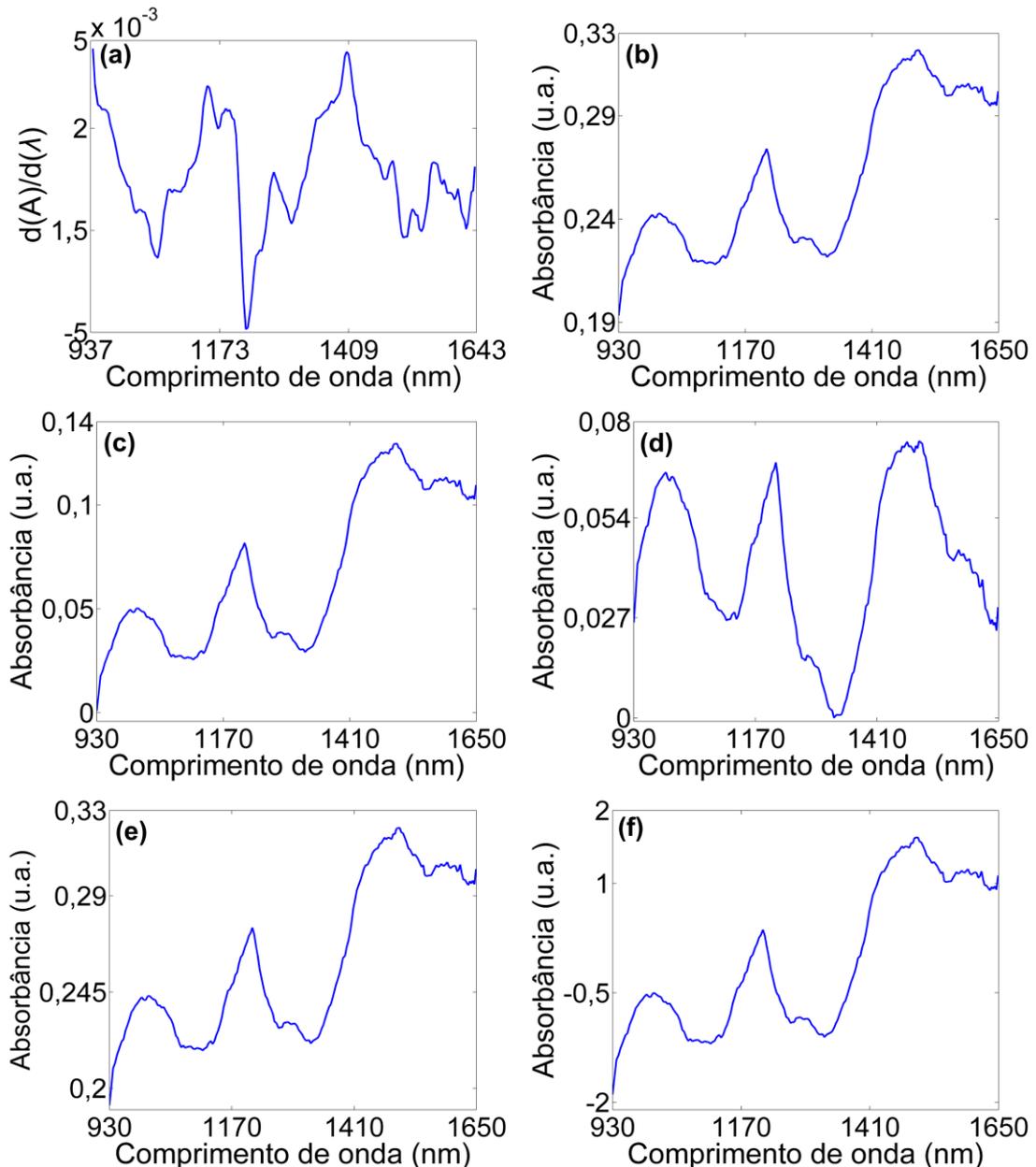


Figura 4.2 – Espectro médio do NIR portátil NIR pré-processado com (a) DSG-1^oD-2^oO com janelas de 3, 5 e 7, (b) LBC; (c) OBC; (d) OBC + LBC; (e) MSC e (f) SNV.

Após a aplicação de diferentes técnicas de pré-processamento nos espectros obtidos a partir do instrumento portátil (**Figura 4.2**), notou-se uma diminuição no espalhamento de luz, na quantidade de ruído e uma correção da linha de base. Depois da etapa de pré-processamento, o algoritmo PCA foi utilizado, a fim de realizar uma análise exploratória dos dados NIR. Essa investigação também foi realizada nos histogramas de cor provenientes das imagens digitais, logo após a remoção dos zeros.

4.2.3 Análise por componentes principais

Nas **Figuras 4.3** e **4.4** é ilustrado os gráficos dos scores das duas primeiras PCs dos espectros NIR portátil, tratados com os diferentes pré-processamentos e dos histogramas de cor dos modelos GS, RGB, HSV, GS+RGB, GS+HSV, RGB+HSV e GS+RGB+HSV das amostras de ovo de galinha em pó, respectivamente, no intuito de avaliar a capacidade de agrupamentos em relação a adulteração com proteína de soja e amido de milho.

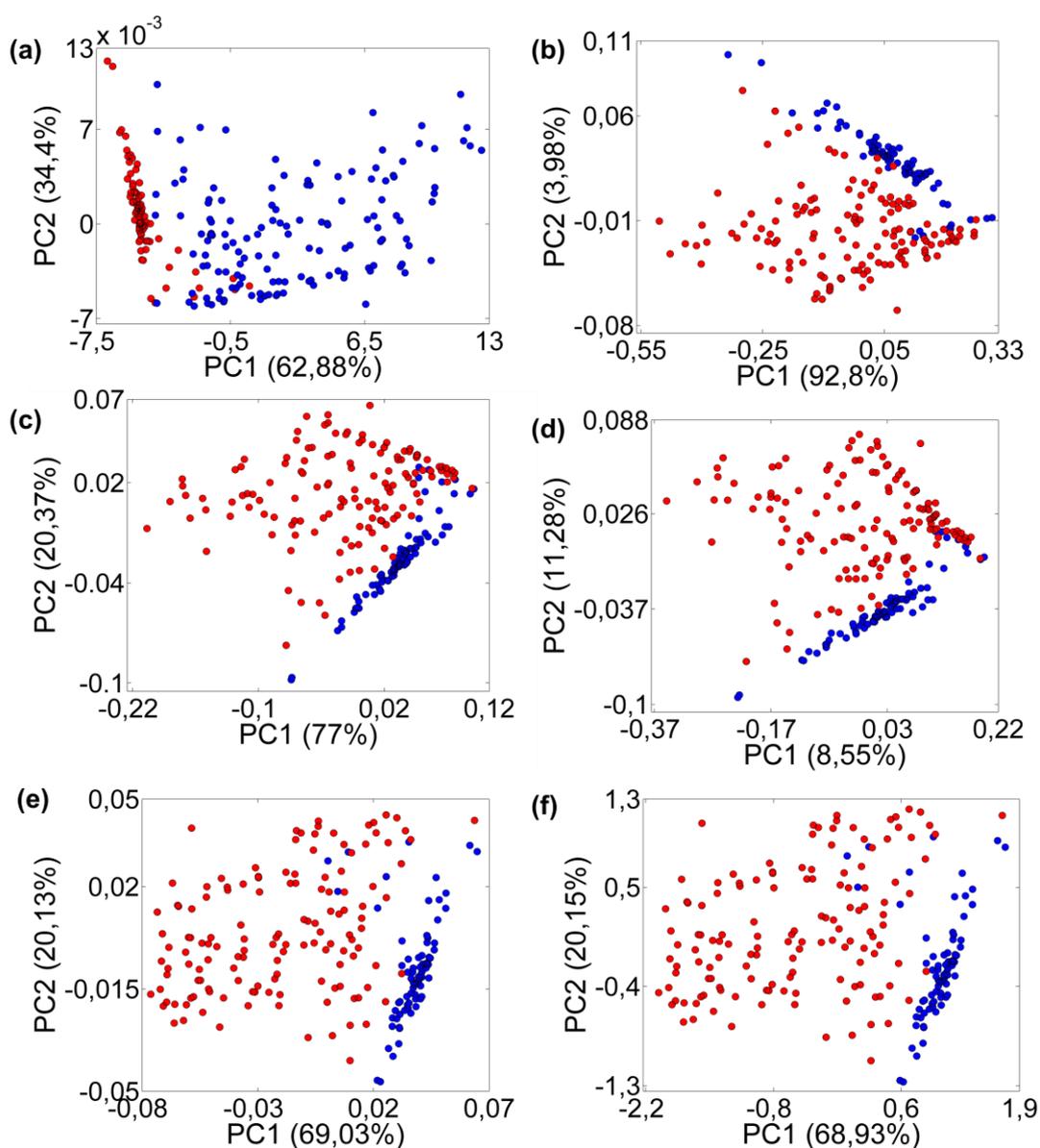


Figura 4.3 – Gráfico dos escores de PC1 versus PC2 para as amostras de ovo em pó puras (●) e adulteradas (●), utilizando os espectros NIR portátil pré-processados com (a) DSG-1°D-2°O com janelas de 3, 5 e 7, LBC (b); OBC (c); OBC + LBC (d); MSC (e) e SNV (f).

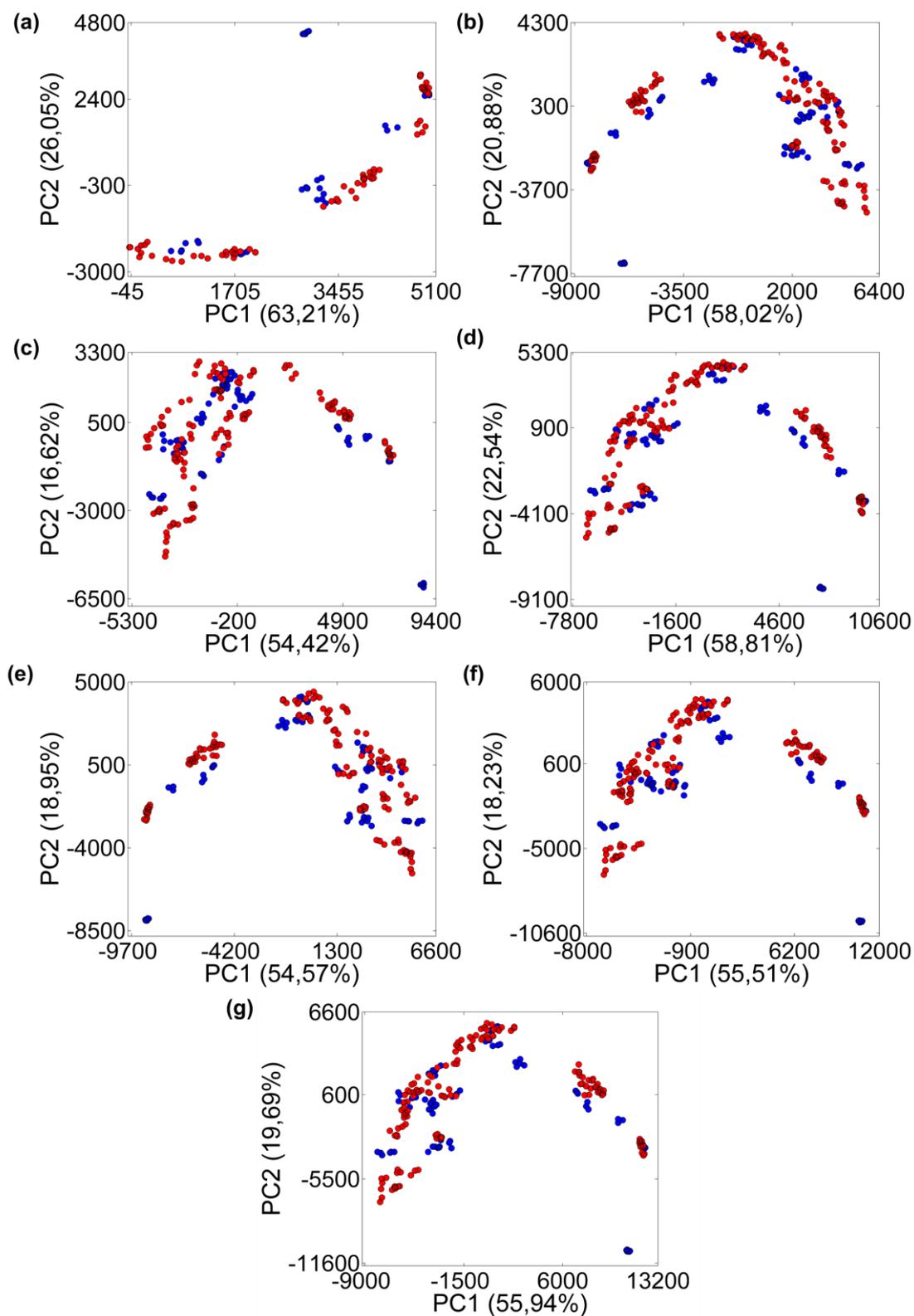


Figura 4.4 – Gráfico dos escores de PC1 versus PC2 para as amostras de ovo em pó puras (●) e adulteradas (●), utilizando os histogramas de cor com os modelos (a) GS, (b) RGB, (c) HSV, (d) GS+RGB, (e) GS+HSV, RGB+HSV e (f) GS+RGB+HSV.



Conforme visualiza-se na **Figura 4.3**, utilizando os espectros NIR portátil, há uma tendência de separação das amostras de ovo em pó puras e adulteradas. No entanto, ainda é possível verificar uma dispersão significativa, principalmente da classe adulterada com ligeira sobreposição com a classe pura. Por outro lado, embora exista uma tendência de separação entre os valores de PC1 e PC2 para os histogramas de cor, há uma forte sobreposição entre as classes estudadas e uma dispersão substancial. Essa grande variabilidade na distribuição das amostras nos gráficos de scores deve-se a diferentes fatores, como a níveis de adulteração, variabilidade nas raças das galinhas, alimentação e manejo.

Desse modo, considerando que essas fontes de variação estão intrinsicamente presentes nas amostras investigadas, a identificação da adulteração pela mistura com proteína de soja e amido de milho, torna-se uma tarefa difícil. Para superar esse inconveniente, o classificador “*one class*” foi utilizado para fins de autenticação, conforme discutido na seção seguinte.

4.2.4 Autenticação de ovos de galinha em pó por DD-SIMCA

A **Tabelas 4.1** apresentada os resultados obtidos para os modelos de classificação DD-SIMCA para autenticação dos ovos de galinha em pó frente a sua adulteração com proteína de soja e amido de milho, em termos de sensibilidade, especificidade e acurácia para os diversos tipos de pré-processamentos nos espectros originados do NIR portátil.

Tabela 4.1 - Resultados dos parâmetros de desempenho sensibilidade, especificidade e acurácia para DD-SIMCA nos conjuntos de treinamento e teste, utilizando os diferentes pré-processamentos dos espectros NIR portátil.

Método de classificação		DD-SIMCA Instrumento NIR portátil			
Pré-Processamento	Conjunto de amostra	PCs	Sensibilidade	Especificidade	Acurácia
DSG-1^aD-2^aO-7P	Treinamento	8	98,1	-	99,4
	Teste		95,7	100,0	
LBC	Treinamento	9	92,5	-	96,3
	Teste		91,3	97,1	
OBC	Treinamento	9	96,2	-	99,4
	Teste		95,7	100,0	
OBC + LBC	Treinamento	8	100,0	-	95,0
	Teste		91,3	95,6	
MSC	Treinamento	9	92,5	-	100,0
	Teste		100,0	100,0	
SNV	Treinamento	9	96,2	-	98,8
	Teste		100,0	98,5	

PCs: componentes principais; **DSG-1^aD-2^aO-7P:** Derivada Savitsky-Golay com 1^a derivada, polinômio de 2^a ordem e janela de 7 pontos; **LBC:** correção de linha de base linear; **OBC:** correção de linha de base offset; **MSC:** correção de espalhamento multiplicativo e **SNV:** variação normal padrão.

Ao analisar a **Tabela 4.1**, verifica-se que com o instrumento NIR portátil a utilização do pré-processamento do filtro Savitzky-Golay com 1^a derivada, polinômio de 2^a ordem e janela de 7 pontos promoveu o melhor resultado com boa sensibilidade no treinamento (98,1%) e um desempenho significativo no teste, com sensibilidade de 95,7%, especificidade de 100% e acurácia 99,4%, mostrando-se particularmente robusto e eficiente, evidenciando a eficácia do DD-SIMCA para fins de autenticação de ovos de galinha em pó adulterados com proteína de soja e amido de milho. Cabe ainda ressaltar que a técnica de OBC também teve um bom desempenho, enquanto o LBC apresentou modelos ligeiramente inferiores.

A escolha do número de componentes principais foi baseada no modelo que incluía o máximo de amostras de treinamento dentro da área de aceitação, ou seja, que proporcionava a máxima taxa de verdadeiros positivos (Neves e Poppi, 2020). Assim, para alcançar valores máximos de sensibilidade, os modelos DD-SIMCA com os espectros NIR portátil pré-processados selecionaram 8 ou 9 componentes principais. A quantidade de PCs selecionada pode estar associada à complexidade da matriz, uma vez que a amostra é biológica.

A **Tabela 4.2** apresenta os resultados obtidos para os modelos de classificação DD-SIMCA para autenticação dos ovos de galinha em pó frente a sua adulteração

com proteína de soja e amido de milho, em termos de sensibilidade, especificidade e acurácia para os diferentes modelos de cor obtidos a partir das imagens digitais.

Tabela 4.2 - Resultados dos parâmetros de desempenho sensibilidade, especificidade e acurácia para DD-SIMCA nos conjuntos de treinamento e teste, utilizando os histogramas de cor dos diferentes modelos de cor provenientes das imagens digitais.

Método de classificação		DD-SIMCA Instrumento NIR portátil			
Modelo de cor	Conjunto de amostra	PCs	Sensibilidade	Especificidade	Acurácia
GS	Treinamento	8	84,9	-	56,9
	Teste		73,9	54,0	
RGB	Treinamento	9	84,9	-	81,9
	Teste		60,9	85,4	
HSV	Treinamento	9	94,3	-	81,3
	Teste		82,6	81,0	
GS+RGB	Treinamento	7	92,5	-	63,1
	Teste		65,2	66,4	
GS+HSV	Treinamento	9	90,6	-	85,6
	Teste		87,0	85,4	
RGB+HSV	Treinamento	10	90,6	-	89,4
	Teste		78,3	91,2	
GS+RGB+HSV	Treinamento	8	94,3	-	72,5
	Teste		78,3	71,5	

PCs: componentes principais; **GS:** Escala de cinzas; **RGB:** Vermelho, verde e azul; **HSV:** Matiz, saturação e valor.

Conforme visualiza-se na **Tabela 4.2**, para as imagens digitais, o modelo de cor RGB+HSV se destacou como o mais eficaz no método de classificação DD-SIMCA. Este modelo demonstrou uma sensibilidade elevada de 90,6% no treinamento e uma sensibilidade (78,3), especificidade (91,2%) e acurácia (89,4%) no teste. Em contrapartida, os modelos individuais como GS e RGB mostraram desempenhos inferiores.

Apesar de apresentar uma sensibilidade, especificidade e acurácia inferior, quando comparadas com os modelos DD-SIMCA utilizando os espectros NIR portátil, as imagens digitais podem ser utilizadas como ferramenta *screening* para avaliar a qualidade dos ovos de galinha em pó quanto a uma possível adulteração, promovendo um método simples, acessível e de baixo custo.

A **Figura 4.5** mostra o gráfico de área de aceitação para o DD-SIMCA nos conjuntos de treinamento e teste para o pré-processamento DSG-1^aD-2^aO-7P nos espectros do NIR portátil (**a** e **b**) e os histogramas do modelo de cor RGB+HSV (**c** e **d**) obtido a partir das imagens digitais dos ovos de galinha em pó.

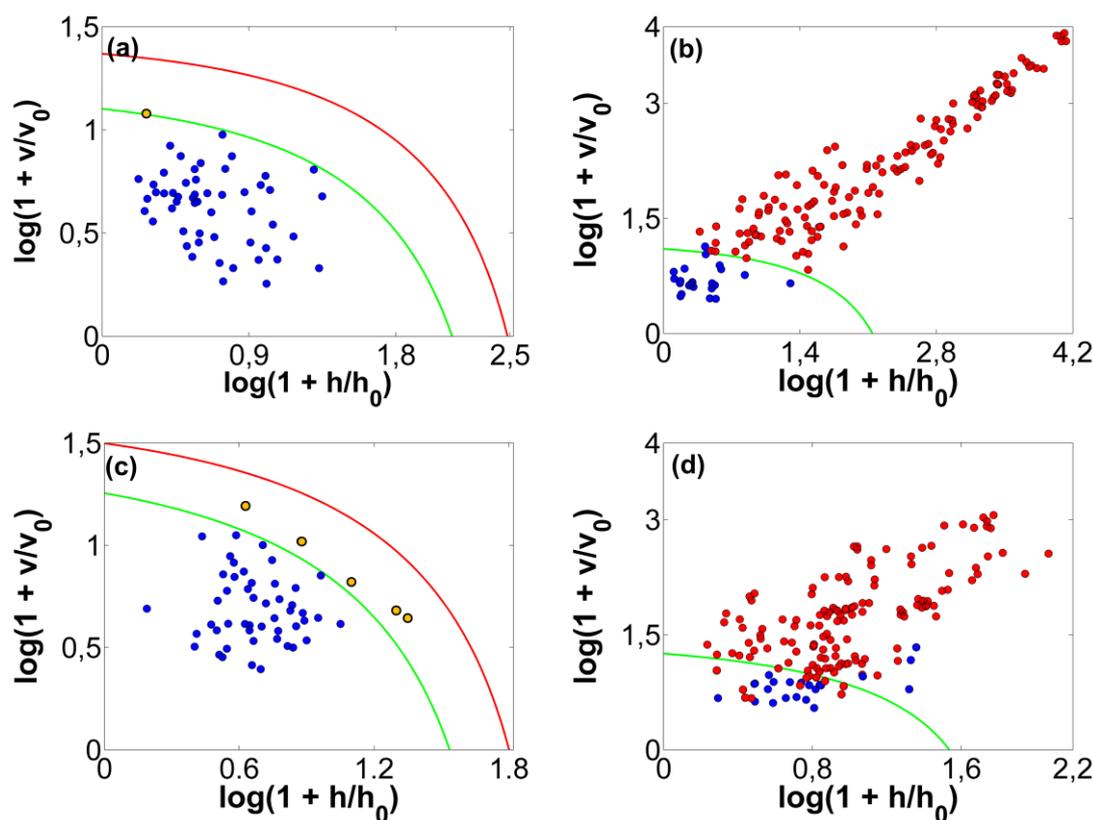


Figura 4.5 – Gráficos de área de aceitação para o conjunto de treinamento (a, c) e teste (b, d) para o algoritmo DD-SIMCA desenvolvidos utilizando os espectros do NIR portátil (a, b) e os histogramas de cor (c, d), das amostras de ovo em pó puras (●) e adulteradas (●). A linha verde (–) representam a borda entre as amostras regulares (puras (●) e extremas (●); a linha vermelha (–) é a fronteira da área de aceitação.

Foi escolhido a escala logarítmica para melhorar a visualização do comportamento do modelo (Zontov et al., 2017). Assim, o gráfico de distância dos scores *versus* as distâncias ortogonais nos permite identificar se uma amostra está ou não adulterada. A combinação dessas duas distâncias resulta em uma distância global, a qual é representada por uma variável c , que segue uma distribuição qui-quadrado e fornece duas áreas com limiares extremos e outliers (Zontov et al., 2017).

Nesse sentido, caso as amostras apresentem valores menores que c , isto é, abaixo da linha verde (Figura 4.5), são consideradas como puras, já aquelas com valores entre as linhas verde e vermelha são consideradas amostras extremas e as que possuem valores maiores que c (acima da linha vermelha) são possíveis *outliers* (Rodionova et al., 2014; Zontov et al., 2017). Nessa perspectiva, ao analisar a Figura 4.5, observa-se a maioria das amostras se posicionaram dentro da área de aceitação do modelo para um valor de significância de 0,05, apenas uma amostra na Figura 4.5 (a) e 5 na Figura 4.5 (c), as quais foram extremas para os modelos de DD-SMICA

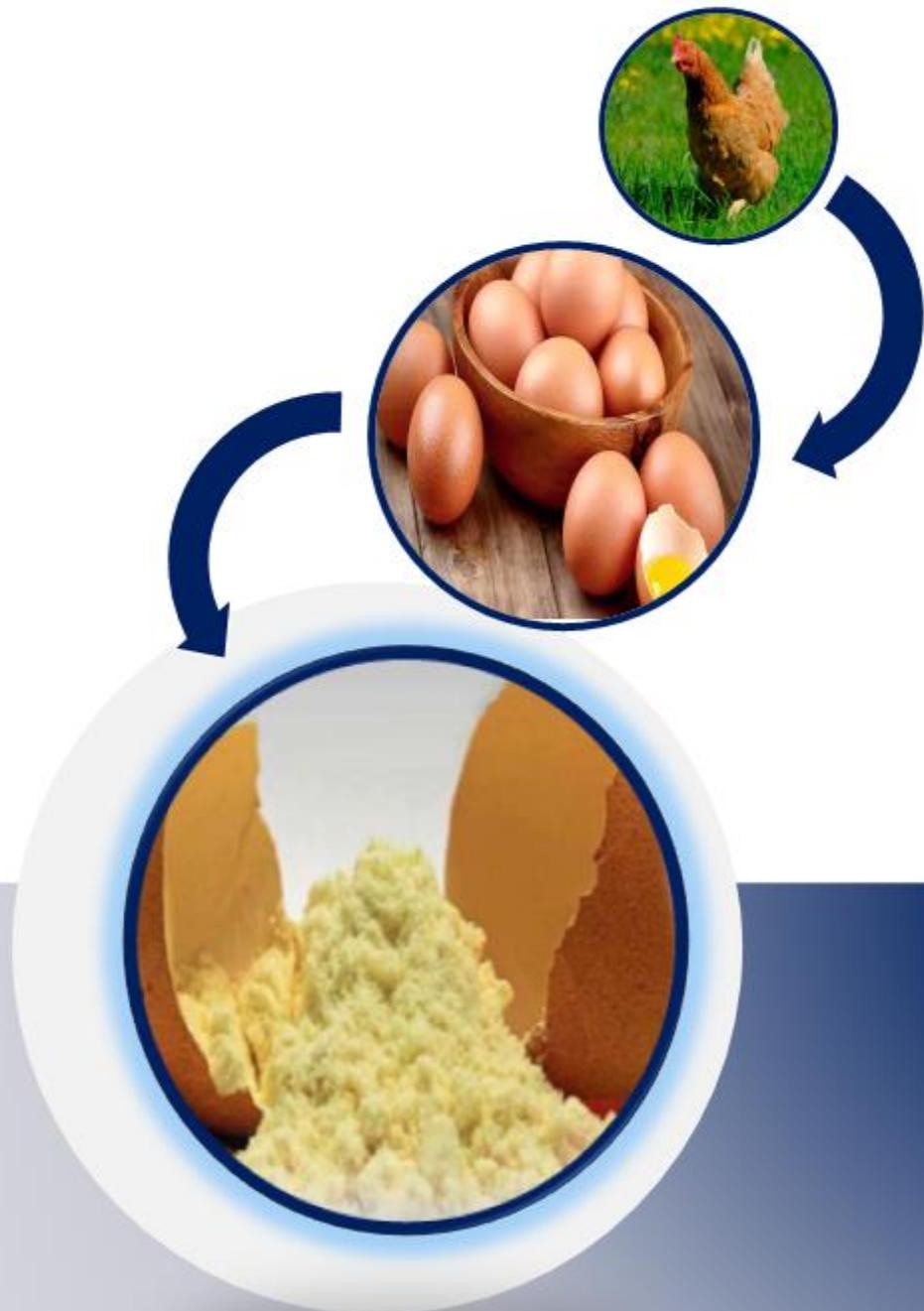


utilizando o NIR portátil e imagens digitais, respectivamente. Em contrapartida, para o NIR portátil todas as amostras adulteradas se apresentaram fora da área de aceitação, resultando em uma taxa de especificidade de 100%. Isso mostra a boa capacidade do modelo em detectar novas amostras adulteradas com proteína de soja e amido de milho.

Nesse sentido, nota-se que ambos os modelos DD-SIMCA construídos para ambos os instrumentos foram capazes de identificar corretamente a maior parte das amostras testadas, indicando que pode ser utilizada como uma alternativa para detectar a adulteração de ovos de galinha em pó. Além disso, os instrumentos são portáteis e de baixo custo, podendo ser utilizado pelos órgãos fiscalizadores.

Capítulo 5

Conclusões e perspectivas





5 CONCLUSÕES E PERSPECTIVAS

5.1 Conclusões

No presente trabalho de tese foram desenvolvidos com sucesso modelos quimiométricos utilizando dados provenientes de imagens digitais e instrumentos comerciais NIR para determinação de umidade e teor de proteínas e fósforos totais, bem como a autenticação de ovos de galinha em pó. As imagens digitais foram capturadas utilizando um *scanner*, já os espectros NIR foram registrados usando espectrômetros de bancada e portátil.

No primeiro estudo, foram avaliadas diferentes técnicas de pré-processamento de dados e algoritmos de calibração multivariada (*full*-PLS, *i*PLS, SPA-MLR e *i*SPA-PLS) para determinar umidade e teor de proteínas e sólidos fósforos em amostras de ovos de galinha em pó. Os resultados mostraram que os métodos de calibração multivariada em termos de r e RPD para o teor de proteínas totais foram melhores usando os histogramas de cor provenientes das imagens digitais do que aqueles obtidos por espectros NIR de bancada e portáteis ($r = 0,995$ e $RPD = 9,55$). Por outro lado, para predição de umidade ($r = 0,974$ e $RPD = 4,35$) e teor de fósforo total ($r = 0,98$ e $RPD = 4,3$), a utilização do equipamento NIR de bancada produziu melhores resultados quando comparados com o instrumento portátil e as imagens digitais.

A utilização da fusão de dados produziu uma performance ligeiramente inferior aos melhores métodos desenvolvidos utilizando espectros NIR pré-processados ou imagens digitais. Em todos os melhores modelos, o SPA permitiu a obtenção de modelos que quantificaram adequadamente umidade e teor de proteínas e fósforos totais quando acoplado ao MLR ou *i*PLS. Verificou-se ainda que o número de intervalos selecionados foi inferior às divisões de intervalos realizadas. Isso permitiu a obtenção de modelos mais parcimoniosos, uma vez que eliminou as variáveis espectrais ou de cor não informativas.

No segundo estudo, foram avaliadas as imagens digitais, NIR portátil, diferentes técnicas de pré-processamento de dados e a utilização do algoritmo DD-SIMCA para autenticar amostras de ovos de galinha em pó quanto a sua adulteração com proteína de soja e amido de milho. Verificou-se que entre os vários métodos construídos, a utilização do pré-processamento Savitsky-Golay com 1ª derivada, polinômio de 2ª ordem e janela de 7 ponto apresentou a melhor capacidade preditiva.

Como resultado, todas as amostras adulteradas foram classificadas corretamente e apenas 1 amostra pura foi classificada como adulterada no conjunto de teste, permitindo ao modelo ajustado resultados de sensibilidade de 95,7%, especificidade de 100% e acurácia 99,4%. Notou-se que os melhores métodos foram aqueles construídos com o NIR portátil quando comparados com as imagens digitais.

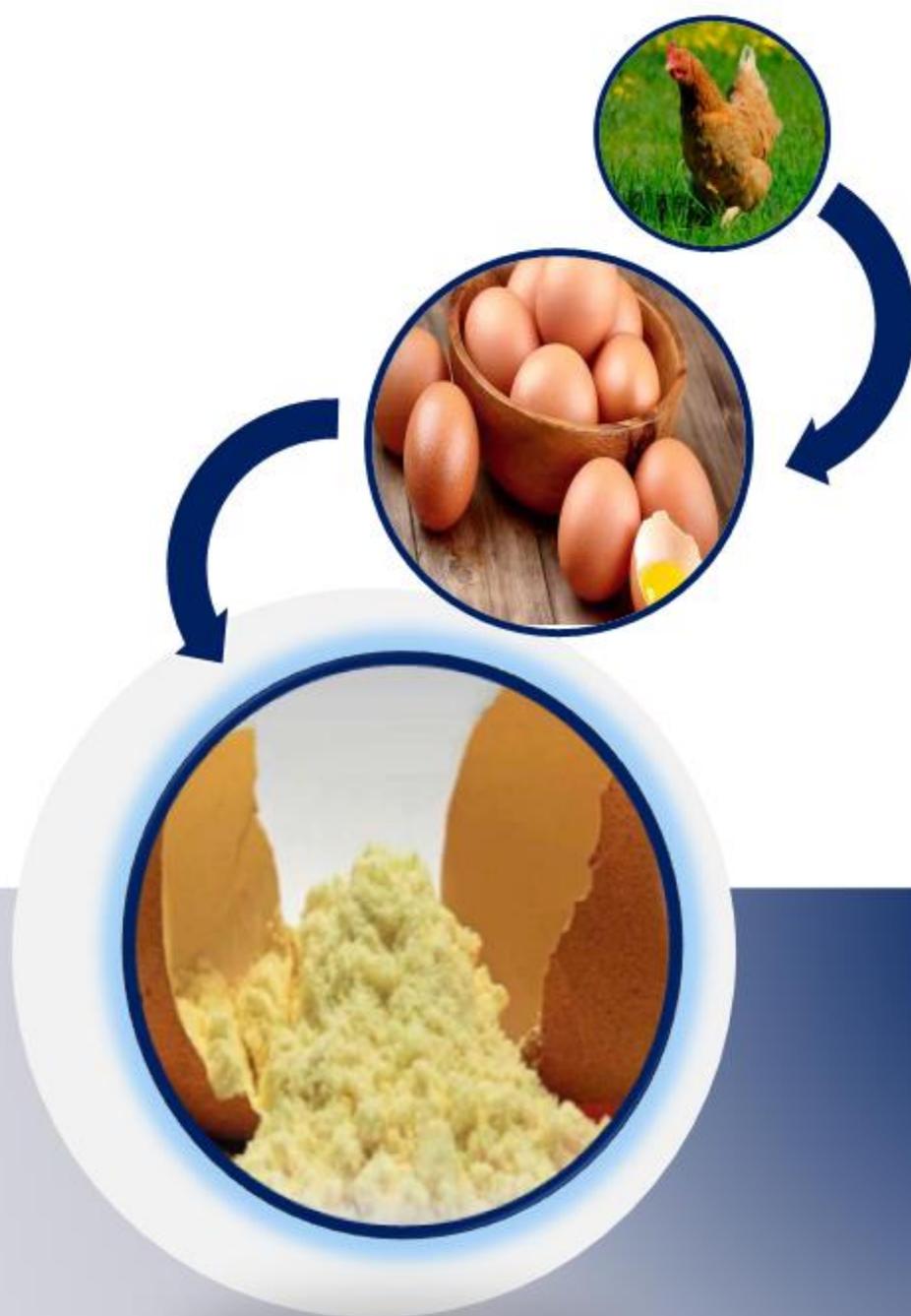
Portanto, os métodos propostos podem ser considerados uma alternativa satisfatória aos métodos de referência padrão para análises de rotina, pois não utilizam reagentes químicos, reduzem o tempo de análise e evitam a geração de resíduos, alinhando-se aos princípios básicos da Química Analítica Verde.

5.2 Perspectivas

Embora esses os métodos desenvolvidos nesse trabalho de tese tenham promovido resultados satisfatórios, as seguintes propostas podem ser vislumbradas futuramente:

- Quantificar o teor de colesterol nos ovos de galinha em pó utilizando os instrumentos NIR de bancada e portátil e imagens digitais;
- Utilizar técnicas discriminantes, como PLS-DA e iSPA-PLS-DA, para identificar o tipo de adulterante (proteína de soja e amido de milho) em ovos de galinha em pó;
- Desenvolver métodos de calibração multivariada para quantificar os teores de amido de milho e proteína de soja em ovos de galinha em pó;
- Desenvolver métodos quimiométricos para autenticação e calibração de ovos de galinha em pó a partir da transferência de calibração.

Referências



REFERÊNCIAS

ARSLAN, M., ZAREEF, M., TAHIR, H. E., XIAODONG, Z., RAKHA, A., ALI, S., SHI, J., XIAOBO, Z. Simultaneous quantitation of free fatty acid in rice by synergetic data fusion of colorimetric sensor arrays, NIR, and MIR spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 292, p. 122359, 2023.

AWADE, A. C., EFSTATHIOU, T., Théo. Comparison of three liquid chromatographic methods for egg-white protein analysis. **Journal of Chromatography B: Biomedical Sciences and Applications**, v. 723, n. 1-2, p. 69-74, 1999.

ABPA - ASSOCIAÇÃO BRASILEIRA DE PROTEÍNA ANIMAL. **Relatório Anual**. 2022. Disponível em: <https://abpa-br.org/wp-content/uploads/2022/05/Relatorio-Anual-ABPA-2022-1.pdf>. Acesso em: 22 set. 2022.

ABREU, E. S.; SPINELLI, M. G. N. **Seleção e preparo de alimentos: gastronomia e nutrição**. São Paulo: Editora Metha, p. 411, 2014.

ALEIXANDRE-TUDO, J. L.; CASTELLO-COGOLLO, L.; ALEIXANDREA, J. L.; ALEIXANDRE-BENAVENT, R. Chemometrics in food science and technology: A bibliometric study. **Chemometrics and Intelligent Laboratory Systems**, v. 222, p. 1-12, 2022.

ALMEIDA, V. E. **Algoritmo das projeções sucessivas associado ao Kernel-PLS para calibração multivariada não linear**. 2017. 93 f. Dissertação (Mestrado em Química) - Universidade Federal da Paraíba, João Pessoa, 2017.

AMARAL, G. F.; GUIMARÃES, D. D.; NASCIMENTO, J. C. O. F.; CUSTODIO, S. **Avicultura de postura: estrutura da cadeia produtiva, panorama do setor no Brasil e no mundo e o apoio do BNDES**. Rio de Janeiro: BNDES Setorial, 2016. p. 167-207.

ANANDHARAMAKRISHNAN, C. **Handbook of drying for dairy products**. 1. ed. John Wiley & Sons, 2017, 336 p.

AOAC, Association of Official Analytical Chemists. **Official Methods of Analysis of AOAC International**, 16th edition. Method 995.11, Phosphorus (Total) in Foods, Colorimetric Method. AOAC International, Gaithersburg, Maryland, USA, 1995.

ARAÚJO, W. M. C.; MONTEBELLO, N. P.; BOTELHO, R. B. A.; BORGIO, L. A. **Alquimia dos alimentos**. 3. ed. Brasília: Editora Senac – DF, 2014. p. 321.

ARAÚJO, T. K. L., NÓBREGA, R. O., SOUSA FERNANDES, D. D., ARAÚJO, M. C. U., DINIZ, P. H. G. D., SILVA, E. C.. Non-destructive authentication of Gourmet ground roasted coffees using NIR spectroscopy and digital images. **Food Chemistry**, v. 364, p. 130452, 2021.

ATTIA, Y. A., AL-HARTHI, M. A., SHIBOUB, M. M. Evaluation of quality and nutrient contents of table eggs from different sources in the retail market. **Italian Journal of Animal Science**, v. 13, n. 2, p. 3294, 2014.

ÁVILA, V. S.; SOARES, J. P. G. **Produção de ovos em sistemas orgânicos**. 2. ed. Concórdia: Embrapa Suínos e Aves, 2010. 100 p.

BIRGHILA, S., BRATU, M. M., PRAJITURA, C., RONCEA, F. N., NEGREANU-PIRJOL, T. Spectrophotometric method for the determination of total proteins in egg white samples. **Rev. Chim.(Bucharest)**, v. 66, p. 378-381, 2015.

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

BELITZ, H. D; GROSCH, W.; SCHIEBERLE, P. **Food chemistry**. 4. ed. Berlin: Springer-Verlag, 2009, 1070 p.

BELYAVIN, C.G. Eggs: Use in the Food Industry. **Encyclopedia of Food and Health**, 2016, pp. 476-479.

BERRAR, D. Cross-Validation. **Journals & Books**, v. 1, p. 542-545, 2019.

BERTECHINI, A. G. Mitos e verdades sobre o ovo de consumo. In: CONFERÊNCIA APINCO DE CIÊNCIA E TECNOLOGIA AVÍCOLAS, 2003, Campinas. **Anais...** Campinas: FACTA, 2003.

BHISE, P. R.; KULKARNI, S. B. Review on analysis and classification techniques of soil study in remote sensing and geographic information system. **Int J Emerg Trends Technol Comput Sci (IJETTCS)**, v. 6, n. 1, p. 124-138, 2017.

BRERETON, R. G. Pattern recognition in chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 149, p. 90-96, 2015.

BIOHAZ. EFSA PANEL ON BIOLOGICAL HAZARDS. Scientific Opinion on the public health risks of table eggs due to deterioration and development of pathogens. **EFSA Journal**, v. 12, n. 7, p. 3782, 2014.

BLOTTA, E.; BOUCHET, A.; BALLARIN, V.; PASTORE, J. Enhancement of medical images in HSI color space. In: Journal of Physics: Conference Series. **IOP Publishing**, v. 332, p. 1-8, 2011.

BONO, V. **Characterization of fibrous fractions from wine industry by-products and their use in baked goods**. 2014. 159 f. Tese (Doutorado em Inovação Tecnológica para a Ciência Agroalimentar e Ambiental) - Università Degli Studi di Milano, Milão, 2014. Disponível em: <https://air.unimi.it/retrieve/dfa8b991-8a12-748b-e053-3a05fe0a3a96/phd_unimi_R09605_01.pdf>. Acesso em 10 set. 2022

BRIDGES, M. **Phosphorus in diet**. Medlineplus. 2021. Disponível em: <<https://medlineplus.gov/ency/article/002424.htm>>. Acesso em 10 set. 2022

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Portaria nº 1, de 21 de fevereiro de 1990**. Aprovar as Normas Gerais de Inspeção de Ovos e Derivados, propostas pela Divisão de Inspeção de Carnes e Derivados. Diário Oficial da União, Brasília, DF, 21 fev. 1990.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Portaria nº 612, de 6 de julho de 2022**. Aprova os requisitos de instalações, equipamentos e os procedimentos para o funcionamento de granjas avícolas e de unidades de beneficiamento de ovos e derivados a registradas no Departamento de Inspeção de Produtos de Origem Animal. Diário Oficial da União, Brasília, DF, 11 jul. 2022.

BRASIL. Presidência da República. **Decreto nº 9.013, de 29 de março de 2017**. Regulamenta a Lei nº 1.283, de 18 de dezembro de 1950, e a Lei nº 7.889, de 23 de novembro de 1989, que dispõem sobre a inspeção industrial e sanitária de produtos de origem animal. Diário Oficial da União, Brasília, DF, 29 mar. 2017.

BRERETON, R. G.; JANSEN, J.; LOPES, J.; F.; POMERANTSEV, A.; RODIONOVA, O.; ROGER, J. M.; WALCZAK, B.; TAULER, R. Chemometrics *in analytical chemistry-part I*: history, experimental design and data analysis tools. *Analytical and Bioanalytical Chemistry*, v. 409, n. 25, p. 5891–5899, 2017.

CABALLERO-AGOSTO, E.R., SIERRA-VEGA, N. O., ROLON-OCASIO, Y., HERNANDEZ-RIVERA, S. P., INFANTE-DEGRÓ, R. A., FONTALVO-GOMEZ, M., PACHECO-LONDOÑO, L. C., INFANTE-CASTILHO, R. Detection and Quantification of Corn Starch and Wheat Flour As Adulterants in Milk Powder by Near- and Mid-infrared Spectroscopy Coupled with Chemometric Routines. **Food Chemistry Advances**, v. 4, p. 100582, 2024.

CARVALHO, F. **Construção e avaliação de desempenho de um spray dryer piloto**. 2010. 167 p. Tese (Doutorado em Ciência dos Alimentos) - Universidade Federal de Lavras, Lavras, 2012.

CHANG, C.; LAHTI, T.; TANAKA, T.; NICKERSON, M. T. Egg proteins: Fractionation, bioactive peptides and allergenicity. **Journal of the Science of Food and Agriculture**, v. 98, n. 15, p. 5547-5558, 2018.

CHERNEY, J. H.; DIGMAN, M. F.; CHERNEY, D. J. Handheld NIRS for forage evaluation. **Computers and Electronics in Agriculture**, v. 190, p. 1-8, 2021.

CHIN, W. W. The partial least squares approach to structural equation modeling. **Modern methods for business research**, v. 295, n. 2, p. 295-336, 1998.

CHOWDHURY, M. Z. I.; TURIN, T. C. Variable selection strategies and its importance in clinical prediction modelling. **Family medicine and community health**, v. 8, n. 1, 2020.

COOPER, K.; GOSNELL, K. **Foundations of Nursing**. Elsevier Health Sciences, p. 1378, 2018.

DEIDDA, R., SACRE, P. Y., CLAVAUD, M., COIĆ, L., AVAHOU, H., HUBERT, P., ZIEMONS, E. Vibrational spectroscopy in analysis of pharmaceuticals: Critical review of innovative portable and handheld NIR and Raman spectrophotometers. **TrAC Trends in Analytical Chemistry**, v. 114, p. 251-259, 2019.

DEVIANTI; SUFARDI; BULAN, R.; SITORUS, A. Vis-NIR spectra combined with machine learning for predicting soil nutrients in cropland from Aceh Province, Indonesia. **Case Studies in Chemical and Environmental Engineering**, v. 6, p. 100268, 2022.

DI DOMENICO, C. N. B.; CONRAD, T. M. Simulação de processos de secagem através dos modelos matemáticos exponencial e de page. **Vivências: Revista Eletrônica de Extensão da URI**, v. 11, n. 20, p. 134-146, 2015.

DINIZ, P. H. G. D. Chemometrics-assisted color histogram-based analytical systems. **Journal of Chemometrics**, v. 34, n. 12, p. 1-27, 2020.

DOMENE, S. M. A. **Técnica dietética: teoria e aplicações**. 2. ed. Rio de Janeiro: Ed. Guanabara Koogan, 2018. 280 p.

DUARTE, E. S. A., ALMEIDA, V. E., COSTA, G. B., ARAÚJO, M. C. U., VÉRAS, G., DINIZ, P. H. G. D., FERNANDES, D. D. S. Feasibility study on quantification and authentication of the cassava starch content in wheat flour for bread-making using NIR spectroscopy and digital images. **Food Chemistry**, v. 368, p. 130843, 2022.

DUKES, H. H.; REECE, W. O. **Fisiologia dos animais domésticos**. 12. ed. Rio de Janeiro: Guanabara Koogan, 2006. 954 p.

DZIKI, D.; POLAK, R.; RUDY, S.; KRZYKOWSKI, A.; GAWLIK-DZIKI, U.; RÓŻYŁO, R.; MIŚ, A.; COMBRZYŃSKI, M. Simulation of the process kinetics and analysis of physicochemical properties in the freeze drying of kale. **International Agrophysics**, v. 32, n. 1, p. 49, 2018.

FAN, Y., LI, J., GUO, Y., XIE, L., ZHANG, G. Digital Image Colorimetry on Smartphone for Chemical Analysis: A Review. **Measurement**, p. 1-10, 2021.

FAO - ORGANIZAÇÃO PARA A ALIMENTAÇÃO E AGRICULTURA. **Definitions for the purposes of the codex alimentarius**. Disponível em: <<https://www.fao.org/3/W5975E/w5975e07.htm>>. Acesso em: 10 set. 2022.

FERNANDES, D. D. S.; GOMES, A. A.; COSTA, G. B. D.; SILVA, G. W. B. D.; VÉRAS, G. Determination of biodiesel content in biodiesel/diesel blends using NIR and visible spectroscopy with variable selection. **Talanta**, v. 87, p. 30-34, 2011.

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. 1. ed. São Paulo: Editora da UNICAMP, 2015.

GIAIO, E. N., MARTINS, V. L., LYRA, W. S., ALMEILDA, L. F., SILVA, E. C., ARAÚJO, M. C. U. Digital Image-based Titrations. **Analytica Chimica Acta**, v. 570, n. 2, p. 283-290, 2006.

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; FRAGOSO, W. D.; SILVA, E. C.; JOSÉ, G. E.; SOARES, S. F. C.; PAIVA, H. M. A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. **Chemometrics and intelligent laboratory systems**, v. 92, n. 1, p. 83-91, 2008.

GALVÃO, R. K. H.; ARAUJO, M. C. U.; JOSÉ, G. E.; PONTES, M. J. C.; SILVA, E. C.; SALDANHA, T. C. B. A method for calibration and validation subset partitioning. **Talanta**, v. 67, n. 4, p. 736-740, 2005

GALVÃO, R. K. H., ARAÚJO, M. C. U., SOARES, S. F. C. Linear Regression Modeling: Variable Selection. *In*: BROWN, S., TAULER, R., WALCZAK **Comprehensive Chemometrics** (org.). 2. ed. Amesterdã: Elsevier, 2020. p. 249-293.

GAUTAM, R.; VANGA, S.; ARIESE, F.; UMAPATHY, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. **EPJ Techniques and Instrumentation**, v. 2, n. 1, p. 1-38, 2015.

GERRETZEN, J.; SZYMAŃSKA, E.; JANSEN, J. J.; BART, J.; MANEN, H. J. V.; HEUVEL, E. R. V. D; BUYDENS, L. M. C. Simple and effective way for data preprocessing selection based on design of experiments. **Analytical Chemistry**, v. 87, n. 24, p. 12096-12103, 2015.

GÉSAN-GUIZIOU, Geneviève. Separation technologies in dairy and egg processing. *In*: **Separation, extraction and concentration processes in the food, beverage and nutraceutical industries**. Woodhead Publishing, 2013. p. 341-380.

GODOI, M. **Conheça melhor as principais extensões de arquivos de imagens**, 2013. Disponível em: <<https://www.desenhoonline.com/site/conheca-melhor-as-principais-extensoes-de-arquivos-de-imagens/>>. Acesso em 10 set. 2022

GOMES, A. A.; AZCARATE, S. M.; DINIZ, P. H. G. D.; FERNANDES, D. D. S.; VERAS, G. Variable selection in the chemometric treatment of food data: A tutorial review. **Food Chemistry**, v. 370, p. 131072, 2022.

GOMES, A. A.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; VÉRAS, G.; DA SILVA, E. C. The successive projections algorithm for interval selection in PLS. **Microchemical Journal**, v. 110, p. 202-208, 2013.

GONZALEZ, R.C., WOODS, R.E. **Digital Image Processing**, 4th ed., Addison-Wesley Publishing, Boston, 2018.

HAN, B., SONG, Q., LI, C., CHEN, Y., ZHOU, M. Fermi resonance: A new way to investigate the planarization of p-quaterphenyl under high pressure. **Chemical Physics Letters**, v. 756, p. 1-5, 2020.

HAYES, E., GREENE, D., & O'DONNELL, C., O'SHEA, N. Spectroscopic technologies and data fusion: Applications for the dairy industry. **Frontiers in Nutrition**, v. 9, p. 1074688, 2023.

HÉBERGER, Karoly. Chemoinformatics—multivariate mathematical—statistical methods for data evaluation. In: **Medical Applications of Mass Spectrometry**. Elsevier, 2008. p. 141-169.

HE, H. J., WANG, Y., ZHANG, M., WANG, Y., OU, X., GUO, J. Rapid determination of reducing sugar content in sweet potatoes using NIR spectra. **Journal of Food Composition and Analysis**, v. 111, p. 104641, 2022.

HOSSEINI, E., GHASEMI, J. B., DARAEI, B., ASADI, G., ADIB, N.. Near-infrared spectroscopy and machine learning-based classification and calibration methods in detection and measurement of anionic surfactant in milk. **Journal of Food Composition and Analysis**, v. 104, p. 104170, 2021.

HUANG, L., ZHAO, W., LIEW, A. W. C., YOU, Y. An evidential combination method with multi-color spaces for remote sensing image scene classification. **Information Fusion**, v. 93, p. 209-226, 2023.

INOBEME, A., NAYAK, V., MATHEW, T. J., OKONKWO, S., EKWOBA, L., AJAI, A. I., SINGH, K. R. Chemometric approach in environmental pollution analysis: A critical review. **Journal of Environmental Management**, v. 309, p. 114653, 2022.

INSTITUTO ADOLFO LUTZ. **Normas analíticas do Instituto Adolfo Lutz**: métodos químicos e físicos para análise de alimentos. 4.ed. São Paulo: Instituto Adolfo Lutz, 2008. p. 1020.

JAIN, A., PRADHAN, B. K., MAHAPATRA, P., RAY, S. S., CHAKRAVARTY, S., PAL, K. Development of a low-cost food color monitoring system. **Color Research & Application**, v. 46, n. 2, p. 430-445, 2021.

JOLLIFFE, I. T., CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, v. 374, n. 2065, p. 20150202, 2016.

KAFLE, B. P. Chapter 7 – Infrared (IR) spectroscopy. *In: KAFLE, B. P. Chemical Analysis and Material Characterization by Spectrophotometry*. Elsevier, 2020, p. 199-243.

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

KHALEGHI, B., KHAMIS, A., KARRAY, F. O., RAZAVI, S. N. Multisensor data fusion: A review of the state-of-the-art. *Information fusion*, v. 14, n. 1, p. 28-44, 2013.

KREPPER, G., ROMEO, F., FERNANDES, D. D. S., DINIZ, P. H. G. D., ARAÚJO, M. C. U., DI NEZIO, M. S., CENTURIÓN, M. E. Determination of fat content in chicken hamburgers using NIR spectroscopy and the Successive Projections Algorithm for interval selection in PLS regression (iSPA-PLS). *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 189, p. 300-306, 2018.

KÖNIG, D. L. **Liofilização aplicada a produtos cárneos avícolas**. 2016. 42 f. TCC (Graduação) - Curso de Tecnologia em Alimentos, Universidade Tecnológica Federal do Paraná, Medianeira, 2016.

KRAEMER, F. B.; HÜTTEN, G. C.; TEIXEIRA, C. E.; PARDI, H. S.; MANO, S. Avaliação da qualidade interna de ovos em função da variação da temperatura de armazenamento. *Revista Brasileira de Ciência Veterinária*, v. 10, n. 3, 2010.

LAVINE, B.; WORKMAN, J. Chemometrics. *Analytical Chemistry*, v. 82, n. 12, p. 4699-4711, 2010.

LEONARDI, J. G.; AZEVEDO, B. M. Métodos de conservação de alimentos. *Revista Saúde em foco*, v. 10, n. 1, p. 51-61, 2018.

LI, Y., HUANG, Y., XIA, J., XIONG, Y., MIN, S. Quantitative analysis of honey adulteration by spectrum analysis combined with several high-level data fusion strategies. *Vibrational Spectroscopy*, v. 108, p. 103060, 2020.

LI, L., GENG, S.; LIN, D.; SU, G.; ZHANG, Y.; CHANG, L.; JI, Y.; WANG, Y.; WANG, L. Accurate modeling of vertical leaf nitrogen distribution in summer maize using in

situ leaf spectroscopy via CWT and PLS-based approaches. **European Journal of Agronomy**, v. 140, p. 126607, 2022.

LIU, C.; WANG, Q.; LIN, W.; YU, C. Origins classification of egg with different storage durations using FT-NIR: A characteristic wavelength selection approach based on information entropy. **Biosystems Engineering**, v. 222, p. 82-92, 2022.

LIU, Y., ZHANG, Z., HU, L. High Efficient Freeze-Drying Technology in Food Industry. **Critical Reviews in Food Science and Nutrition**, v. 62, p.3370-88, 2021.

LYRA, W. S. **Um método espectrométrico de emissão em chama baseado em imagens digitais para determinação indireta de fármacos e determinação simultânea de sódio e cálcio**. 2012. 109 f. Tese (Doutorado em Química) - Universidade Federal da Paraíba, João Pessoa, 2012.

LYRA, W. S., SHACHES, F. A. C., CUNHA, F. A., DINIZ, P. H. G. D., LEMOS, S. G., SILVA, E. C., ARAÚJO, M. C. U. Indirect determination of sodium diclofenac, sodium dipyrone and calcium gluconate in injection drugs using digital image-based (webcam) flame emission spectrometric method. **Analytical Methods**, v. 3, n. 9, p. 1975-1980, 2011.

MAHGOUB, Y. A., SHAWKY, E., ABDALLAH, I. I.. Authentication of oats and discrimination from their gluten-containing adulterants using NIR diffuse reflectance spectroscopy and multivariate analysis. **Food Bioscience**, v. 60, p. 104271, 2024.

MARCELINO NETTO, J. M., Honorato, F. A., Celso, P. G., & Pimentel, M. F. Authenticity of almond flour using handheld near infrared instruments and one class classifiers. **Journal of Food Composition and Analysis**, v. 115, p. 104981, 2023.

MANCINI, F. R.; AFFRET, A.; DOW, C.; BALKAU, B.; CLAVEL-CHAPELON, F.; BONNET, F.; BOUTRON-RUAULT, M.; FAGHERAZZI, G. High dietary phosphorus intake is associated with an increased risk of type 2 diabetes in the large prospective E3N cohort study. **Clinical Nutrition**, v. 37, n. 5, p. 1625-1630, 2018.

MEDINA, S., PERESTRELO, R., SILVA, P., PEREIRA, J. A. M., CÂMARA, J. S. Current trends and recent advances on food authenticity technologies and chemometric approaches. **Trends in Food Science & Technology**, v. 85, p. 163-176, 2019.

MISHRA, P.; RUTLEDGE, D. N.; ROGER, J.; WALI, K.; KHAN, H. A. Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction. **Talanta**, v. 229, p. 1-7, 2021.

MOREIRA, L.N. **Técnica dietética**. 1. ed. Rio de Janeiro: SESES, 2016. 240 p.

MORENO, J.; DÍAZ-GÓMEZ, J.; FUENTES-FONT, L.; ANGULO, E.; GOSÁLVEZ, L.; SANDMANN, G.; PORTERO-OTIN, M.; CAPELL, T.; ZHU, C.; CHRISTOU, P.; NOGAREDA, C. Poultry diets containing (keto) carotenoid-enriched maize improve egg yolk color and maintain quality. **Animal Feed Science and Technology**, v. 260, p. 114334, 2020.

MU, Q., WANG, X., WEI, Y., LI, Z. Low and non-uniform illumination color image enhancement using weighted guided image filtering. **Computational Visual Media**, p. 1-18, 2021.

NELSON LEHNINGER, D.; COX, M. M. **Princípios de bioquímica de Lehninger**. 8ed. Porto Alegre: Artmed, 2022.

NEVES, M. G.; POPPI, R. J.; BREITKREITZ, M. C. Authentication of plant-based protein powders and classification of adulterants as whey, soy protein, and wheat using FT-NIR in tandem with OC-PLS and PLS-DA models. **Food Control**, v. 132, p. 108489, 2022.

NEVES, M. G.; POPPI, R. J. Authentication and identification of adulterants in virgin coconut oil using ATR/FTIR in tandem with DD-SIMCA one class modeling. **Talanta**, v. 219, p. 121-338, 2020.

NØRGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. **Applied spectroscopy**, v. 54, n. 3, p. 413-419, 2000.

NÓBREGA, R. O., DA SILVA, S. F., FERNANDES, D. D., LYRA, W. S., ARAÚJO, T. K., DINIZ, P. H. D., ARAÚJO, M. C.. Classification of instant coffees based on caffeine content and roasting degree using NIR spectrometry and multivariate analysis. **Microchemical Journal**, v. 190, p. 108624, 2023.

NÓBREGA, R. O. **Classificação de cafés solúveis usando espectroscopia NIR e quimiometria**. 2021. 77f. Dissertação (Programa de Pós-Graduação em Química - PPGQ) - Universidade Federal da Paraíba, João Pessoa, 2021.

NUNES, J. L. S. **Secagem, Beneficiamento e Armazenagem**. Agrolink. 2016. Disponível em: <https://www.agrolink.com.br/sementes/tecnologia-sementes/secagem--beneficiamento-e-armazenagem_361343.html> Acesso em 20 de set. 2022.

ONWUDE, D. I.; HASHIM, N.; JANIUS, R.; ABDAN, K.; CHEN, G.; OLADEJO, A. O. Non-thermal hybrid drying of fruits and vegetables: A review of current technologies. **Innovative Food Science & Emerging Technologies**, v. 43, p. 223-238, 2017.

ORNELLAS, L.H. **Técnica dietética: seleção e preparo de alimentos**. 7. ed. São Paulo: Editora Metha, 2001. 330 p.

OZAKI, Y., GENKAWA, T., FUTAMI, Y. Near-Infrared Spectroscopy. *In*: LINDON, J. C., TRANTER, G. E., KOPPENAAL. **Encyclopedia of Spectroscopy and Spectrometry**. Academic Press, 2017, p. 40-49.

OZAKI, Y., HUCK, C. W., BEĆ, K. B. Near-IR Spectroscopy and Its Applications. *In*: GUPTA, V. P., OZAKI, Y. **Molecular and Laser Spectroscopy**. Elsevier, 2018, p. 11-38.

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

PATEL, N., TOLEDO-ALVARADO, H., BITTANTE, G. Performance of different portable and hand-held near-infrared spectrometers for predicting beef composition and quality characteristics in the abattoir without meat sampling. **Meat Science**, v. 178, p. 108518, 2021.

PEREIRA, E. V. D. S.; FERNANDES, D. D. D. S.; DE ARAÚJO, M. C. U.; DINIZ, P. H. G. D.; MACIEL, M. I. S. Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms. **LWT**, v. 127, p. 109427, 2020.

PEREIRA, E. V. S., FERNANDES, D. D. S., ARAÚJO, M. C. U., DINIZ, P. H. G. D., MACIEL, M. I. S. In-situ authentication of goat milk in terms of its adulteration with cow milk using a low-cost portable NIR spectrophotometer. **Microchemical Journal**, v. 163, p. 105885, 2021.

PHILIPPI, S. T. **Nutrição e técnica dietética**. 3. ed. Barueri: Editora Manole, 2014, 424 p.

PUERTAS, G., VÁZQUEZ, M. Cholesterol determination in egg yolk by UV-VIS-NIR spectroscopy. **Food Control**, v. 100, p. 262-268, 2019.

PUERTAS, G., VÁZQUEZ, M. UV-VIS-NIR spectroscopy and artificial neural networks for the cholesterol quantification in egg yolk. **Journal of Food Composition and Analysis**, v. 86, p. 103350, 2020.

PUERTAS, G., CAZÓN, P., VÁZQUEZ, M. Application of UV-VIS-NIR spectroscopy in membrane separation processes for fast quantitative compositional analysis: A case study of egg products. **LWT**, v. 174, p. 114429, 2023.

RADY, A.; WATSON, N. J. Detection and quantification of peanut contamination in garlic powder using NIR sensors and machine learning. **Journal of Food Composition and Analysis**, v. 114, p. 104820, 2022.

RAMOS, B.F.S. **Gema de ovo composição em aminas biogênicas e influência da gema na fração volátil de creme de pasteleiro**. 2008. 111f. Dissertação (Mestrado em Controle de qualidade) – Faculdade de farmácia. Universidade do Porto, 2008.

REILE, C. G., RODRÍGUEZ, M. S., FERNANDES, D. D. S., GOMES, A. A., DINIZ, P. H. G. D., ANIBAL, C. V. D. Qualitative and quantitative analysis based on digital images to determine the adulteration of ketchup samples with Sudan I dye. **Food Chemistry**, v. 328, p. 1-8, 2020.

REY, L. Glimpses into the realm of freeze-drying: classical issues and new ventures. *In: Freeze-drying/lyophilization of pharmaceutical and biological products*. CRC Press, 2016. p. 15-42.

RIBEIRO, P. F. F. A. C. **Processo de Liofilização de Produtos Alimentares Perecíveis**. 2012. 117 f. Tese (Doutorado) - Curso de Engenharia Mecânica, Departamento de Engenharia Mecânica, Instituto Superior de Engenharia de Lisboa, Lisboa, 2012.

RIU, J.; RIUS, F. X. Assessing the accuracy of analytical methods using linear regression with errors in both axes. **Analytical Chemistry**, v. 68, n. 11, p. 1851-1857, 1996.

Rinnan, Å., Berg, F. V. D., & Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC Trends in Analytical Chemistry**, v. 28, n. 10, p. 1201-1222, 2009.

ROBERT, G.; GOSELIN, R. Evaluating the impact of NIR pre-processing methods via multiblock partial least-squares. **Analytica Chimica Acta**, v. 1189, p. 1-11, 2022.

RODIONOVA, O. Y.; BALKLOVA, K. S.; TITOVA, A. V.; POMERANTSEV, A. L. Quantitative risk assessment in classification of drugs with identical API content. **Journal of Pharmaceutical and Biomedical Analysis**, v. 98, p. 186-192, 2014.

ROWLANDS, D. A. Color conversion matrices in digital cameras: a tutorial. **Optical Engineering**, v. 59, n. 11, p. 1-36, 2020.

SÁEZ-HERNÁNDEZ, R., RUIZ, P., MAURI-AUCEJO, A. R., YUSA, V., CERVERA, M. L. Determination of acrylamide in toasts using digital image colorimetry by smartphone. **Food Control**, v. 141, p. 109163, 2022.

SANTOS FILHO, J. I.; MIELE, M.; MARTINS, F. M.; TALAMINI, D. J. D. Os 35 anos que mudaram a avicultura brasileira. **Sonho, Desafio e Tecnologia**, Embrapa, 2011.

CELESTINO, S. M. C. **Princípios de secagem de alimentos**. 1ª ed. Planaltina: EMBRAPA, 2010, 51 p.

SILVA, A. F. S., GONCALVES, I. C., ROCHA, F. R. Smartphone-based digital images as a novel approach to determine formaldehyde as a milk adulterant. **Food Control**, v. 125, p. 107956, 2021.

SINGH, S., SZOSTAK, R., CZARNECKI, M. A. Vibrational intensities and anharmonicity in MIR, NIR and Raman spectra of liquid CHCl_3 , CDCl_3 , CHBr_3 and CDBr_3 : Spectroscopic and theoretical study. **Journal of Molecular Liquids**, v. 336, p. 1-16, 2021.

SOARES, S. F. C.; GOMES, A. A.; ARAUJO, M. C. U.; FILHO, A. R. G.; GALVÃO, R. K. H. The successive projections algorithm. **TrAC Trends in Analytical Chemistry**, v. 42, p. 84-98, 2013.

STADELMAN, W. J.; COTTERILL, O. J. **Egg science and technology**. 4. ed. CRC Press, 2017, 592 p.

STEIDLE NETO, A. J., MOURA, L. D. O., LOPES, D. D. C., CARLOS, L. D. A., MARTINS, L. M., FERRAZ, L. D. C. L. Non-destructive prediction of pigment content in lettuce based on visible–NIR spectroscopy. **Journal of the Science of Food and Agriculture**, v. 97, n. 7, p. 2015-2022, 2017.

TEXAS INSTRUMENTS. **DLP® NIRscan™ Nano EVM User's Guide**, 2017.
Disponível em: <https://www.ti.com/lit/ug/dlpu030g/dlpu030g.pdf>. Acesso em: 16 out. 2021.

TONON, R. V., BARONI, A. F., HUBIGNER, M. D. Outros Processos de Desidratação. In: TADANI, C. C. (org). **Operações Unitárias na Indústria de Alimentos**. 2. ed. Rio de Janeiro: LTC, 2016.

UYSAL, R. S., BOYACI, I. H.. Authentication of liquid egg composition using ATR-FTIR and NIR spectroscopy in combination with PCA. **Journal of the Science of Food and Agriculture**, v. 100, n. 2, p. 855-862, 2020.

VISCONTI, L. G., VARGAS, S. M., RODRÍGUEZ, M. S., DI ANIBAL, C. V., DELRIEUX, C. Determination of grated hard cheese adulteration by digital image analysis and multivariate analysis. **International Dairy Journal**, v. 138, p. 105539, 2023.

VOHLAND, M., JUNG, A., REU, B., RAMIREZ, J. A. Quantification of non-structural carbohydrates in different tissues of trees with diffuse reflectance spectroscopy: Is there a benefit from a fusion of near-and mid-infrared data. **Talanta**, v. 269, p. 125406, 2024.

WALSH, G. **Proteins: biochemistry and biotechnology**. 2. ed. West Sussex: John Wiley & Sons, 2014.

WATANABE, L. S., BOVOLENTA, Y. R., JUNIOR, V. R. A., BARBIN, D. F., MADEIRA, T. B., NIXDORF, S. L. Investigation of NIR spectra pre-processing methods combined with multivariate regression for determination of moisture in powdered industrial egg. **Acta Scientiarum. Technology**, v. 40, 2018.

WARE, C. Chapter Four - Color. In: WARE, C. **Encyclopedia of Spectroscopy and Spectrometry**. Morgan Kaufmann, 2021, p. 96-141.

WESTAD, F., MARINI, F. Validation of chemometric models—a tutorial. **Analytica Chimica Acta**, v. 893, p. 14-24, 2015.

WIEDERSTEIN, M., BAUMGARTNER, S., LAUTER, K. Soybean (Glycine max) allergens—A Review on an Outstanding Plant Food with Allergenic Potential. **ACS Food Science & Technology**, v. 3, n. 3, p. 363-378, 2023.

WŁODARSKA, K., PIASECKI, P., LOBO-PRIETO, A., PAWLAK-LEMAŃSKA, K., GÓRECKI, T., SIKORSKA, E. Rapid screening of apple juice quality using ultraviolet, visible, and near infrared spectroscopy and chemometrics: A comparative study. **Microchemical Journal**, v. 164, p. 106051, 2021.

WORKMAN, J.; WEYER, L. **Practical guide to interpretive near-infrared spectroscopy**. CRC press, 2007.

XIAO, N.; ZHAO, N.; YAO, Y.; WU, N.; XU, M.; DU, H. TU, Y. Biological activities of egg yolk lipids: A review. **Journal of agricultural and food chemistry**, v. 68, n. 7, p. 1948-1957, 2020.

YANG, X., ZHU, L., HUANG, X., ZHANG, Q., LI, S., CHEN, Q., LI, J. Determination of the soluble solids content in korla fragrant pears based on visible and near-infrared spectroscopy combined with model analysis and variable selection. **Frontiers in Plant Science**, v. 13, p. 938162, 2022.

YU, H., LIU, H., WANG., RUTH, S. V. Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. **LWT**, v. 128, p. 1-8, 2020.

ZHAO, Q., LV, X., JIA, Y., CHEN, Y., XU, G., QU, L. Rapid Determination of the Fat, Moisture, and Protein Contents in Homogenized Chicken Eggs Based on Near-infrared Reflectance Spectroscopy. **Poultry Science**, v. 97, n. 6, pp. 2239-2245, 2018

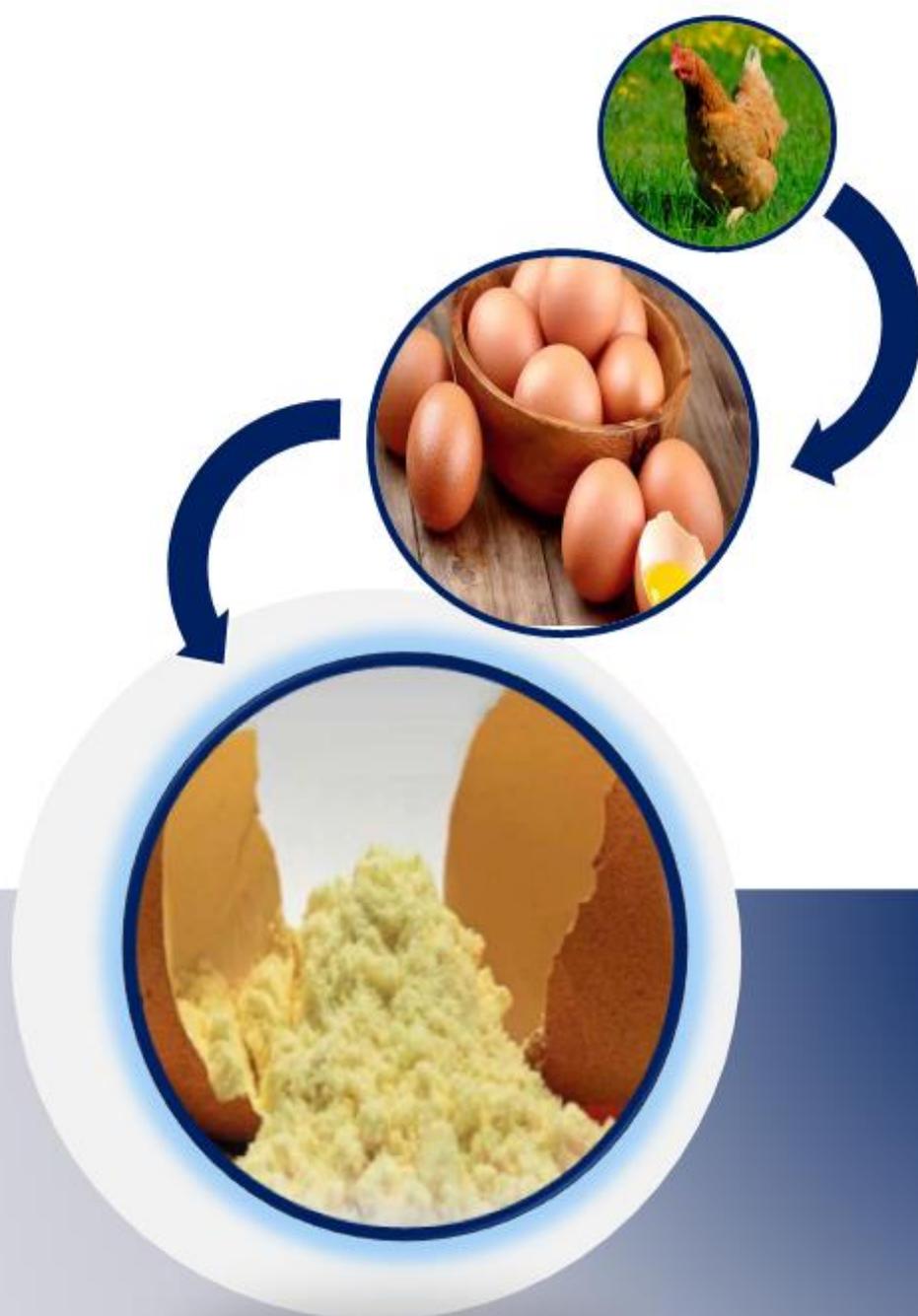
ZHOU, L., TAN, L., CHANG, C., ZHAO, N., HE, Y., QIU, Z. A portable NIR-system for mixture powdery food analysis using deep learning. **LWT**, p. 1-7, 2021.

ZHU, J., FAN, X., HAN, L., ZHANG, C., WANG, J., PAN, L., ZHANG, M. Quantitative analysis of caprolactam in sauce-based food using infrared spectroscopy combined with data fusion strategies. **Journal of Food Composition and Analysis**, v. 104, p. 104130, 2021.



ZONTOV, Y. V.; RODIONOVA, O.Y.; KUCHERYAVSKIV, S.V.; POMERANTSEY, A.L. DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach. **Chemometrics and Intelligent Laboratory Systems**, v. 167, p. 23–28, 2017.

Anexo



ANEXO I – ARTIGO DA TESE

Journal of Food Composition and Analysis 127 (2024) 105940



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca



Determination of moisture and total protein and phosphorus contents in powdered chicken egg samples using digital images, NIR spectra, data fusion, and multivariate calibration

Luciano B. Ramo^a, Rossana O. Nobrega^a, David D.S. Fernandes^a, Wellington S. Lyra^a, Paulo H. G.D. Diniz^b, Mario C.U. Araujo^{a,*},¹

^a Universidade Federal da Paraíba, Departamento de Química, Laboratório de Automação e Instrumentação em Química Analítica/Quimiometria (LAQA), Caixa Postal 5093, CEP 58051 –970 João Pessoa, PB, Brazil

^b Universidade Federal do Oeste da Bahia, Campus Reitor Edgard Santos, CEP 47.810–059 Barreiras, BA, Brazil