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

ALINE COELHO SILVA FONSECA

# IDENTIFICAÇÃO DE VESTÍGIOS DE SANGUE HUMANO DEPOSITADOS EM DIFERENTES SUBSTRATOS E ENVELHECIDOS EM TEMPOS DISTINTOS UTILIZANDO ESPECTROSCOPIA E IMAGENS NO INFRAVERMELHO E FERRAMENTAS QUIMIOMÉTRICAS

Recife
2024

ALINE COELHO SILVA FONSECA

# IDENTIFICAÇÃO DE VESTÍGIOS SANGUE HUMANO DEPOSITADOS EM DIFERENTES SUBSTRATOS E ENVELHECIDOS EM TEMPOS DISTINTOS UTILIZANDO ESPECTROSCOPIA E IMAGENS NO INFRAVERMELHO E FERRAMENTAS QUIMIOMÉTRICAS

Tese de Doutorado apresentada ao Programa de Pós-graduação em Química da Universidade Federal de Pernambuco como requisito parcial à obtenção do Título de Doutor em Química.

**Área de concentração**: Química Analítica

Orientadora: Profª Drª Maria Fernanda Pimentel Avelar

Coorientador: Dr. José Francielson Queiroz Pereira

Recife
2024

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Fonseca, Aline Coelho Silva.
Identificação de vestígios de sangue humano depositados em diferentes substratos e envelhecidos em tempos distintos utilizando espectroscopia e imagens no infravermelho e ferramentas quimiométricas / Aline Coelho Silva Fonseca. - Recife, 2024.
133f.: il.

Tese (Doutorado) - Universidade Federal de Pernambuco, Centro de Ciências Exatas e da Natureza, Programa de Pós Graduação em Química, 2024.
Orientação: Maria Fernanda Pimentel Avelar.
Coorientação: José Francielson Queiroz Pereira.
Inclui referências e apêndices.

1. Quimiometria; 2. Infravermelho; 3. Química forense; 4. Manchas de sangue. I. Avelar, Maria Fernanda Pimentel. II. Pereira, José Francielson Queiroz. III. Título.

UFPE-Biblioteca Central

**ALINE COELHO SILVA FONSECA**

# "IDENTIFICAÇÃO DE VESTÍGIOS DE SANGUE HUMANO DEPOSITADOS EM DIFERENTES SUBSTRATOS E ENVELHECIDOS EM TEMPOS DISTINTOS UTILIZANDO ESPECTROSCOPIA E IMAGEM NO INFRAVERMELHO E TÉCNICAS QUIMIOMÉTRICAS"

Tese apresentada ao Programa de Pós-Graduação no Departamento de Química Fundamental da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Química.

Aprovada em: 16/12/2024

## BANCA EXAMINADORA

Documento assinado digitalmente
gov.br
MARIA FERNANDA PIMENTEL AVELAR
Data: 17/12/2024 12:40:10-0300
Verifique em https://validar.iti.gov.br

_____

**Profª. Maria Fernanda Pimentel Avelar (Orientadora)**
Universidade Federal de Pernambuco

Documento assinado digitalmente
gov.br
CLAUDETE FERNANDES PEREIRA
Data: 17/12/2024 12:21:10-0300
Verifique em https://validar.iti.gov.br

_____

**Profª. Claudete Fernandes Pereira**

Universidade Federal de Pernambuco

Documento assinado digitalmente
gov.br
MARCELO MARTINS DE SENA
Data: 16/12/2024 22:40:57-0300
Verifique em https://validar.iti.gov.br

_____

**Prof. Marcelo Martins de Sena**

Universidade Federal de Minas Gerais

Documento assinado digitalmente
gov.br
JOSE LICARION PINTO SEGUNDO NETO
Data: 17/12/2024 10:14:16-0300
Verifique em https://validar.iti.gov.br

_____

**Prof. José Licarion Pinto Segundo Neto**

Universidade Estadual do Rio de Janeiro

Documento assinado digitalmente
gov.br
PATRICIA VALDERRAMA
Data: 16/12/2024 14:12:55-0300
Verifique em https://validar.iti.gov.br

_____

**Profª. Patrícia Valderrama**

Universidade Tecnológica Federal do Paraná

Dedico a todas as mulheres da minha família que vieram antes de mim e abriram os caminhos para eu conseguir trilhar esses passos;

À minha avó Nicéa (*in memorian*) que me ensinou a ser forte;

À minha avó Cândida (*in memorian*) que me ensinou a ter fé;

À minha tia Adarc (*in memorian*) que me ensinou a ser resiliente;

À minha madrinha, Fátima, que me ensinou a ter dedicação;

À minha mãe, Arlete, que me ensinou a acreditar em mim mesma.

## AGRADECIMENTO

É uma verdade universalmente reconhecida que fazer uma pesquisa de doutorado não é uma tarefa fácil. Pense em tudo o que já ouviu falar sobre fazer um doutorado... Pensou? É tudo verdade: o medo da desaprovação, de não ser bom o suficiente, a tensão dos experimentos não darem certo, a incerteza com relação ao futuro e outras coisas mais... Agora, pense em tudo isso e adicione uma pandemia na história, o que já era desafiador, fica ainda mais assustador, certo? É por isso que, apesar de parecer, fazer uma pesquisa de doutorado, não é uma tarefa solitária. Ainda que ela seja defendida por uma única pessoa (sob a orientação e coorientação de duas pessoas incríveis), ela é, na verdade, fruto de uma grande comunidade invisível, que segurou a mão da doutoranda que vos escreve e não deixou ela desistir e fraquejar nos momentos de medo e desespero. A minha rede de apoio foi a melhor que eu poderia ter e, foi tão grande, que é possível que eu esqueça alguém no meio do caminho dos agradecimentos, mas eu não posso deixar de agradecer os que sempre estiveram mais próximos.

Minha gratidão a Deus que, nos momentos que eu duvidei de mim mesma, deu sinais e enviou pessoas incríveis para me dá força e coragem de continuar.

Minha gratidão à minha orientadora, professora Dra. Maria Fernanda Pimentel, que acreditou em mim quando eu apareci na sala dela - no final da graduação em Engenharia Química - com a ideia de fazer um mestrado na área de Química Forense no departamento de Química Fundamental, porque eu pensava em um dia – quem sabe - fazer o concurso para área de perícia criminal. Agradeço por ter segurado a minha mão no mestrado e não ter soltado em nenhum momento, mesmo quando eu precisei ficar distante, por ter confiado e acreditado em mim quando eu disse que ia estudar para um concurso durante o doutorado e ter me dado todo o apoio e suporte para que eu tivesse êxito nesse objetivo, por toda orientação, ideias, sugestões, correções, conselhos e incentivos. Professora, a senhora é um exemplo não só de profissional, mas de pessoa, e uma grande inspiração.

Minha gratidão ao meu coorientador, o Dr. José Francielson Queiroz Pereira que, além de todas as ideias, orientações, treinamentos, correções, sugestões, também me deu muitos conselhos e virou um grande amigo.

Minha gratidão aos meus pais, Américo e Arlete, que sempre acreditaram em mim, e me apoiaram em todas as decisões. Que nos dias que eu duvidei de mim mesma, levantaram a minha cabeça, enxugaram as minhas lágrimas e me ajudaram a enxergar a minha força.

Minha gratidão ao meu companheiro de vida, João Roberto, que esteve ao meu lado em todos os momentos: os de alegria por ter tido um resultado positivo, e o de medo por não ter

atingido a expectativa. E que sempre segurou minha mão e, nos momentos que eu pensei em desistir - e não foram poucos, só eu e você, sabemos quantos - me abraçou e me deu o apoio e suporte necessário para continuar e seguir em frente.

Minha gratidão ao meu irmão Arthur, por ser um porto seguro de confiança e paz, e a minha cunhada Gabriela, por toda torcida e confiança.

Minha gratidão à minha psicóloga, Luzia Compasso, que foi essencial durante toda a trajetória. Ter o seu suporte profissional foi essencial para que eu conseguisse atravessar essa jornada com mais leveza e enxergar todos os meus questionamentos com mais clareza.

Minha gratidão aos meus amigos Bruno e Erklaylle que trilharam comigo essa jornada desde o mestrado, e estiveram ao meu lado durante todos os trabalhos em grupo, os estudos pré-provas, e sempre trouxeram boas risadas nos momentos de cansaço e desespero. Ter vocês do meu lado durante os últimos seis anos tornaram tudo mais leve e divertido, que sorte eu tenho de fazer parte dessa geração de *"Fernandetes*"!

Minha gratidão a minha madrinha Fátima e ao meu padrinho Henrique que sempre acreditaram e torceram por mim durante todas as etapas da minha vida, seja acadêmica, profissional ou pessoal.

Minha gratidão aos meus sogros, Silvana e João, ao meu cunhado, Rodrigo, e a minha cunhada Laura, que se tornaram minha segunda família e me incentivaram e torceram por mim durante toda essa jornada.

Minha gratidão a todos os membros do grupo de pesquisa e do Laboratório de Combustíveis que possibilitaram o desenvolvimento desse trabalho.

Minha gratidão às minhas amigas Lívia, Marianna e Jacqueline que entenderam minha ausência, mas também acompanharam os momentos de tensão e alegria, e me deram muito apoio, também escutaram muitos desabafos e me distraíam nos momentos de descontração.

Minha gratidão aos amigos que fiz durante o curso de formação que torceram por mim nessa reta final do doutorado, quando eu duvidei que fosse conseguir conciliar e tive muito medo de não ser capaz de vencer os desafios, não só me incentivaram, como me motivaram e me deram muita força para eu acreditar no meu potencial. Agradeço, em especial, Gabi, Marina, Lorena, que seguraram a minha mão em muitos momentos e me deram muita força.

Minha gratidão a todos os membros do grupo de pesquisa, do Laboratório de Combustíveis (LAC), e do Instituto de Pesquisa em Petróleo e Energia (LITPEG) que possibilitaram o desenvolvimento desse trabalho.

Minha gratidão a todos os voluntários que aceitaram participar desse estudo e que, literalmente, deram o sangue pela ciência! Agradeço a confiança na pesquisa científica

brasileira e pernambucana. Em tempos obscuros com tanta desinformação, ter pessoas que apoiam e acreditam no pensamento científico trazem luz para a escuridão.

Minha gratidão às agências de fomento: à CAPES, pela bolsa de doutorado concedida a mim. Ao NUQAAPE, NEQUIFOR, INCTAA, FACEPE e ao CNPq pelo incentivo ao projeto, à UFPE pelo suporte institucional.

Minha gratidão a cada integrante da Pós-Graduação e do Departamento de Química Fundamental - UFPE, aos professores, aos alunos, e aos servidores, por toda dedicação empregada ao programa e ao departamento. Em especial, à Patrícia (Pati), que sempre foi muito disponível e solícita ao tirar minhas dúvidas, e resolver qualquer problema e questões administrativas. Pati, sua dedicação é admirável, você transmite paz e segurança para todos os alunos!

Minha gratidão a todos que contribuíram de alguma forma para esse trabalho.

## RESUMO

Manchas de sangue são vestígios comuns em investigações forenses e evidências essenciais, pois podem revelar informações críticas para auxiliar na resolução de crimes. Frequentemente encontradas em superfícies como pisos e tecidos, essas manchas podem ter sido depositadas em momentos variados. Consequentemente, desenvolver métodos analíticos que possam identificar manchas de sangue humano e estimar seu tempo de deposição é crucial para a ciência forense. O presente trabalho propõe o uso combinado de tecnologias baseadas espectroscopia no infravermelho próximo (NIR) e ferramentas quimiométricas, para identificar e estimar a idade de manchas de sangue humano depositadas em diferentes intervalos de tempo. Para o problema voltado em distinguir manchas de sangue humano e animal depositadas em diferentes pisos e tempos de deposição, foi utilizado um espectrômetro NIR portátil (região espectral entre 908 e 1676 nm), e modelos hierárquicos de classificação foram construídos combinando métodos quimiométricos, como Modelagem Independente e Flexível de Analogia de Classe (SIMCA) e Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA). Esses modelos foram aplicados a amostras de sangue humano e animal, envelhecidas em intervalos variados (0, 1, 9, 20, 30 e 162 dias para sangue humano; e 0, 1, 10, 13, 20, 29, 105 e 176 dias para sangue animal), depositadas em nove diferentes pisos. O modelo com melhor desempenho alcançou sensibilidade superior a 0,92 e especificidade acima de 0,7 para amostras envelhecidas entre 0 e 30 dias. Para o segundo problema, manchas de sangue também foram depositadas nos nove tipos de pisos e submetidas a envelhecimento por até 15 dias, com coleta de espectros NIR diretamente nas superfícies usando um outro equipamento portátil que abrange outra região espectral (1350 a 2500 nm). Foram testadas diferentes ferramentas de pré-processamento, e modelos hierárquicos foram desenvolvidos, combinando PLS-DA e Regressão de Mínimos Quadrados Parciais (PLSR). O melhor modelo estimou a idade das manchas de sangue depositadas em um intervalo de 2 a 10 horas com um erro quadrático médio de predição (RMSEP) próximo de 1 hora, e para manchas depositadas em intervalos de 1 a 15 dias, com um RMSEP de aproximadamente 2 dias. Além disso, foram comparadas duas abordagens para construção dos modelos, uma global, desenvolvida para ser utilizada em qualquer local de crime, e uma caso-específica, desenvolvida baseada nas condições em que a mancha foi encontrada. Para o terceiro problema, foram estudadas manchas de sangue depositadas em oito tecidos diferentes (quatro tecidos sintéticos e quatro tecidos de algodão, todos com cores distintas). Foram coletadas imagens hiperespectrais durante o intervalo de 0 a 15 dias e modelos de regressão foram construídos utilizando PLSR. Também foram comparadas duas abordagens, a global e a caso-específica. Os modelos específicos construídos para as manchas de sangue depositadas em um único tipo de tecido obtiveram RMSEP de 44 h (menos de dois dias) para as manchas de sangue depositadas em algodão, e RMSEP de 39 h para as manchas de sangue depositadas em tecido sintético, enquanto o modelo que englobava as manchas depositadas em ambos os tipos de fibra de tecido obteve um valor de RMSEP de 51 h.



Palavras-Chave: Sangue Humano. Modelagem Hierárquica. Espectrocopia no Infravermelho Próximo. Classificação. Datação.

**ABSTRACT**

Bloodstains are common forensic evidence that can reveal critical information for solving crimes. Often found on surfaces such as floors and fabrics, these stains may be deposited at different times, making it crucial to develop analytical methods capable of identifying human bloodstains and estimating their deposition time. This study proposes the combined use of near-infrared (NIR) spectroscopy and chemometric techniques to identify and estimate the age of human bloodstains deposited at various intervals. To address the challenge of distinguishing between human and animal bloodstains deposited on different floor types and at different times, a portable NIR spectrometer (operating between 908 and 1676 nm) was used. Hierarchical classification models were constructed by integrating chemometric methods, including Soft Independent Modeling of Class Analogy (SIMCA) and Partial Least Squares Discriminant Analysis (PLS-DA). These models were applied to human blood samples aged at 0, 1, 9, 20, 30, and 162 days, and to animal blood samples aged at 0, 1, 10, 13, 20, 29, 105, and 176 days, all deposited on nine different types of flooring. The best model achieved a sensitivity greater than 0.92 and specificity greater than 0.7 for samples aged between 0 and 30 days. For the second objective, bloodstains were deposited on the same nine floor types and aged for up to 15 days. NIR spectra were collected directly from the surfaces using a different portable spectrometer covering a spectral range of 1350 to 2500 nm. Various preprocessing techniques were tested, and hierarchical models combining PLS-DA and Partial Least Squares Regression (PLSR) were developed. The optimal model estimated the age of bloodstains in the 2- to 10-hour interval with a root mean square error of prediction (RMSEP) close to 1 hour, and for stains aged between 1 and 15 days, an RMSEP of approximately 2 days was achieved. Two modeling approaches were compared: a global model, designed for use at any crime scene, and a case-specific model tailored to the specific conditions where the stain was found. In a third study, bloodstains were deposited on eight different fabrics (four synthetic and four cotton, each with different colors) and monitored over a 0- to 15-day period. Hyperspectral images were collected, and PLSR regression models were constructed. The specific models developed for single fabric types achieved an RMSEP of 44 hours (less than two days) for bloodstains on cotton fabrics and 39 hours for synthetic fabrics, while the model encompassing stains on both fabric types had an RMSEP of 51 hours. These findings demonstrate the potential of NIR spectroscopy combined with chemometric techniques in forensic science, offering promising strategies for estimating bloodstain deposition times on various surfaces and materials.

Keywords: Human Blood. Hierarchical Modelling. Near Infrared Spectroscopy. Classification. Dating.

## LISTA DE ILUSTRAÇÕES


**Figura 1 -** Mudanças cinéticas da hemoglobina fora do corpo humano. 16

**Figura 2 -** Esquema do processo de evaporação da secagem de uma gota de sangue na superfície de um substrato: a) Fase pré-gelificação; b) Fase de gelificação. 18

**Figura 3 -** Diagrama esquemático de uma imagem hiperespectral (hipercubo) para uma hipotética mancha de sangue depositada em tecido mostrando a relação entre as dimensões espectrais e espaciais. Cada pixel na imagem hiperespectral é representado por um espectro individual contendo informações sobre a composição química deste pixel. 22

**Figura 4 -** Representação esquemática do princípio de funcionamento de a) uma ferramenta de análise discriminante e b) uma ferramenta de modelagem de classes, em um exemplo ilustrativo envolvendo duas classes de amostras (pontos azuis e pontos vermelhos). 27

**Figura 5 -** Esquema ilustrativo da distância ortogonal (OD) e da distância de *scores* (SD) de uma amostra em relação a uma classe hipotética. 28

**Figura 6 -** Média dos espectros: a) originais de sangue humano; b) pré-processados de sangue humano; c) originais de sangue animal; d) pré-processados de sangue animal. 43

**Figura 7 -** Modelo de PCA construído a partir dos espectros de sangue humano pré-processados com 1ª Derivada de Savitzky-Golay; a) Scores de PC1xPC2; b) Scores de PC1xPC3; c) *Loadings* de PC1 e PC2; d) *Loading*s de PC1 e PC3. 44

**Figura 8 -** Modelo de PCA construído com os espectros pré-processados de sangue humano e sangue animal. a) *Scores* de PC1xPC2 (coloridas de acordo com as classes de doadores); b) *Scores* de PC1xPC3 (coloridas de acordo com as classes de doadores); c) *Scores* de PC1xPC2 (coloridas de acordo com o tempo de deposição) d) *Scores* de PC1xPC3 (coloridas de acordo com o tempo de deposição); e) Gráfico de *Loadings* de PC1 e PC2; f) Gráfico de *Loadings* de PC1 e PC3. 45

**Figura 9 -** Esquema representativo dos modelos hierárquicos construídos. 46

**Figura 10 -** Gráficos de influência; a) Soma quadrada dos resíduos versus gráfico T² de Hotelling do primeiro modelo de PCA construído com amostras de sangue com 6 dias (Modelo Original); b) Gráfico de resíduos do segundo modelo de PCA construído com amostras de sangue humano com 6 dias (Modelo Original). 47

**Figura 11 -** Resultados de Sensibilidade e Especificidade do Modelo Hierárquico 1 (MH1) construído com a) amostras de 6 dias; b) amostras de 6 + 0 dias; c) amostras de 6 + 1 dias e d) amostras de 6 + 0 +1 dias. 48

**Figura 12 -** Resultados de Sensibilidade e Especificidade do Modelo Hierárquico 2 (MH2) construído com a) amostras de 6 dias; b) amostras de 6 + 0 dias; c) amostras de 6 + 1 dias ed) amostras de 6 + 0 +1 dias. 50

**Figura 13 -** Resultados de Sensibilidade e Especificidade do Modelo Hierárquico 3 (MH2) construído com a) amostras de 6 dias; b) amostras de 6 + 0 dias; c) amostras de 6 + 1 dias ed) amostras de 6 + 0 +1 dias. 51

**Figura 14 -** Manchas de sangue depositadas em pisos de cerâmica. 58

**Figura 15 -** Manchas de sangue depositadas em pisos de porcelanato. 58

**Figura 16 -** a) Imagem da aquisição espectral; b) Acessório de Teflon. 59

**Figura 17 -** a) Média dos espectros originais; b) Médio dos espectros pré-processadosa (1ª Derivada, 5 pontos e polinômio do 2º grau). 63

**Figura 18 -** Modelo de Análise de Componentes Principais construído com todas as amostras de sangue e de pisos (1ª Derivada de Savitzky-Golay - janela de 5 pontos e polinômio do 2° grau - e centragem na média): a) *Scores* de PC1xPC2 coloridos de acordo com o tipo de amostra; b) *Scores* de PC1xPC2 colorido de acordo com o tipo de piso; c) *Loadings* de PC1e PC2. 64

**Figura 19 -** Gráfico da Absorbância. a) Média dos espectros de sangue originais; b) Média dos espectros de sangue pré-processados (Suavização de Savitzky-Golay com janela de 5 pontos e polinômio de 2°grau e SNV). 65

**Figura 20 -** Modelo de PCA construído com todas as amostras de sangue: a) *Scores* de PC1xPC2; b) *Loadings* de PC1 e PC2. 65

**Figura 21 -** Modelo de PCA a) *Scores* de PC1XPC2 e b) *Loadings* de PC1xPC2 das amostras de sangue com tempo de deposição entre 0.5 h e 10 h; c) *Scores* de PC1xPC2 e d) *Loadings* de PC1xPC2 das amostras de sangue com tempo de deposição entre 24 h e 360 h; 67

**Figura 22 -** Esquema do Modelo Hierárquico. 68

**Figura 23 -** Modelo de regressão PLS construído com CT4 (amostras com tempo de deposição entre 2h e 10h); a) Referência versus previsto; b) Gráfico de *VIP Scores* pré-processado com Suavização, SNV e MC; c) Referência versus previsto; d) Gráfico *VIP Scores pré*-processado com Suavização, SNV, GLSW ($\alpha$ = 0,2) e MC; e) Referência versus previsto e f) Gráfico de pontuações VIP pré-processado com Suavização, SNV, GLSW ($\alpha$ = 0,03) e MC. 73

**Figura 24 -** Modelo de regressão PLS construído com CT4 (amostras com tempo de depósito entre 24h e 360h); a) Referência versus previsto e b) Gráfico de *VIP Scores* pré-processado com Suavização, SNV e CM; c) Referência versus previsto e d) Gráfico *VIP Scores* pré-processado com Suavização, SNV, GLSW ($\alpha$ = 0,2) e MC; e) Referência versus previsto e d) Gráfico de pontuações VIP pré-processado com Suavização, SNV, GLSW ($\alpha$ = 0,03). 75

**Figura 25 -** Gráficos de Probabilidade de Predição de Classe do modelo PLS-DA com projeção das amostras PS4 (cinza) com a) tempo de deposição entre 2h e 10h (Recente); b) horário de depósito entre 24h e 360h (Antigo). 76

**Figura 26 -** Modelo de PCA model construído com os espectros das triplicatas das manchas de sangue depositadas no piso PO5, pré-processados com Suavização, SNV e MC; a) coloridas de acordo com os doadores; b) coloridas de acordo com o tempo de deposição; c) *Loadings* de PC1 e PC2. 77

**Figura 27 -** Modelo de PCA model construído com os espectros das triplicatas das manchas de sangue depositadas no piso PO5, pré-processados 78

com Suavização, SNV, GLSW e MC; a) coloridas de acordo com os doadores; b) coloridas de acordo com o tempo de deposição; c) *Loadings* de PC1 e PC2.

**Figura 28 -** Modelo de regressão PLS construído com amostras depositadas em piso PO5; a) Gráfico de referência versus previsto e b) Gráfico de VIP *Scores* pré-processado com Suavização, SNV, GLSW (α = 0,03) e MC (amostras com tempo de deposição entre 2h e 10h); c) Referência versus previsto ed) VIP *Score*s pré-processado com Suavização, SNV, GLSW (α = 0,03) e MC (amostras com tempo de deposição entre 24 h e 360 h). 79

**Figura 29 -** Manchas de sangue depositadas em tecidos de algodão e sintético. 85

**Figura 30 -** a) Imagem falsa da mancha de sangue no tecido após a compressão; b) Histograma de frequência; c) Imagem falsa evidenciando o recorte dos pixels selecionados pela máscara. 87

**Figura 31 -** a) Espectros médios originais das manchas de sangue depositadas nos tecidos de algodão e sintético; b) Espectros médios originais das manchas de sangue depositadas nos diferentes tipos de tecido. 89

**Figura 32 -** a) Espectros originais obtidos a partir da média dos pixels das manchas de sangue coloridos de acordo com a escala de tempo; b) Média dos espectros obtidos a partir da média dos pixels das manchas de sangue coloridos de acordo com a escala de tempo; 90

**Figura 33 -** Modelo de PCA das amostras depositadas nos tecidos de algodão e sintéticos pré-processadas usando Suavização de Savitzky-Golay com 7 pontos e polinômio do 2°grau, SNV e centragem na média; a) *Scores* de PC1xPC2; b) *Loadings* de PC1 e PC2. 90

**Figura 34 -** Modelo de PCA das amostras depositadas nos tecidos de algodão e sintéticos pré-processadas usando Suavização de Savitzky-Golay com 7 pontos e polinômio do 2°grau, SNV, GLSW (α = 0.02) e centragem na média; a) *Scores* de PC1xPC2; b) *Loadings* de PC1 e PC2; 91

**Figura 35 -** Modelos de regressão PLS construídos utilizando a abordagem global. a) Y Predito x Y Medido e b) Gráfico de VIP *Scores* do modelo construído com as manchas de sangue depositadas em 93

tecido sintético e algodão; c) Y Predito x Y Medido e d) Gráfico de VIP *Scores* do modelo construído com as manchas de sangue depositadas em tecido de algodão; e) Y Predito x Y Medido e f) Gráfico de VIP *Scores* do modelo construído com as manchas de sangue depositadas no tecido sintético.

**Figura 36 -** Modelos de regressão PLS construídos utilizando a abordagem caso-específico nas amostras do tecido de algodão (CBE). a) Y Predito x Y Medido e b) Gráfico de VIP Scores do modelo 96

**Figura 37 -** Modelos de regressão PLS construídos utilizando a abordagem caso-específico nas amostras do tecido sintético (SBE). a) Y Predito x Y Medido e b) Gráfico de VIP Scores do modelo. 97

## LISTA DE TABELAS


| | | |
|---|---|---|
| **Tabela 1 -** | Quantidade de espectros de sangue humano e sangue animal coletados em cada dia. | 39 |
| **Tabela 2 -** | Resumo dos modelos hierárquicos construídos. | 41 |
| **Tabela 3 -** | Distribuição dos espectros utilizados na validação externa (predição). | 42 |
| **Tabela 4 -** | Resumo dos diferentes pisos usados nos diferentes conjuntos de treinamento e predição. | 69 |
| **Tabela 5 -** | Resumo das figuras de mérito dos modelos PLS construídos com amostras com tempo de depósito entre 2h e 10h e os diferentes pré-processamentos. | 71 |
| **Tabela 6 -** | Resumo das figuras de mérito dos modelos PLS construídos com amostras com tempo de depósito entre 24h e 360h e os diferentes pré-processamentos. | 74 |
| **Tabela 7 -** | Valores obtidos das figuras de mérito para os modelos. | 92 |
| **Tabela 8 -** | Valores obtidos das figuras de mérito para os modelos utilizando a 2ª abordagem nas amostras do tecido de algodão CBE. | 95 |
| **Tabela 9 -** | Valores obtidos das figuras de mérito para os modelos utilizando a 2ª abordagem nas amostras do tecido sintético SBE. | 96 |

## LISTA DE ABREVIATURAS E SIGLAS

| | |
|---|---|
| ATR | Reflectância Total Atenuada |
| CAAE | Certificado de Apresentação de Apreciação Ética |
| CAPES | Coordenação de Aperfeiçoamento de Pessoal de Nível Superior |
| CEP | Comitê de Ética em Pesquisa |
| CLS | Método Clássico de Mínimos Quadrados |
| MC | Centragem na Média |
| CP | Conjunto de Predição |
| CT | Conjunto de Treinamento |
| DA | Análise Discriminante |
| DQF | Departamento de Química Fundamental |
| FN | Falso Negativo |
| FP | Falso Positivo |
| FTIR | Infravermelho Próximo com Transformada de Fourier |
| GA-LDA | Análise Discriminante Linear baseada em Algoritmo Genético |
| GLSW | Mínimos Quadrados Parciais Ponderados Generalizados |
| HSI | Imagens Hiperespectrais |
| HSI-NIR | Imagens Hiperespectrais no Infravermelho Próximo |
| INCTAA | Instituto Nacional de Ciência e Tecnologia de Tecnologias Analíticas Avançadas |
| IR | Infravermelho |
| LAC | Laboratório de Combustíveis |
| LITPEG | Instituto de Pesquisa em Petróleo e Energia |
| LV | Variável Latente |
| MH | Modelo Hierárquico |
| MLR | Regressão Linear Múltipla |
| NS | Não-Sangue |
| NIR | Infravermelho Próximo |
| NUQAAPE | Núcleo de Química Analítica Avançada de Pernambuco |
| OD | Distância Ortogonal |
| PF | Polícia Federal |
| PC | Componente Principal |
| PCA | Análise de Componente Principal |

| | |
|---|---|
| PCR | Regressão por Componentes Principais |
| PLS | Mínimos Quadrados Parciais |
| PLS-DA | Análise Discriminante por Mínimos Quadrados Parciais |
| PLSR | Regressão por Mínimos Quadrados Parciais |
| RMSE | Raiz do Erro Quadrático Médio |
| RMSEC | Raiz do Erro Quadrático Médio de Calibração |
| RMSECV | Raiz do Erro Quadrático Médio de Validação Cruzada |
| RMSEP | Raiz do Erro Quadrático Médio de Predição |
| RoI | Região de Interesse |
| SA | Sangue Animal |
| SD | Distância de *Scores* |
| SPA-LDA | Análise Discriminante Linear baseado em Algoritmo de Projeções Sucessivas |
| SH | Sangue Humano |
| SIMCA | Modelagem Independente e Flexível por Analogia de Classe |
| Sn | Sensibilidade |
| SNV | Padronização Normal de Sinal |
| Sp | Especificidade |
| SVM | Máquina de Vetores de Suporte |
| TCLE | Termo de Livre Consentimento Esclarecido |
| UFPE | Universidade Federal de Pernambuco |
| UFRPE | Universidade Federal Rural de Pernambuco |
| UV-Vis | Ultravioleta Visível |
| VIP | Importância das Variáveis na Projeção |
| VN | Verdadeiro Negativo |
| VP | Verdadeiro Positivo |

**SUMÁRIO**


| | | |
|---|---|---|
| **1** | **INTRODUÇÃO GERAL** | **12** |
| 1.1 | OBJETIVOS GERAIS | 13 |
| **2** | **FUNDAMENTAÇÃO TEÓRICA** | **15** |
| 2.1 | VESTÍGIOS DE SANGUE | 15 |
| 2.1.1 | Composição e propriedades do sangue | 15 |
| 2.1.2 | Dinâmica de secagem do sangue | 16 |
| 2.1.3 | Manchas de sangue em pisos | 17 |
| 2.1.4 | Manchas de sangue em tecidos | 18 |
| **2.2** | **ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO** | **19** |
| 2.2.1 | Equipamentos Portáteis no Infravermelho Próximo | 20 |
| 2.2.2 | Imagens Hiperespectrais no Infravermelho Próximo | 21 |
| **2.3** | **FERRAMENTAS QUIMIOMÉTRICAS** | **23** |
| 2.3.1 | Ferramentas de Pré-Processamentos | 24 |
| 2.3.2 | Análise de Componente Principais (PCA) | 25 |
| 2.3.3 | Métodos de Classificação | 26 |
| 2.3.4 | Métodos de Calibração | 31 |
| 2.3.5 | Modelos Hierárquicos | 32 |
| 2.3.6 | Validação e Figuras de Mérito | 33 |
| **3** | **MODELOS HIERÁRQUICOS PARA IDENTIFICAÇÃO DE AMOSTRAS DE SANGUE COM TEMPOS DIFERENTES** | **36** |
| 3.1 | INTRODUÇÃO | 36 |
| **3.1.1** | **Objetivos** | **37** |
| 3.2 | MATERIAIS E MÉTODOS | 38 |
| **3.2.1** | **Preparação da amostra** | **38** |
| **3.2.2** | **Aquisição dos Espectros** | **38** |
| **3.2.3** | **Análise e Pré-Processamento de Dados** | **39** |
| **3.2.4** | **Conjuntos de Treinamento e Predição** | **39** |
| **3.2.5** | **Avaliação dos Modelos Hierárquicos para Classificação** | **40** |
| 3.3 | RESULTADOS E DISCUSSÃO | 42 |
| **3.3.1** | **Análise Espectral** | **42** |
| **3.3.2** | **Avaliação dos Modelos Hierárquicos** | **46** |
| 3.3.2.1 | Modelo Hierárquico 1 | 46 |

3.3.2.2 Modelo Hierárquico 2 ..... 49
3.3.2.3 Modelo Hierárquico 3 ..... 51
3.4 CONCLUSÃO ..... 53
**4 MODELOS QUIMIOMÉTRICOS PARA ESTIMATIVA DO TEMPO DE DEPOSIÇÃO DE AMOSTRAS DE SANGUE EM DIFERENTES PISOS ..... 55**
4.1 INTRODUÇÃO ..... 55
**4.1.1 Objetivos ..... 57**
4.2 MATERIAIS E MÉTODOS ..... 58
**4.2.1 Aquisição dos Espectros ..... 59**
**4.2.2 Análise de Dados ..... 60**
4.3 RESULTADOS E DISCUSSÃO ..... 62
**4.3.1 Abordagem Global ..... 62**
4.3.1.1 Análise Exploratória ..... 62
4.3.1.2 Classificação das Amostras e Construção do Modelo Hierárquico ..... 67
4.3.1.3 Conjuntos de Treinamento e Predição ..... 69
4.3.1.4 Modelos de Regressão ..... 70
4.3.1.5 Resultados do Modelo Hierárquico ..... 76
**4.3.2 Abordagem Caso-Específico ..... 77**
4.3.2.1 Análise Exploratória ..... 77
4.3.2.2 Modelos de Regressão ..... 78
4.4 CONCLUSÃO ..... 80
**5 ESTIMATIVA DO TEMPO DE DEPOSIÇÃO DE MANCHAS DE SANGUE EM TECIDOS UTILIZANDO IMAGENS HIPERESPECTRAIS NO INFRAVERMELHO PRÓXIMO E FERRAMENTAS QUIMIOMÉTRICAS ..... 82**
5.1 INTRODUÇÃO ..... 82
**5.1.1 Objetivos ..... 84**
5.2 MATERIAIS E MÉTODOS ..... 85
5.3 RESULTADOS E DISCUSSÃO ..... 88
**5.3.1 Abordagem Global ..... 88**
5.3.1.1 Análise Espectral e Exploratória ..... 88

5.3.1.2 Modelos de Regressão..... 91
**5.3.2 Abordagem Caso-Específico..... 94**
5.4 CONCLUSÃO..... 97
6 PERSPECTIVAS FUTURAS..... 99
**REFERÊNCIAS..... 100**
**APÊNDICE A – ARTIGO PUBLICADO..... 114**
**APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO..... 115**
**APÊNDICE C – GRÁFICOS DE CONTORNO DAS SUPERFÍCIES DE RESPOSTAS PARA OS DADOS COLETADOS UTILIZANDO O NEOSPECTRA..... 117**
**APÊNDICE D – NOTA DE IMPRENSA..... 121**

# 1 INTRODUÇÃO GERAL

Os vestígios de sangue são uma das evidências mais importantes encontradas em cenas de crimes violentos, posto que elas podem ajudar a responder uma série de questionamentos. Idealmente, essas evidências devem ser identificadas e coletadas ainda no local do crime, de modo a agilizar a ação dos peritos para solucionar a ordem dos acontecimentos. A maioria dos métodos utilizados são destrutivos, o que dificulta que análises posteriores, como identificação de DNA sejam realizadas quando o vestígio coletado é limitado (BEVEL; GARDNER, 2008).

Nesse sentido, o uso de equipamentos portáteis, como os baseados em técnicas vibracionais como o infravermelho próximo, são uma alternativa para o desenvolvimento de métodos capazes de preservar a integridade dos vestígios (FONSECA et al., 2022; MORILLAS; GOOCH; FRASCIONE, 2018; PEREIRA et al., 2017) e permitir que análises de contraprova ou complementares sejam realizadas. Para extrair informações de interesse a partir dos espectros coletados é necessário associar o uso de ferramentas estatísticas multivariados, também conhecidos como ferramentas quimiométricas.

Contudo, muitas vezes, não é suficiente apenas identificar os vestígios de sangue, mas também determinar o tempo de deposição daquele material. Nos últimos anos, vários pesquisadores têm buscado desenvolver métodos capazes de não somente identificar o sangue, mas também extrair outras informações que possam colaborar com a elucidação de crimes, tal como estimar o tempo de sua deposição. Uma revisão escrita por Zadora e Menzyk, em 2018, destaca as dificuldades encontradas por pesquisadores para encontrar um método confiável para estimar o tempo decorrido desde a deposição de um vestígio de sangue. Fatores como umidade, temperatura, e a própria composição sanguínea são limitações a serem vencidas no desenvolvimento de métodos para realizar essa estimativa (ZADORA; MENZYK, 2018). Entre 2011 e 2024, diversos métodos foram propostos usando técnicas espectroscópicas para datar nãos só manchas de sangue como demais fluidos biológicos, utilizando Espectroscopia Raman, Espectroscopia no Infravermelho Próximo com Transformada de Fourier e Reflectância Total Atenuada (ATR-FTIR), Imagens Hiperespectrais (HSI), entre outros (DOTY; MCLAUGHLIN; LEDNEV, 2016; DOTY; MURO; LEDNEV, 2017; LI et al., 2013; LIN et al., 2017, 2018; MANIS et al., 2022; TASTEKIN et al., 2024).

Edelman *et al.* (2012) utilizaram imagens empregando câmeras hiperespectrais obtidas na faixa entre 1000 e 2500 nm para identificar e datar vestígios de sangue. Em

um estudo para amostras de sangue com idades diferentes, os autores constataram que as bandas das regiões acima de 1690 nm foram as que mais sofreram variações ao longo do tempo (EDELMAN; VAN LEEUWEN; AALDERS, 2012). Um trabalho realizado por Kuenstner e Norris, em 1994, mostrou que foi possível detectar bandas de absorção associadas à hemoglobina centralizadas em 1690, 1740, 2056, 2170, 2290 e 2350 nm (KUENSTNER; NORRIS, 1994). Malegori *et al.* (2020) usaram Imagens Hiperespectrais no Infravermelho Próximo (HSI-NIR, do inglês, *Hyperspectral Imaging Near Infrared*) na faixa entre 1000 e 2500 nm para identificar diferentes fluidos biológicos, incluindo o sangue humano. Eles atribuíram a absorção em 2280 nm ao segundo sobretom de estiramento da ligação -NH associada a proteínas presentes no sangue (MALEGORI et al., 2020).

Outro fato que dificulta o desenvolvimento de ferramentas para identificação e estimativa do tempo de depósito de vestígios de sangue, é o fato de que podem ser encontrados em diferentes tipos de superfície, sendo que pisos e tecidos são as superfícies mais comuns. O estudo realizado por Edelman *et al.* (2012) utilizou espectroscopia na região do NIR combinada à reflectância no visível para identificar e datar manchas de sangue em tecidos de algodão cru e colorido. Para isso, eles construíram modelos de Regressão por Mínimos Quadrados Parciais (PLS, do inglês, *Partial Least Squares*) e os resultados obtidos evidenciaram que vestígios de sangue podem ser distinguidos de demais substâncias com coloração avermelhada com 100% de sensibilidade e especificidade. Contudo, em relação à datação, eles obtiveram resultados satisfatórios apenas para manchas mais recentes (EDELMAN *et al.*, 2012).

Um estudo recente elaborado por Pereira e colaboradores (2021) utilizou HSI-NIR para identificar manchas de sangue. Modelos hierárquicos foram desenvolvidos para unir os modelos PCA e PLS-DA construídos com as imagens das amostras de sangue depositados em tecido de algodão obtendo resultados de 95% de sensibilidade e 98% de especificidade. Já para amostras de sangue depositadas em tecidos sintéticos, os valores de sensibilidade e especificidade foram 100% e 98%, respectivamente (PEREIRA et al., 2021). O trabalho demonstrou o excelente potencial da aplicação dos modelos hierárquicos para identificação e classificação de amostras de sangue. Com o intuito de expandir os estudos iniciados por Pereira *et al.* (2021) e Fonseca *et al.* (2022), o presente trabalho visou aumentar a variabilidade das amostras estudadas, incluindo amostras com

tempos de deposição distintos, além de buscar desenvolver técnicas para estimar o tempo de deposição de manchas de sangue nas diferentes superfícies.

## 1.1 OBJETIVOS GERAIS

Essa pesquisa teve como objetivo desenvolver técnicas analíticas para identificação de manchas de sangue humano depositadas em diferentes substratos (pisos e tecidos) com tempos de deposição distintos, a partir da construção de modelos quimiométricos, buscando simular vestígios de sangue que poderiam ser encontrados em diferentes superfícies em possíveis cenas de crime. Para adquirir os espectros foram utilizados diferentes espectrômetros portáteis no infravermelho próximo (NIR) e imagens adquiridas por uma câmera hiperespectral na região NIR.

# 2 FUNDAMENTAÇÃO TEÓRICA

## 2.1 VESTÍGIOS DE SANGUE

O sangue é um tecido conjuntivo cuja principal função é transportar oxigênio e nutrientes para todos os órgãos do corpo. Correspondendo a cerca de 8% do peso corporal humano, ele é composto de glóbulos vermelhos (eritrócitos), glóbulos brancos (leucócitos) e plaquetas (trombócitos), que estão suspensos em uma matriz chamada plasma (LI, 2011; SOBAC; BRUTIN, 2014). Do ponto de vista forense, encontrar evidências como manchas de sangue significa encontrar uma fonte de grande variedade de informações que podem ajudar a solucionar um crime, tais como: material genético, ordem de acontecimentos, presença de substâncias tóxicas, uso de medicamentos, entre outros (BEVEL; GARDNER, 2008; ELKINS, 2019; VIRKLER; LEDNEV, 2009; ZADORA; MENZYK, 2018b).

### 2.1.1 Composição e propriedades do sangue

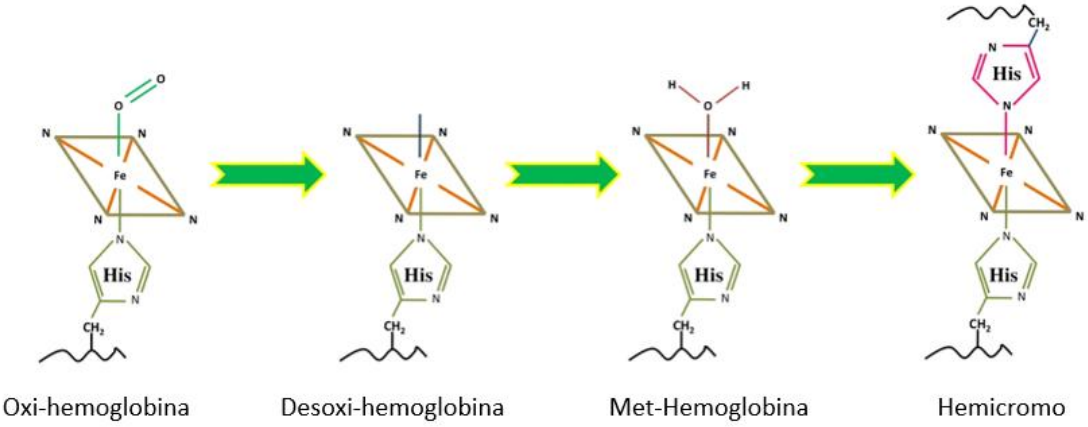
Em relação ao seu aspecto físico-químico, o sangue pode ser descrito como um fluido composto por uma fase líquida (plasma) na qual partículas sólidas (glóbulos vermelhos, glóbulos brancos e plaquetas) estão em suspensão. O plasma é composto por água, proteínas, ácidos nucléicos, lipídeos, fatores de coagulação e outros compostos metabólicos em suspensão. Ele representa cerca de 55% do volume sanguíneo (sendo que 95% do plasma é água, e os outros 5% são os demais componentes). O volume restante do sangue é constituído pelos elementos figurados - ou partículas sólidas – sendo que cerca de 97% desses elementos são compostos por hemácias (eritrócitos), 2% são glóbulos brancos e 1% são plaquetas (SOBAC; BRUTIN, 2014).

A hemoglobina é a proteína responsável pelo transporte de oxigênio intracelular, e ela se encontra dentro das hemácias. Quando está dentro do corpo humano, a hemoglobina pode ser encontrada na forma de duas variantes: a oxi-hemoglobina e a desoxi-hemoglobina, sendo a primeira a combinação da hemoglobina saturada de oxigênio, e a desoxi-hemoglobina o seu estado após a saída do oxigênio (SOBAC; BRUTIN, 2014).

Quando o sangue deixa o corpo humano, ele sofre diversas mudanças físico-químicas. Primeiramente, a hemoglobina ficará saturada de oxigênio atmosférico e se transformará em oxi-hemoglobina. Com o tempo, ela continuará a sofrer um processo

oxidativo e se transformará em metemoglobina (caracterizada pela presença do ferro no estado férrico, $Fe^{+3}$, incapaz de se ligar ao oxigênio) e, por fim, em hemicromo (Das et al., 2020). Os hemicromos são formados por meio de mudanças na conformação da proteína, de modo que os átomos endógenos à proteína se ligam ao ferro como sexto ligante (MARRONE; BALLANTYNE, 2009). As mudanças cinéticas que a molécula de hemoglobina pode sofrer podem ser observadas na Figura 1.

**Figura 1 -** Mudanças cinéticas da hemoglobina fora do corpo humano.




**Fonte:** Adaptado de (Das et al., 2020)

Além dos processos oxidativos descritos acima, durante a secagem do sangue outros fenômenos físicos ocorrem e são capazes de influenciar na dinâmica de secagem das manchas de sangue. É importante destacar que todas as modificações também influenciam nas variações espectrais que ocorrem durante a absorção de radiação na região do infravermelho próximo, conforme pode ser observado no Anexo A (KUENSTNER et al., 1994).

### 2.1.2 Dinâmica de secagem do sangue

Para entender a dinâmica de secagem do sangue, é importante considerar que o sangue é um fluido heterogêneo, que tem um comportamento similar a uma suspensão coloidal (SOBAC; BRUTIN, 2011) e que diversos fenômenos começam a ocorrer quando o sangue deixa o corpo humano: as plaquetas dão início ao processo de coagulação sanguínea e outros componentes começam um processo de degradação. Esses fenômenos ainda são pouco estudados e pouco considerados no estudo do envelhecimento do sangue (BENABDELHALIM; BRUTIN, 2022), isso porque as plaquetas representam apenas 1% das partículas sólidas presentes no sangue (LI, 2011).

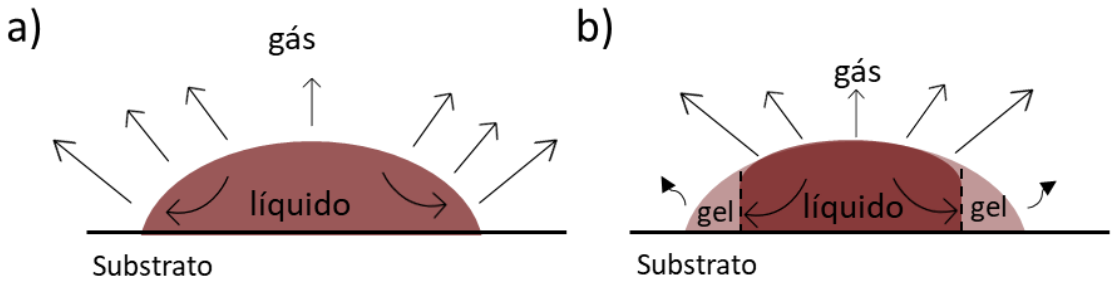
Um dos principais fenômenos que ocorrem é a evaporação da água presente no plasma. Alguns estudos demonstram que o sangue apresenta um comportamento similar ao de uma suspensão coloidal, e à medida que a água do plasma evapora, o sangue começa a assumir uma consistência de gel (BRUTIN et al., 2011; SMITH; NICLOUX; BRUTIN, 2020; SOBAC; BRUTIN, 2011, 2014). É importante ressaltar que a dinâmica de secagem do sangue também está sujeita à influência de diversos fatores, como as condições ambientais de temperatura, umidade, fluxo do ar, exposição solar, porosidade do substrato (BENABDELHALIM; BRUTIN, 2022; BRUTIN et al., 2011; SOBAC; BRUTIN, 2011, 2014); o volume de sangue depositado (SMITH; NICLOUX; BRUTIN, 2020); e a própria composição sanguínea, visto que distúrbios e doenças podem afetar na concentração de hemácias, leucócitos e plaquetas. Todos esses fatores, tornam a dinâmica do processo de secagem do sangue bastante complexa (ATTINGER et al., 2013).

Outro fator importante que influencia na dinâmica de secagem é a molhabilidade da superfície do substrato, assim como a sua porosidade. Uma mancha de sangue depositada em um material pouco poroso, como o piso, tem uma dinâmica de secagem diferente de uma mancha de sangue depositada sob um tecido, por exemplo. Nesse caso, o fluido irá se espalhar ao longo das fibras do tecido por capilaridade (DICKEN et al., 2019a; TAYLOR et al., 2016; WANG et al., 2021).

### 2.1.3 Manchas de Sangue em Pisos

Quando gotas de sangue caem em um substrato pouco poroso, elas não são absorvidas pela superfície do substrato. As hemácias se espalham uniformemente ao longo da mancha e o processo de coagulação começa a ocorrer. O processo de coagulação ocorre quando as plaquetas e o fibrinogênio começam a estabelecer um aspecto de teia e coincide com o início do processo de gelificação da mancha de sangue. Durante a gelificação, a evaporação da água inicia-se pelas extremidades mais externas da mancha onde há um maior fluxo de evaporação, pois as partículas sólidas começam a se concentrar nessa linha de contato (Figura 2a). Essa região, onde há uma maior concentração de partículas sólidas, começa a sofrer um aumento na sua viscosidade, o que leva a formação de um gel iniciando-se nas extremidades, enquanto o centro da mancha continua líquido (Figura 2b)(BENABDELHALIM; BRUTIN, 2022; SOBAC; BRUTIN, 2014).

**Figura 2 -** Esquema do processo de evaporação da secagem de uma gota de sangue na superfície de um substrato: a) Fase pré-gelificação; b) Fase de gelificação.




**Fonte:** Adaptado de (SOBAC; BRUTIN, 2014)

Após o início do processo de gelificação, a mancha se encontra com uma textura um pouco mais sólida. No entanto, como ainda há solvente retido em seus poros, o processo de evaporação continua, algo que poderá resultar na formação de fissuras, um processo também chamado de craquelamento. Esse fenômeno pode gerar padrões radiais de rachaduras que começam a ser formadas radialmente em torno do centro da mancha (BENABDELHALIM; BRUTIN, 2022; BRUTIN et al., 2011; SOBAC; BRUTIN, 2014).

Além da formação de fissuras devido ao craquelamento, também pode ocorrer um outro processo físico, chamado de delaminação. Esse fenômeno ocorre, principalmente, nas extremidades das manchas devido à maior concentração de tensões mecânicas nessa região, as bordas começam a levantar e descolar da superfície do substrato (BENABDELHALIM; BRUTIN, 2022; SOBAC; BRUTIN, 2014). Todos esses fenômenos revelam que as manchas formadas a partir de gota de sangue podem adquirir diferentes formatos, o que torna seu estudo bastante complexo. De modo que mesmo gotas de sangue fornecidas por um mesmo doador podem formar manchas com padrões diferentes entre si.

### 2.1.4 Manchas de Sangue em Tecidos

Quando uma gota de sangue semelhante à descrita no item anterior cai sobre a superfície de um substrato têxtil, os mesmos processos descritos para a secagem de manchas de sangue em pisos podem vir a ocorrer. No entanto, diferente da superfície dos

pisos, o processo de absorção do sangue no tecido adiciona uma outra complexidade na formação de manchas (WILLIAMS et al., 2016)

Estudos mostram que diversos fatores relacionados aos tipos de tecidos podem influenciar na absorção, secagem e no espalhamento das manchas de sangue, como o tipo de fibra (se natural, sintética, ou uma mistura de ambos), a rugosidade do tecido, a massa distribuída por área de tecido, se há aplicação de algum produto em sua superfície, se os tecidos foram lavados previamente, entre diversos outros fatores (DE CASTRO et al., 2013; DICKEN et al., 2019b; KARGER; RAND; BRINKMANN, 1998)

A maioria dos estudos relacionados à dinâmica de secagem de manchas de sangue em tecido estão interessados em analisar os padrões de manchas de sangue que podem ser formados para correlacionar com a forma como aconteceu um crime, de modo a inferir se as manchas de sangue foram causadas por contato direto, por gotículas, se foi em movimento, se caiu em uma trajetória linear, entre outros (DE CASTRO et al., 2013, 2015; KARGER; RAND; BRINKMANN, 1998; WANG et al., 2021; WILLIAMS et al., 2016).

Estudos recentes mostram a aplicação de outras técnicas analíticas para identificação de manchas de sangue em tecidos, como o uso de espectroscopia no infravermelho próximo através de câmeras hiperespectrais (MANIS et al., 2022; PEREIRA et al., 2021).

## 2.2 ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIR)

A espectroscopia de absorção no infravermelho próximo é uma técnica analítica não-destrutiva que se baseia na interação entre a radiação na região do infravermelho próximo (NIR) e a matéria. A radiação eletromagnética na região NIR compreende a faixa entre 780 nm e 2500 nm. A energia transportada pela radiação nessa região é relativamente baixa, de modo que sua interação com a matéria costuma promover apenas transições de estados vibracionais, e em alguns casos, como em gases, rotacionais (BURNS; CIURCZAK, 2009; PASQUINI, 2003; SKOOG; HOLLER; CROUCH, 2018).

Quando consideramos as frequências de radiação na região do infravermelho próximo, a energia associada a essa região excede a energia necessária para as transições vibracionais fundamentais, que ocorrem devido à interação e absorção da radiação na região do infravermelho médio. Dessa forma, as características espectrais na região NIR terão um comportamento diferente da região do MIR e estão normalmente associadas a

sobretons, combinações de bandas de absorção, ressonância e acoplamento de diferentes tipos de vibrações de mesmos grupos funcionais (BEĆ; GRABSKA; HUCK, 2021; BURNS; CIURCZAK, 2009; O'BRIEN et al., 2012; PASQUINI, 2003).

A complexidade da interpretação dos espectros NIR fez com que apenas após a década de 1980, com os avanços da tecnologia, em particular da quimiometria, a técnica recebesse a devida atenção e sua aplicação começou a ser mais difundida em diversas áreas do meio acadêmico (BURNS; CIURCZAK, 2009; PASQUINI, 2003). Atualmente, o estudo da ciência forense vem implementando o uso da espectroscopia no infravermelho próximo para análise de manchas de sangue (GARIGLIO et al., 2024; KUMAR; SHARMA; SHARMA, 2020; MENŻYK et al., 2023) entre outras aplicações... É esperado que a associação das abordagens espectroscópicas juntamente com as ferramentas quimiométricas melhorem os resultados das investigações criminais, visto que as evidências podem ser analisadas de maneira não destrutivas, de modo a permitir que análises posteriores sejam realizadas (CANO-TRUJILLO et al., 2023).

O avanço da tecnologia permitiu não só a evolução dos equipamentos de bancada que operam na região do infravermelho próximo, como também o desenvolvimento de equipamentos portáteis que podem ser levados mais facilmente para qualquer lugar (BEĆ; GRABSKA; HUCK, 2021; MORILLAS; GOOCH; FRASCIONE, 2018; ZHI; LUO; LIU, 2020), e de câmeras hiperespectrais que fornecem informações espaciais das amostras, além das informações espectrais (AMIGO; BABAMORADI; ELCOROARISTIZABAL, 2015; MAJDA et al., 2018; SUN, 2010; VIDAL; AMIGO, 2012).

### 2.2.1 Equipamentos Portáteis no Infravermelho Próximo

Os espectrômetros portáteis, ou miniaturizados, vêm sendo amplamente utilizados em diversas áreas da química analítica, como na análise de alimentos, monitoramento industrial, na área forense, entre outras (BEĆ; GRABSKA; HUCK, 2021). Há diversas vantagens em sua aplicação, pois, eles são leves (facilitando sua portabilidade), podem ser conectados online, via *bluetooth*, via Barramento Serial Universal (USB, do inglês *Universal Serial Bus*) ou via *Wifi*, e vêm aumentando seu desempenho e eficiência permitindo análises mais robustas e confiáveis (BEĆ; GRABSKA; HUCK, 2021; ZHI; LUO; LIU, 2020).

Geralmente, por serem muito mais compactos que os equipamentos de bancadas, os equipamentos portáteis possuem faixas espectrais mais reduzidas – porém, podem funcionar em toda região do NIR – e com menor resolução quando comparados aos instrumentos de bancada. O surgimento de espectrômetros NIR portáteis pode ser visto como um grande avanço, pois proporcionou o desenvolvimento de aplicações inteiramente novas e anteriormente inatingíveis, especialmente, quando se considera a possibilidade de realização de análises *in situ* (BEĆ; GRABSKA; HUCK, 2021; O'BRIEN et al., 2012; ZHI; LUO; LIU, 2020). Com o avanço das ferramentas quimiométricas, o uso de equipamentos portáteis se tornou cada vez mais interessante, pois permitem o desenvolvimento de modelos de classificação e calibração que tornam seu uso rápido, eficiente e automatizado (CANO-TRUJILLO et al., 2023; KUMAR; SHARMA; SHARMA, 2020; KUMAR; SHARMA, 2018; SILVA; BRAZ; PIMENTEL, 2019).

A facilidade de realização dessas análises é de grande interesse para área forense, pois permite que o equipamento seja levado até o local do crime, facilitando a coleta de evidências e a realização de testes preliminares. Além disso, por ser uma técnica não destrutiva, permite que análises confirmatórias sejam realizadas posteriormente em laboratório (CANO-TRUJILLO et al., 2023; KUMAR; SHARMA; SHARMA, 2020; KUMAR; SHARMA, 2018; SILVA; BRAZ; PIMENTEL, 2019). Muitos estudos vêm sendo realizados utilizando equipamentos portáteis para aplicações em diversas áreas da ciência forense, desde análise de adulteração de alimentos até discriminação de amostras de fluidos biológicos (FONSECA et al., 2022; FULGÊNCIO et al., 2022; MORILLAS; GOOCH; FRASCIONE, 2018; NETTO et al., 2023; O'BRIEN et al., 2012; PEREIRA et al., 2017).

### 2.2.2 Imagens Hiperespectrais no Infravermelho Próximo (HSI-NIR)

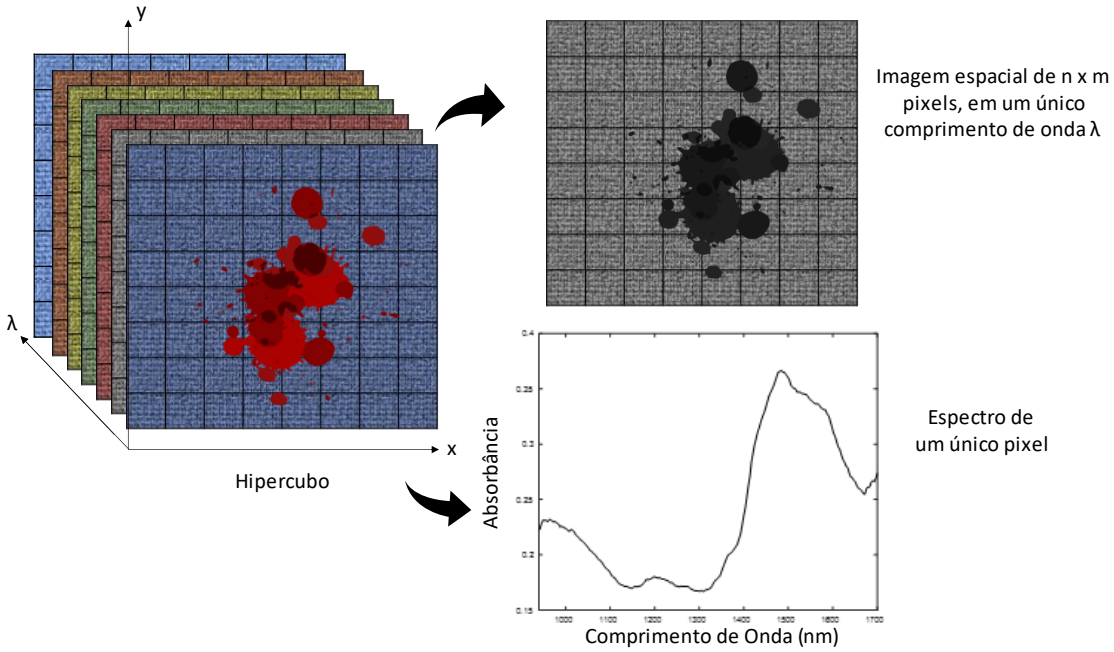
Uma das limitações do uso de espectrômetros portáteis em cenas de crime é que as medições obtidas com esses equipamentos são pontuais, de modo que, muitas vezes, é necessário adquirir diversos espectros da mesma amostra para obter um espectro representativo. Nesse sentido, as imagens hiperespectrais no infravermelho próximo (HSI-NIR, do inglês Hyperspectral Imaging in the Near Infrared) apresentam a vantagem de fornecer informações espaciais e espectrais dos objetos que estiverem em seu campo de amostragem, eliminando a necessidade de encontrar a posição exata das manchas de

sangue, por exemplo (AMIGO; BABAMORADI; ELCOROARISTIZABAL, 2015; VIDAL; AMIGO, 2012).

Os dados adquiridos através das imagens encontram-se na forma de hipercubo (também chamado de cubo hiperespectral), em que cada pixel contém duas dimensões de informações espaciais (composta por "n" linhas e "m" colunas) e uma informação que corresponde a um perfil espectral (CHANG, 2003; MAJDA et al., 2018; SUN, 2010). Devido à quantidade e complexidade do conjunto de dados, a quimiometria também é essencial para extrair informação relevante das imagens (AMIGO; BABAMORADI; ELCOROARISTIZABAL, 2015; VIDAL; AMIGO, 2012).

A Figura 3 ilustra um hipercubo de uma hipotética imagem hiperespectral de uma mancha de sangue depositada em um tecido (SUN, 2010).

**Figura 3** - Diagrama esquemático de uma imagem hiperespectral (hipercubo) para uma hipotética mancha de sangue depositada em tecido mostrando a relação entre as dimensões espectrais e espaciais. Cada pixel na imagem hiperespectral é representado por um espectro individual contendo informações sobre a composição química deste pixel.




**Fonte:** Adaptado de (SUN, 2010).

A imagem hiperespectral bruta consiste em uma série de imagens contíguas, sendo que cada imagem representa a intensidade e a distribuição espacial do objeto em uma determinada faixa de comprimentos de onda. Todas as imagens espaciais individuais podem ser captadas a partir do hipercubo em qualquer comprimento de onda que cubra a faixa espectral do sistema. Trabalhar com as imagens nesse formato de hipercubo é

bastante complexo, e o usual é desdobrar essas imagens em uma matriz de dados bidimensional, de modo que cada pixel será tratado como uma amostra e os comprimentos de onda são as variáveis (λ) (BASANTIA; NOLLET; KAMRUZZAMAN, 2019; CHANG, 2003; DE JUAN et al., 2009).

Além disso, muitas vezes, não é possível trabalhar com as imagens hiperespectrais inteiras, e é necessário selecionar uma região de interesse (ROI, do inglês, *Region of Interest*) que consiste na seleção de pixels que contém informações espaciais e espectrais que interessem. Essa seleção pode ser realizada computacionalmente, ou de forma manual, supervisionada por um usuário (BASANTIA; NOLLET; KAMRUZZAMAN, 2019). Em uma imagem de mancha de sangue em tecido, por exemplo, os pixels referentes à mancha de sangue são selecionados para identificar seus pixels vizinhos semelhantes. Assim, esses métodos utilizam informações espaciais e de intensidade como referência manual para segmentar uma região. Um estudo realizado por Edelman e colaboradores (2012) demonstra que imagens hiperespectrais de manchas de sangue com aproximadamente 1 mm de diâmetro e obtidas com até 1,5 m de distância puderam ser utilizadas para identificar sangue em possíveis cenas de crime (EDELMAN; VAN LEEUWEN; AALDERS, 2012).

Quando se considera o cenário forense, há uma grande diversidade de aplicações no uso de imagens hiperespectrais, como seu uso para identificação de manchas de sêmen em tecidos (SILVA et al., 2017); diferenciar manchas de sangue humano (SH), de manchas de sangue animal (SA) e possíveis falsos positivos comuns (FPC) depositadas na superfície de diferentes tecidos (PEREIRA et al., 2021); detectar e identificar folhas de *Cannabis sativa L.*(PEREIRA et al., 2020); além de diversas outras aplicações na área de documentoscopia, resíduos de tiros, identificação de traços e impressões digitais (CARNEIRO; SILVA; WEBER, 2023; MARIOTTI; ORTIZ; FERRÃO, 2023; RODRIGUES E BRITO et al., 2019). Fica claro, portanto, que esta técnica oferece aos investigadores um potencial significativo para detecção, visualização, identificação e estimativa de idade de vestígios forenses. As características de possibilitar aquisições rápidas, não destrutivas e sem contato do uso de HSI marcam sua adequação como ferramenta analítica para a ciência forense.

## 2.3 FERRAMENTAS QUIMIOMÉTRICAS

Os avanços das técnicas de instrumentação, como o desenvolvimento de equipamentos portáteis e câmeras hiperespectrais, ocasionaram não apenas o aumento do volume de dados gerados, mas também a necessidade do avanço de novas ferramentas de tratamento de dados. Nesse sentido, a Quimiometria é ferramenta essencial para auxiliar na exploração dos dados e no desenvolvimento de ferramentas e métodos capazes de criar modelos adequados para diferentes tipos de aplicação (BEEBE; PELL; SEASHOLTZ, 1998; BRERETON, 2003; BRERETON et al., 2017; SHARMA; KUMAR, 2018). Conceitualmente, as ferramentas quimiométricas podem ser compreendidas como a aplicação de ferramentas estatísticas e matemáticas a dados químicos (BROWN, 2017).

### 2.3.1 Ferramentas de Pré-Processamentos

O primeiro passo na aplicação das ferramentas quimiométricas envolve o exame dos dados coletados a fim de entender o problema e verificar a existência de amostras anômalas ou de possíveis efeitos indesejados como a presença de ruídos, ou variação de linha de base, por exemplo. Esses efeitos indesejados, muitas vezes, levam a fontes de variação aleatórias ou sistemáticas que podem mascarar as variações de interesse. Dessa forma, faz-se necessário o uso de ferramentas de pré-processamento (BEEBE; PELL; SEASHOLTZ, 1998; VIDAL; AMIGO, 2012). Essas técnicas modificam os dados e podem tanto melhorar ou piorar o desempenho dos modelos construídos a partir deles, quando são usados de maneira inapropriada (BEEBE; PELL; SEASHOLTZ, 1998).

Para espectros de infravermelho próximo é muito comum observar a presença de ruídos, e variação de linha de base, devido ao efeito do espalhamento da radiação. Os pré-processamentos mais utilizados para amenizar essas observações são a Padronização Normal de Sinal (SNV), Correção Multiplicativa de Sinal (MSC), Suavização, Derivada, entre outros (BEEBE; PELL; SEASHOLTZ, 1998; BUENDIA GARCIA et al., 2022; JIAO et al., 2020).

A Padronização Normal de Sinal (SNV, do inglês, *Standard Normal Variate*) é uma ferramenta de pré-processamento muito usada para amenizar a variabilidade causada por efeitos de espalhamento da radiação muito comuns em espectros de reflectância. Essa técnica funciona a partir de uma normalização dos espectros, subtraindo-se de cada

espectro a média dos espectros e dividindo-se pelo desvio padrão dos valores de intensidade de cada espectro (FEARN et al., 2009).

A suavização é muito utilizada para melhorar a relação sinal-ruído dos espectros. Existem diferentes algoritmos que podem ser utilizados, sendo que um dos mais comuns é o de Savitsky-Golay que utiliza uma função polinomial. Através desse método, seleciona-se um número ímpar de pontos que serão suavizados, dando-se o nome de janela a esse intervalo de pontos. A partir dessa janela, calcula-se uma função polinomial (geralmente de primeira ou segunda ordem) que irá se adequar a esses dados. A partir do polinômio calculado, estima-se o valor do ponto central da janela, e esse valor irá substituir o valor original daquele ponto. Esse processo é repetido para todo o conjunto de dados movendo-se a janela ao longo de todo o espectro (BEEBE; PELL; SEASHOLTZ, 1998; GEMPERLINE, 2006; OTTO, 2017).

A aplicação da derivada de Savitzky-Golay é a ferramenta mais utilizado de pré-processamento de derivada, e é usada para remover variação de linha de base, como a transformação da derivada é linear, as curvas produzidas retêm os aspectos quantitativos do sinal original, a diferença é que as regiões de máxima absorbância se tornarão pontos de inflexão, no caso da 1ª derivada. O algoritmo realiza, inicialmente, o mesmo processo descrito para a suavização, mas irá aplicar a derivada do polinômio na janela de pontos selecionada (BEEBE; PELL; SEASHOLTZ, 1998; GEMPERLINE, 2006; OTTO, 2017).

Algumas ferramentas de pré-processamento são mais voltadas para atenuar problemas relacionados a diferenças de amostras com características semelhantes (mesma classe) ou a transferência de calibração, como é o caso da ferramenta de pré-processamento chamada de Mínimos Quadrados Generalizados Ponderados (GLSW, do inglês, *Generalized Least Squares Weighting*). Ela funciona como um filtro de atenuação de interferentes, a partir da estimativa de uma matriz calculada a partir das diferenças entre as amostras que pertencem a uma mesma classe (EIGENVECTOR, 2015; WISE et al., 2001).

Ao aplicar esse pré-processamento, é necessário escolher um parâmetro $\alpha$ que é inversamente proporcional ao efeito do filtro, de modo que quanto menor o valor deste parâmetro, maior o efeito do filtro. Para selecionar o valor ideal de $\alpha$ podem ser utilizadas estratégias como a construção de superfícies de respostas, que se baseiam na construção de superfícies a partir dos valores da Raiz do Erro Quadrático Médio de Calibração (RMSEC) e da Raiz do Erro Quadrático Médio de Validação Cruzada (RMSECV) obtidos

a partir da construção de modelos de Regressão por Mínimos Quadrados Parciais (PLSR) (SILVA G.; SILVA S.; PIMENTEL, 2022).

### 2.3.2 Análise de Componentes Principais (PCA):

Após a realização da etapa de aplicação dos pré-processamentos, a grande dimensionalidade dos dados requer, muitas vezes, a utilização de ferramentas multivariadas para análise exploratória. Análise de Componentes Principais (PCA, do inglês, *Principal Component Analysis*) é uma das ferramentas mais utilizadas, pois realiza redução da dimensionalidade dos dados, evidenciando as informações relevantes facilitando sua interpretação (BEEBE; PELL; SEASHOLTZ, 1998; BRERETON, 2003; BRO; SMILDE, 2014; OTTO, 2017).

A PCA representa o conjunto de dados em um novo sistema de eixos, não correlacionados, posicionados na direção de maior variância dos dados. A matriz de dados **X** é decomposta em duas outras matrizes: **T** e $\mathbf{P^T}$, que são matrizes de *scores* e *loadings*, respectivamente.

$$\mathbf{X} = \mathbf{T}\mathbf{P^T} + \mathbf{E_X} \quad (1)$$

A matriz **X** é organizada de forma que as *i* linhas contenham as informações das amostras e as *j* colunas apresentam as variáveis originais do conjunto de dados. Já a matriz de *scores* **T** corresponde às coordenadas das amostras no novo sistema de eixos, também chamados de Componentes Principais (PC, do inglês, *Principal Component*). Essa matriz possui o mesmo número de linhas *i* da matriz X, mas o número de colunas mudará, e estará relacionado ao *k* número de componentes principais necessários para descrever a informação. A matriz de *loadings* $\mathbf{P^T}$ tem o mesmo número *j* de colunas da matriz de dados, mas o número de linhas será igual ao número *k* de componentes principais. A matriz $\mathbf{E_X}$ é a matriz de resíduos e terá as mesmas dimensões da matriz **X**.

### 2.3.3 Métodos de Classificação

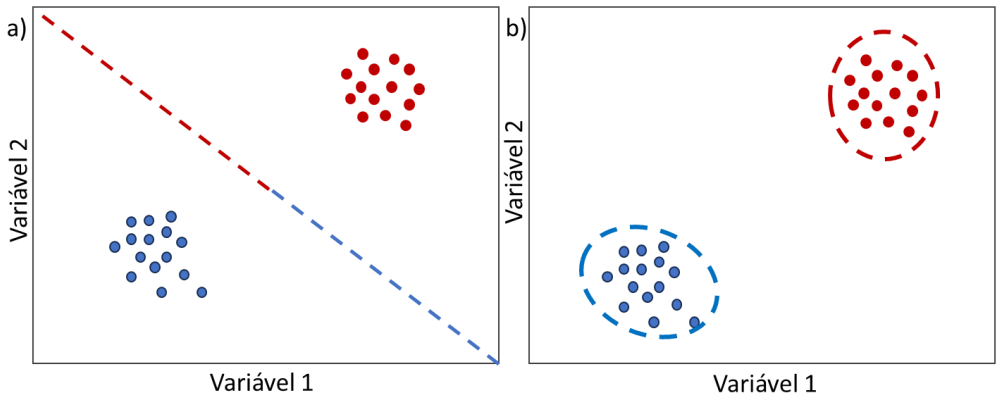
São ferramentas de reconhecimento de padrão supervisionado que utilizam classes para agrupar indivíduos (amostras) com características semelhantes. Cada grupo de amostras similares pode ser representado por um mesmo valor de variáveis discretas, e a partir disso, os modelos de classificação estabelecem relações matemáticas entre o vetor

de classes e as variáveis do conjunto de dados (OLIVERI, 2017; OLIVERI; DOWNEY, 2012).

Na perspectiva forense, as ferramentas de classificação são muito úteis para solucionar problemas de identificação, diferenciação e de adulteração de amostras (KUMAR; SHARMA, 2018). Há diversos estudos utilizando métodos de classificação para discriminar fluidos biológicos como diferentes tipos de sangue, sêmen, urina e saliva (CANO-TRUJILLO et al., 2023; FONSECA et al., 2022; MALEGORI et al., 2020; PEREIRA et al., 2021; SILVA et al., 2017; ZHANG et al., 2016).

As ferramentas de classificação podem ser subdividas em duas categorias: as ferramentas de modelagem de classes, e as de análises discriminantes. A principal diferença entre essas duas categorias é a forma como as fronteiras das classes são definidas. A diferença entre ambas as ferramentas pode ser compreendida a partir da observação da Figura 4.

**Figura 4 -** Representação esquemática do princípio de funcionamento de a) uma ferramenta de análise discriminante e b) uma ferramenta de modelagem de classes, em um exemplo ilustrativo envolvendo duas classes de amostras (pontos azuis e pontos vermelhos).




**Fonte:** Adaptado de (VITALE et al., 2023)

Nas ferramentas de análise discriminante, as fronteiras das classes são definidas usando informação das diferenças entre as classes, na busca de identificar as características que mais contribuem para a separação de classes para criar uma função discriminante. Dessa forma, nessas ferramentas, as classes (pontos azuis e vermelhos, na Figura 4a) são estritamente separadas e, portanto, quando se projeta uma amostra externa, ela sempre será atribuída a apenas uma classe. Há, ainda, a possibilidade de desenvolver modelos *soft* discriminantes, os quais podem, por exemplo, combinar ferramentas de

detecção de outliers para detectar amostras não pertencendo a nenhuma classe. (OLIVERI, 2017; OLIVERI; DOWNEY, 2012; VITALE et al., 2023).

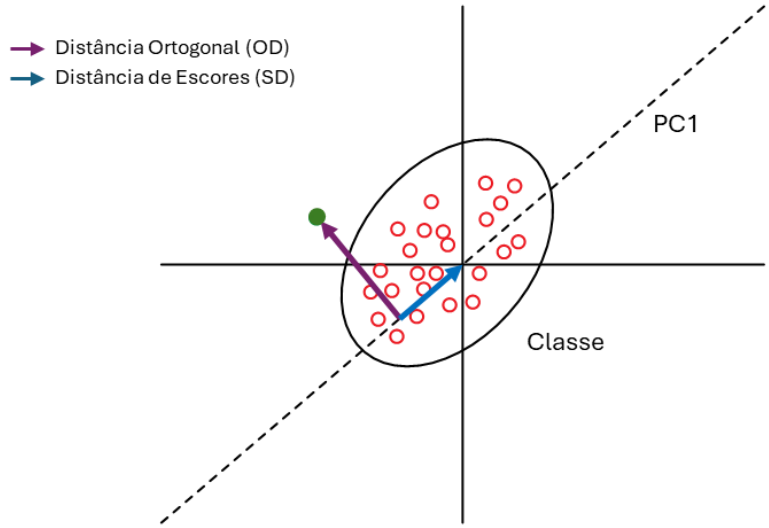
Já nas ferramentas de modelagem de classe, as fronteiras são determinadas a partir das características das amostras que pertencem à mesma classe, e nenhuma informação de amostras pertencentes a outras classes é utilizada para delimitar esses parâmetros, de modo que as amostras pertencem a classe em que têm maior probabilidade de serem encontradas (DURANTE; BRO; COCCHI, 2011; OLIVERI, 2017; POMERANTSEV; RODIONOVA, 2020; VITALE et al., 2023). Como consequência, se mais de uma classe for modelada, as amostras podem ser reconhecidas como membros de nenhuma, uma ou múltiplas categorias modeladas, o que torna a aplicação dessas ferramentas adequada quando se espera que as categorias investigadas constituam apenas uma parte daqueles que poderiam ser potencialmente encontrados e explorados (OLIVERI, 2017).

#### 2.3.3.1 Modelagem Independente e Flexível de Analogia de Classe (SIMCA)

A Modelagem Independente e Flexível de Analogia de Classe (SIMCA, do inglês, *Soft Independent Modelling of Class Analogy*) é uma ferramenta de modelagem de classes. Cada grupo de classes é modelado individualmente através de um modelo PCA que pode, inclusive, ter número de PCs diferentes para cada classe. As componentes principais vão definir o subespaço da classe, e quanto maior for a distância de uma observação em relação ao subespaço, maior será a probabilidade de a respectiva amostra não pertencer à classe que está sendo modelada (BEEBE; PELL; SEASHOLTZ, 1998).

Duas métricas de distância são comumente utilizadas para avaliar se uma amostra pertence a uma determinada classe: a distância euclidiana quadrada, também conhecida como distância ortogonal (OD, do inglês *Orthogonal Distance*, ou Q-Residual) e a distância quadrada de Mahalanobis também conhecida como distância de *scores* (SD, do inglês *Score Distance*, ou $T^2$ de Hotteling). A principal diferença entre essas distâncias é que a OD desconsidera a correlação entre as variáveis, enquanto a SD considera a correlação entre as variáveis (OLIVERI, 2017; VITALE et al., 2023). Simplificadamente, a distância ortogonal é a distância euclidiana perpendicular entre a amostra e a projeção da amostra em uma classe (ou o subespaço de uma classe), enquanto a distância de *scores* é a distância entre a projeção da amostra na classe e o centro da classe. Visualmente, essas distâncias podem ser representadas em um espaço vetorial conforme a Figura 5.

**Figura 5** - Esquema ilustrativo da distância ortogonal (OD) e da distância de *scores* (SD) de uma amostra em relação a uma classe hipotética.




**Fonte:** (Adaptado de Vitale et al., 2023).

Os valores dessas distâncias podem ser calculados a partir das Equações (2) e (3), respectivamente:

$$\boldsymbol{OD_{new}} = \mathbf{Q} = \|\mathbf{x_{new}^T}(\mathbf{I} - \mathbf{PP^T})\|^2 \quad (2)$$

$$\boldsymbol{SD_{new}} = \mathbf{T^2} = \mathbf{x_{new}}\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P^T}\mathbf{x_{new}^T} \quad (3)$$

Sendo $OD_{\text{new}}$, a distância ortogonal (Q-Residual) e $SD_{\text{new}}$ a distância de *scores* ($T^2$ de Hotteling) de um objeto, $\mathbf{x_{new}}$ é a medida do vetor (1xJ) da projeção do objeto predito, I é a matriz identidade, é **P** matriz de *loadings*, **Λ** (AxA) é uma matriz diagonal de autovalores, igual a $\frac{\mathbf{T^T T}}{\mathbf{N-1}}$ (sendo **T** a matriz de *scores* e N o número de amostras) e $\| \|^2$ simbolizada a norma. O símbolo $^{\mathbf{T}}$ sobreposto, indica que a matriz é transposta, e o $^{-1}$ sobreposto indica a inversa da matriz (VITALE et al., 2023).

Uma vez que as distâncias são calculadas, é necessário estabelecer um critério de decisão para estabelecer se a amostra pertence ou não a uma determinada classe. No geral, a grande maioria dos algoritmos que usam SIMCA definem o critério de decisão como um volume dentro do espaço em que as amostras têm maior probabilidade de serem localizadas baseados nos valores de ambas as distâncias (OD e SD), e distância é calculada conforme a Equação (4).

$$\mathbf{d_{new}} = \sqrt{\left(\frac{\mathbf{OD_{new}}}{\mathbf{OD_{crit}}}\right)^2 + \left(\frac{\mathbf{SD_{new}}}{\mathbf{SD_{crit}}}\right)^2} \quad (4)$$

Sendo que $d_{new}$ é a distância reduzida, $OD_{crit}$ e $SD_{crit}$ são os valores críticos da distância ortogonal e da distância dos *scores* da classe modelada, respectivamente. Existem diversas outras maneiras de combinar ambas as distâncias para definir o critério de decisão, e o SIMCA acaba recebendo outras denominações, como SIMCA alternativo, *Combined Index* SIMCA, *Data-Driven* SIMCA, entre outros. Além disso, o critério de decisão também pode ser estabelecido utilizando uma única distância, conforme as equações (5) e (6) (VITALE et al., 2023):

$$\boldsymbol{OD_{new}} = \mathbf{Q} \leq \mathbf{OD_{crit}} \quad (5)$$

$$\mathbf{SD_{new}} = \mathbf{T^2} \leq \mathbf{SD_{crit}} \quad (6)$$

Para aplicações forenses, o uso da modelagem SIMCA é bastante interessante, pois permite que a uma classe seja modelada independente das possíveis demais classes existentes. Além de possibilitar a classificação flexível, sem a obrigatoriedade de atribuir uma classe a uma amostra que não se adequa aos modelos de classes pré-estabelecidos.

#### 2.3.3.2 Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)

A ferramenta de Mínimos Quadrados Parciais (PLS, do inglês, *Partial Least Squares*) pode ser utilizada tanto para construir modelos de classificação quanto modelos de regressão (CHEVALLIER et al., 2006; WOLD et al., 2001). Assim como a PCA, a ferramenta de Mínimos Quadrados Parciais (PLS) é uma ferramenta multivariada de redução de dimensionalidade. No entanto, ao contrário da PCA, PLS usa, além das informações da matriz de dados **X**, informações fornecidas por uma outra matriz **Y,** que pode conter uma ou mais propriedades relacionadas a matriz de dados **X**, para determinar a direção do novo sistema de coordenadas garantindo a máxima covariância entre **X** e **Y** (BALLABIO; CONSONNI, 2013; RUIZ-PEREZ et al., 2020; WOLD et al., 2001).

O uso da informação proveniente da matriz **Y** faz com que PLS ao definir as novas variáveis, preserve na sua Primeira Variável Latente (LV1), o máximo de covariância possível entre os dados originais e as propriedades da matriz **Y**. Dessa forma, as direções

do sistema de coordenadas definido pela ferramenta PLS não são, necessariamente, as mesmas direções do sistema de coordenadas definido pela PCA, visto que na PCA, a transformação preserva em sua primeira componente principal o máximo possível de variação nos dados originais (RUIZ-PEREZ et al., 2020).

No caso da Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA, do inglês, *Partial Least Squares Discriminant Analysis*), o número de linhas da matriz **Y** é definida de acordo com as *n* amostras e o número de colunas serão as *m* classes e cada entrada $y_{nm}$ da matriz **Y** está relacionada a n-ésima amostra e a m-ésima classe e é expressa com um código binário (1 ou 0). Após a calibração do modelo PLS-DA, a probabilidade de uma amostra pertencer a uma classe específica pode ser calculada com base nos valores estimados da classe. De modo que, uma probabilidade é calculada para cada classe e a classificação da amostra é realizada escolhendo a classe que tem a maior probabilidade (CHEVALLIER et al., 2006).

### 2.3.4 Métodos de Calibração

A calibração multivariada é uma das áreas da quimiometria mais utilizadas cujo principal objetivo é estabelecer uma relação entre as variáveis de entrada e as variáveis de saída de uma forma que permita previsões das variáveis de saída quando apenas as variáveis de entrada estiverem disponíveis (OLIVIERI, 2018; BRAGA; POPPI, 2004). Existem diferentes ferramentas de calibração multivariada, sendo que as principais são a Regressão por Mínimos Quadrados Parciais (PLS), a Regressão Linear Múltipla (MLR, do inglês *Multiple Linear Regression*), Regressão por Componentes Principais (PCR, do inglês *Principle Component Regression*), o Método Clássico de Mínimos Quadrados (CLS, do inglês *Classical Least Squares*), entre outras (BEEBE; PELL; SEASHOLTZ, 1998; FEUDALE et al., 2002; FORINA; LANTERI; CASALE, 2007; BRAGA; POPPI, 2004).

Na perspectiva forense, as ferramenta s de calibração podem ser utilizadas não só para realizar a estimativa do tempo de deposição ou degradação de uma evidência, como manchas de sangue, fluidos biológicos, papéis e tintas de caneta, por exemplo (DAS et al., 2020; EDELMAN et al., 2012; EDELMAN; VAN LEEUWEN; AALDERS, 2012; KUMAR; SHARMA; SHARMA, 2020; MANIS et al., 2022; RODRIGUES E BRITO et al., 2019; SHARMA; KUMAR, 2018; SILVA et al., 2018; SILVA G.; SILVA S.; PIMENTEL, 2022; TAKAMURA et al., 2019). Também podem ser usadas para

determinar a concentração de substâncias de interesses, como quantificar cocaína e outras substâncias em drogas apreendidas (GROBÉRIO et al., 2015), e demais problemas forenses envolvendo adulteração de alimentos e outras substâncias (ABDULLAH SANI et al., 2021; DA SILVA et al., 2015).

2.3.4.1 Regressão por Mínimos Quadrados Parciais (PLSR)

A Regressão por Mínimos Quadrados Parciais (PLSR) é uma das ferramentas de calibração mais utilizadas (FORINA; LANTERI; CASALE, 2007; OLIVIERI, 2018; WOLD et al., 2001). Diferentemente da ferramenta do PLS-DA, em que a matriz **Y** é, uma matriz de variáveis discretas relacionadas às diferentes classes da amostra, na PLS-R, a matriz **Y** será, geralmente, um vetor **y** (PLS1), relacionado a uma propriedade quantitativa dos dados, como, por exemplo, a concentração de uma espécie.

De maneira simplificada, a regressão consiste em uma relação entre as matrizes de *scores* **T** e **U**, sendo que **T** é a matriz de *scores* de **X** e **U**, matriz de *scores* de **Y**. O modelo PLS pode ser considerado como composto por relações externas (bloco **X** e **Y** individualmente) e uma relação interna (ligando ambos os blocos). Matematicamente, essas relações são expressas através das Equações 7, 8 e 9 (GEMPERLINE, 2006).

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E}_{\mathbf{X}} \quad (7)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{L}^{\mathbf{T}} + \mathbf{F}_{\mathbf{Y}} \quad (8)$$

$$\mathbf{U} = \mathbf{T}\mathbf{W} \quad (9)$$

Sendo que $\mathbf{P}^{\mathbf{T}}$ e $\mathbf{L}^{\mathbf{T}}$ representam as matrizes de *loadings* de **X** e **Y**, respectivamente. **W** é a matriz de pesos e $\mathbf{E}_{\mathbf{X}}$ e $\mathbf{F}_{\mathbf{Y}}$ as matrizes residuais de **X** e **Y**, respectivamente. Essa relação entre as matrizes dos *scores* **T** e **U** estabelece que as novas variáveis do modelo devem ser obtidas de forma a maximizar a correlação entre as matrizes **X** e **Y**, como já mencionado. A partir dessas equações, podem ser calculado os coeficientes de regressão do modelo $\widehat{\mathbf{B}}$ que irão ser utilizados para realizar a predição dos valores a propriedade de interesse $\widehat{\mathbf{Y}}$, conforme as Equação 10 e 11 (GEMPERLINE, 2006).

$$\widehat{\mathbf{B}} = \mathbf{P}(\mathbf{P}^{\mathbf{T}}\mathbf{P})^{-1}\mathbf{W}\mathbf{L}^{\mathbf{T}} \quad (10)$$

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

### 2.3.5 Modelos Hierárquicos

Os modelos hierárquicos consistem em uma estratégia para associar mais de uma ferramenta quimiométrica em um formato de árvore de decisão, em que cada nó é representado por um modelo quimiométrico diferente e uma regra de decisão é definida de modo que um objeto só seguirá para a etapa seguinte se atender a essa regra (FONSECA et al., 2022; PEREIRA et al., 2021)

Muitas vezes, ao construir modelos voltados para problemas mais complexos - com muitas classes ou classes muito semelhantes, ou ainda, modelos que envolvam classificação e calibração -, as abordagens quimiométricas tradicionais não são tão eficazes e podem falhar. De modo que, faz-se necessário associar mais de uma ferramenta quimiométrica para solucionar o problema (MARCHI et al., 2022).

Os modelos hierárquicos podem ser utilizados combinando tanto modelos de classificação quanto modelos de regressão quando é necessário inicialmente classificar as amostras antes de realizar a estimativa de um determinado parâmetro através de um modelo de calibração. Essa estratégia facilita a construção de modelos mais fidedignos que representem melhor os dados. Além disso, as amostras externas só precisam ser testadas em um único modelo em que todas essas etapas são executadas em sequência.

### 2.3.6 Validação e Figuras de Mérito

Para que todos os modelos quimiométricos discutidos acima possam ser utilizados é necessário que eles sejam validados. A validação é o procedimento que garante que o modelo foi testado e é confiável. Esse procedimento envolve a estimativa da capacidade preditiva em novas amostras, não utilizadas para a construção do modelo, e a partir disso, são calculados alguns parâmetros – as Figuras de Mérito – que irão indicar a confiabilidade do modelo (OLIVERI, 2017; SILVA et al., 2012).

#### 2.3.6.1 Figuras de Mérito para Classificação

A validação de um modelo de classificação está estritamente relacionada à sua capacidade de prever corretamente a classe em que uma amostra pertence. Para construir o modelo, e validá-lo, o conjunto de dados é subdividido em conjunto de treinamento –

usado para construir o modelo - e conjunto de predição/previsão – usado para validar o modelo (OLIVERI, 2017). Geralmente, um percentual de 70% do conjunto de dados é utilizado como conjunto de treinamento e 30% do conjunto de dados é usado no conjunto de predição, e a seleção desses dados pode ser feita de maneira arbitrária, ou através de algum algoritmo de seleção de amostras. Ambos os conjuntos devem conter amostras de membros de classes conhecidas, e as amostras do conjunto de predição não podem ser utilizadas para a construção do modelo, para que seja evitado sobreajuste (OLIVERI, 2017; VITALE et al., 2023).

Para validar o modelo, as amostras do conjunto de predição são projetadas nele, e se as amostras pertencentes à classe de interesse forem corretamente classificadas pelo modelo, elas são consideradas verdadeiro positivo (VP), no entanto, se elas forem rejeitadas erroneamente, são chamadas de falso negativo (FN). Já as amostras que não pertencem à classe de interesse são rotuladas como falso positivo (FP) se forem erroneamente atribuídas à classe, e verdadeiro negativo (VN) se forem corretamente recusadas (OLIVERI, 2017; VITALE et al., 2023).

A partir dessa designação podem ser calculadas as principais figuras de mérito: sensibilidade e especificidade. A Sensibilidade (Sn) pode ser definida como a fração de amostras da classe de interesse que resultou em uma atribuição verdadeira positiva das amostras dentro de sua classe, e pode ser calculada através da Equação 12 (OLIVERI, 2017; VITALE et al., 2023).

$$\mathrm{Sn} = \frac{\mathrm{VP}}{\mathrm{VP} + \mathrm{FN}} \tag{12}$$

Enquanto a especificidade, mostra a fração de amostras estranhas à classe modelada que foram corretamente recusadas pelo modelo para a classe observada. O seu cálculo é obtido através da Equação 13 (OLIVERI, 2017; VITALE et al., 2023).

$$\mathrm{Sp} = \frac{\mathrm{VN}}{\mathrm{VN} + \mathrm{FP}} \tag{13}$$

2.3.6.2 Figuras de Mérito para Calibração

Em relação aos modelos de calibração multivariada, a capacidade preditiva de um modelo é geralmente expressa em termos da Raiz do Erro Quadrático Médio de Predição (RMSEP, do inglês *Root Mean Square Error of Prediction*) determinado para um Conjunto de Predição (CP), composto por amostras externas ao conjunto usado para construir o modelo (Conjunto de Treinamento, CT) (FABER, 1999). O RMSEP é um parâmetro derivado da Raiz do Erro Quadrático Médio (RMSE, do inglês *Root Mean Square Error*), que também pode ser calculado para a calibração (RMSEC, do inglês *Root Mean Square Error of Calibration*) e para validação cruzada (RMSECV, do inglês *Root Mean Square Error of Cross Validation*) (FERREIRA, 2015; NICOLAAS; FABER, 1999; SILVA et al., 2012). A forma utilizada para calcular o RMSEP pode ser observado na Equação (14):

$$\mathrm{RMSEP} = \sqrt{\frac{\sum(\mathrm{y_i} - \widehat{\mathrm{y_i}})^2}{\mathrm{N}}} \tag{14}$$

Sendo que $y_i$ é o valor correspondendo a propriedade de interesse referente a amostra *i*, $\widehat{y_i}$ é o valor dessa propriedade predito pelo modelo de calibração, e N é o número de amostras de previsão (FERREIRA, 2015).

Uma forma de comparar e avaliar os modelos de regressão é através do teste estatístico de Fischer, conhecido como teste F, o qual é frequentemente utilizado na comparação de duas variâncias populacionais para testar a hipótese de que ambas as variáveis não possuem relação. Nesse caso, o teste F pode ser utilizado para comparar a exatidão de dois modelos diferentes, verificando se suas variâncias de erro são significativamente diferentes, ou seja, comparam-se os valores de RMSEP.

Outro parâmetro importante para avaliar a existência de erros sistemáticos no modelo de calibração é o *bias*, ou viés, que indica o desvio sistemático do valor calculado em relação ao valor real, ele pode ser calculado através da Equação (15):

$$bias = \frac{\sum(\mathrm{y_i} - \widehat{\mathrm{y_i}})}{\mathrm{N}} \tag{15}$$

Do mesmo modo, o $y_i$ é o valor da propriedade de interesse da amostra i, $\widehat{y_i}$ é o valor dessa propriedade predito pelo modelo de calibração, e N é o número de amostras de previsão ou calibração.

# 3 MODELOS HIERÁRQUICOS PARA IDENTIFICAÇÃO DE AMOSTRAS DE SANGUE ENVELHECIDAS EM TEMPOS DISTINTOS

## 3.1 INTRODUÇÃO

A identificação de manchas de sangue é essencial para o trabalho do perito criminal em cenas de crimes violentos. Elas podem ser encontradas em diferentes tipos de superfícies, sendo que uma das superfícies mais comuns de serem encontradas são os pisos (BEVEL; GARDNER, 2008; ELKINS, 2019; TAYLOR et al., 2016). Estudos recentes demonstraram a utilidade de equipamentos portáteis para identificar manchas de sangue *in situ,* simulando locais e condições de crime (FONSECA et al., 2022; MORILLAS; GOOCH; FRASCIONE, 2018; PEREIRA et al., 2017). No entanto, ainda há muitos desafios a serem superados para uma aplicação prática, como a diversidade de superfícies onde essas manchas são encontradas, as condições ambientais às quais as evidências podem ser submetidas, além dos efeitos de degradação ao longo do tempo de deposição.

Em um contexto forense, manchas de sangue podem ser encontradas tanto frescas, com tempos de deposição de minutos ou frações de horas, quanto envelhecidas por um período indeterminado. Contudo, a maioria dos trabalhos desenvolvidos usando espectroscopia vibracional para identificação de manchas de sangue humano utilizam amostras frescas ou com um mesmo tempo de deposição para construção dos modelos de discriminação (FONSECA et al., 2022; MALEGORI et al., 2020; MORILLAS; GOOCH; FRASCIONE, 2018; PEREIRA et al., 2017).

O estudo pioneiro realizado por Pereira e colaboradores (2017) utilizou um equipamento portátil no infravermelho próximo (MicroNIR) para identificar e discriminar manchas de sangue humano, sangue animal e de sete substâncias com coloração avermelhada comumente confundidas com sangue, depositadas em diferentes superfícies (vidro, metal, cerâmica e porcelanato). Esse estudo avaliou modelos de classificação utilizando SIMCA, Análise Discriminante Linear combinada a Algoritmo Genético (GA-LDA, do inglês *Genetic Algorithm – Linear Discriminant Analysis)*, Análise Discriminante Linear combinada a Algoritmo de Projeções Sucessivas (SPA-LDA, do inglês *Successive Projection Algorithm – Linear Discriminant Analysis)*) e PLS-DA. Por ser um estudo preliminar, o tempo de deposição foi fixado em três dias, e os espectros foram coletados nesse mesmo intervalo de tempo, de modo que, mesmo

obtendo excelentes resultados de sensibilidade e especificidade utilizando modelos individuais para cada superfície, não se pode afirmar que esses modelos poderiam ser utilizados para identificar manchas de sangue com diferentes tempos de deposição (PEREIRA et al., 2017).

O trabalho desenvolvido por Morillas *et al.* (2018) utilizou um dispositivo portátil no infravermelho próximo (SCiO®NIR) para diferenciar manchas de sangue de substâncias semelhantes ao sangue que também podem estar presentes na cena de um crime. Doze modelos foram construídos utilizando PLS e diferentes ferramentas de pré-processamento, sendo que um dos modelos foi capaz de identificar o sangue corretamente em 94% dos casos, com uma taxa de falsos positivos de 14% das amostras semelhantes a sangue. No entanto, os espectros das manchas também foram coletados em um mesmo intervalo de tempo (duas horas), de modo que um maior desenvolvimento e ampliação do estudo é necessário para que essa metodologia esteja apta a ser usada em cenas de crime (MORILLAS; GOOCH; FRASCIONE, 2018).

O trabalho realizado por Fonseca e colaboradores (2022) utilizou o mesmo equipamento portátil do trabalho descrito por Pereira *et al.* (2017) e construiu modelos hierárquicos capazes de discriminar manchas de sangue humano, sangue animal e falsos positivos comuns, depositadas em nove pisos diferentes. O tempo de deposição das manchas também foi constante, seis dias, e os modelos tiveram excelentes resultados de sensibilidade e especificidade (FONSECA et al., 2022). Da mesma forma que os estudos anteriormente mencionados, faz-se necessário que a metodologia seja ampliada para diferentes tempos de deposição para que ela possa vir a ser utilizada em um contexto forense real.

### 3.1.1 Objetivos

O objetivo principal do presente trabalho foi desenvolver modelos hierárquicos capazes de identificar manchas de sangue depositadas em diferentes pisos e tempos. Nesse sentido, os objetivos específicos deste trabalho foram:

- Avaliar o potencial dos modelos hierárquicos para identificar amostras de sangue humano e sangue animal com tempo de deposição entre 0 dias e 176 dias.

- ❖ Desenvolver modelos hierárquicos otimizados incorporando amostras com tempo de deposição entre 0 dias e 1 dia ao conjunto de treinamento para identificar amostras de sangue humano e sangue animal com tempo de deposição entre 0 dias e 176 dias.

## 3.2 MATERIAIS E MÉTODOS

### 3.2.1 Preparação da amostra

Foram depositadas na superfície de nove pisos diferentes (cinco cerâmicas e quatro porcelanatos), 132 amostras de sangue humano (SH) coletadas com agulhas individuais esterilizadas diretamente nos capilares dos dedos de 22 voluntários saudáveis (dez homens e doze mulheres). Também foram depositadas nos substratos, amostras de sangue animal (SA) de cinco animais (três carneiros e dois cavalos) cujo sangue foi fornecido pelo hospital veterinário da Universidade Federal Rural de Pernambuco (UFRPE). O volume do sangue e o número de gotas de sangue depositados nos substratos não foram controlados. Todas as amostras de sangue foram coletadas de acordo com as exigências do Comitê de Ética em Pesquisa envolvendo Seres Humanos da Universidade Federal de Pernambuco (nº 1.059.225).

Mais detalhes da coleta de amostras podem ser vistos no procedimento operacional descrito por Pereira e colaboradores (2017).

### 3.2.2 Aquisição dos Espectros

Para a aquisição dos espectros, foi utilizado um espectrômetro MicroNir 1700 fabricado pela VIAVI. O equipamento tem dimensões de 45 mm de diâmetro e 42 mm de altura e 60 g de peso e um filtro linear variável (LVF) acoplado a uma matriz de detectores de arseneto de gálio e índio (InGaAs). Os espectros foram registrados na faixa de 908–1676 nm com resolução de 1,25% do comprimento de onda central e obtidos com 64 varreduras e tempo de integração de 5 milissegundos.

Os espectros das amostras de sangue humano foram coletados em 0, 1, 9, 20, 30 e 162 dias e os espectros de sangue animal foram coletados em 0, 1, 2, 9, 20, 29, 105 e 176 dias. Quatro espectros foram adquiridos de diferentes localizações de cada mancha. Obteve-se um total de 4320 espectros sanguíneos: 2941 foram espectros de SH e 1379 foram espectros de SA.

A distribuição da quantidade de espectros coletados por dia e por doador pode ser observada na Tabela 1. Em sua dissertação de mestrado, a autora desenvolveu modelos hierárquicos combinando ferramentas quimiométricas e utilizou três diferentes conjuntos de treinamento para avaliar a influência do piso (substrato) no qual as manchas encontravam-se depositadas, o intuito desse trabalho era diferenciar manchas de sangue humano, animal e substâncias comerciais comumente confundidas com sangue.

Considerando que um dos objetivos desta etapa da tese foi avaliar a capacidade de classificação do melhor modelo construído por Fonseca e colaboradores (2022) com relação a influência de amostras em diferentes tempos de deposição, o conjunto de dados de treinamento para análise quimiométrica e a construção de modelos de classificação hierárquica foram estabelecidos e detalhados neste trabalho anterior, sendo aqui apresentados de forma resumida.

**Tabela 1 -** Quantidade de espectros de sangue humano e sangue animal coletados em cada dia.

| **Dia** | **0** | **1** | **9** | **10** | **13** | **20** | **29** | **30** | **105** | **162** | **176** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **SH** | 492 | 487 | 492 | - | - | 486 | - | 484 | - | 500 | - |
| **SA** | 201 | 184 | - | 200 | 193 | 201 | 201 | - | 200 | - | 200 |

### 3.2.3 Análise e Pré-Processamento de Dados

O tratamento dos dados e a análise estatística foram realizados no software Matlab (MATLAB® R2010a 7.10.0.499, MathWorks) e os modelos hierárquicos, assim como os modelos PCA, SIMCA e PLS-DA foram realizados no PLS_Toolbox (Eigenvector Research, Inc).

Seguindo o procedimento descrito e estabelecido pela autora em sua dissertação de mestrado (FONSECA, 2020) que também foi discutido em seu artigo publicado em 2022 (FONSECA, 2022), os espectros de todas as amostras foram cortados na faixa de 970-1600 nm. A ferramenta de 1ª derivada de Savitzky Golay com 15 pontos de suavização e polinômio de 2º grau foi aplicada a todos os dados.

### 3.2.4 Conjuntos de Treinamento e Predição

Os conjuntos de treinamento e predição foram selecionados a partir de um estudo prévio realizado durante o mestrado da autora (FONSECA, 2020) em que foram selecionadas para compor o Conjunto de Treinamento 1 (CT1) amostras depositadas em

seis diferentes pisos (CE1, CE3, CE5, PO1, PO3 e PO4) e para o Conjunto de Predição 1 (CP1) foram usadas amostras depositadas em três diferentes pisos (CE4, PO5 e CE6) (FONSECA, 2020). Essa seleção de amostras foi respeitada para as amostras com mesmo tempo de deposição incluídas no conjunto de treinamento.

Em relação as amostras com tempos de deposição diferentes (que não estavam incluídos no conjunto de treinamento), todas foram incluídas no Conjunto de Predição independente do piso no qual a amostra estava depositada - por exemplo, se a mancha de sangue tinha tempo de deposição diferente das manchas de sangue usadas no CT (por exemplo, 13 dias), e estava depositada em um dos pisos usados no CP (CE4, PO5 ou CE6), ainda assim, ela foi incluída no Conjunto de Predição.

### 3.2.5 Avaliação dos Modelos Hierárquicos para Classificação

Como já mencionado, a construção de modelos de classificação hierárquica é uma estratégia quimiométrica utilizada quando uma única ferramenta de classificação, como SIMCA e PLS-DA, não é suficiente para resolver o problema de classificação. Esses modelos funcionam como uma árvore de decisão onde diferentes ferramentas de classificação são conectadas em série sendo separadas por uma regra de decisão (FONSECA et al., 2022; PEREIRA et al., 2021).

No estudo desenvolvido por Fonseca et *al*. (2022) foram construídos diferentes modelos hierárquicos utilizando amostras de sangue depositadas durante seis dias, e duas regras de decisão foram estabelecidas para o desenvolvimento de cada modelo hierárquico. A primeira regra utilizou como critério de decisão o Q-Residual obtido através de um modelo de PCA construído utilizando apenas amostras de sangue humano e sangue animal. Esta regra é responsável por separar os falsos positivos comuns (FPC) - substâncias que podem ser confundidas com sangue - das amostras de sangue e foi a mesma para todos os modelos. Já a segunda regra de decisão, que separou as amostras de sangue humano das amostras de sangue animal, foi diferente para cada modelo: o primeiro modelo hierárquico (MH1) utilizou como critério de decisão o Q-Residual, mas obtido usando um modelo de PCA construído utilizando amostras de sangue humano, o segundo (MH2) utilizou um modelo de SIMCA e o terceiro (MH3) utilizou um modelo de PLS-DA.

Para avaliar o efeito do tempo nos resultados da classificação, os espectros de manchas de sangue depositadas em diferentes tempos foram projetados no modelo

hierárquico original e as figuras de mérito, sensibilidade e especificidade, foram calculadas.

Outros três modelos foram desenvolvidos e avaliados com o objetivo de otimizar os modelos para melhorar sua performance em relação à identificação de amostras com diferentes tempos de deposição. Nestes novos modelos, espectros de sangue humano e sangue animal com zero dias (aproximadamente com 45 min de tempo de deposição) e um dia foram inseridos separadamente e simultaneamente, no conjunto de treinamento. Estes foram selecionados com base nos tipos de piso utilizados no conjunto de treinamento original (CE1, CE3, CE5, PO1, PO3 e PO4) e conjunto de previsão (CE4, CE6 e PO5), conforme descrito na metodologia detalhada na dissertação de mestrado da autora (FONSECA, 2020).

A Tabela 2 mostra a distribuição dos espectros no conjunto de treinamento e os diferentes tipos de modelos hierárquicos que foram construídos. Enquanto a Tabela 3, mostra a distribuição dos espectros de validação externa utilizados para cada modelo.

**Tabela 2 -** Resumo dos modelos hierárquicos construídos.

| **Modelo Hierárquico*** | **Tempo de Deposição do Conjunto de Treinamento** | **Conjunto de Treinamento** | | **1ª Regra de Decisão** | **2ª Regra de Decisão** |
|---|---|---|---|---|---|
| | | **SH** | **SA** | | |
| MH1.a | 6 dias | 365 | 168 | PCA | PCA |
| MH1.b | 6 + 0 dias | 721 | 338 | | |
| MH1.c | 6 + 1 dias | 716 | 323 | | |
| MH1.d | 6 + 0 + 1 dias | 1072 | 493 | | |
| MH2.a | 6 dias | 365 | 168 | PCA | SIMCA |
| MH2.b | 6 + 0 dias | 721 | 338 | | |
| MH2.c | 6 + 1 dias | 716 | 323 | | |
| MH2.d | 6 + 0 + 1 dias | 1072 | 493 | | |
| MH3.a | 6 dias | 365 | 168 | PCA | PLS-DA |
| MH3.b | 6 + 0 dias | 721 | 338 | | |
| MH3.c | 6 + 1 dias | 716 | 323 | | |
| MH3.d | 6 + 0 + 1 dias | 1072 | 493 | | |

* O primeiro número dos modelos refere-se à ferramenta utilizada como segunda regra (MH1 – PCA, MH2 – SIMCA e MH3 – PLS-DA). A letra refere-se à idade do conjunto de treinamento utilizado (MH*.a – amostras com 6 dias; MH*.b – 6 dias e amostras frescas (6 + 0); MH*.c – amostras de seis e um dias (6 + 1 ) dias); MH*.d – amostras de seis, zero e um dia (6 + 0 + 1) dias.

**Tabela 3 -** Distribuição dos espectros utilizados na validação externa (predição).

| **Validação Externa** | | | | |
|---|---|---|---|---|
| | **MH com 6 dias** | **MH com 6 + 0 dias** | **MH com 6 + 1 dia** | **MH com 6 + 0 + 1 dias** |
| **Dias** | **N° de Espectros de Sangue Humano** | | | |
| 0 | 492 | 136 | 492 | 136 |
| 1 | 487 | 487 | 136 | 136 |
| 9 | 492 | 492 | 492 | 492 |
| 20 | 486 | 486 | 486 | 486 |
| 30 | 484 | 484 | 484 | 484 |
| 162 | 500 | 500 | 500 | 500 |
| **Dias** | **N° de Espectros de Sangue Animal** | | | |
| 0 | 201 | 31 | 201 | 31 |
| 1 | 184 | 184 | 32 | 32 |
| 10 | 200 | 200 | 200 | 200 |
| 13 | 193 | 193 | 193 | 193 |
| 20 | 201 | 201 | 201 | 201 |
| 29 | 201 | 201 | 201 | 201 |
| 105 | 200 | 200 | 200 | 200 |
| 176 | 200 | 200 | 200 | 200 |

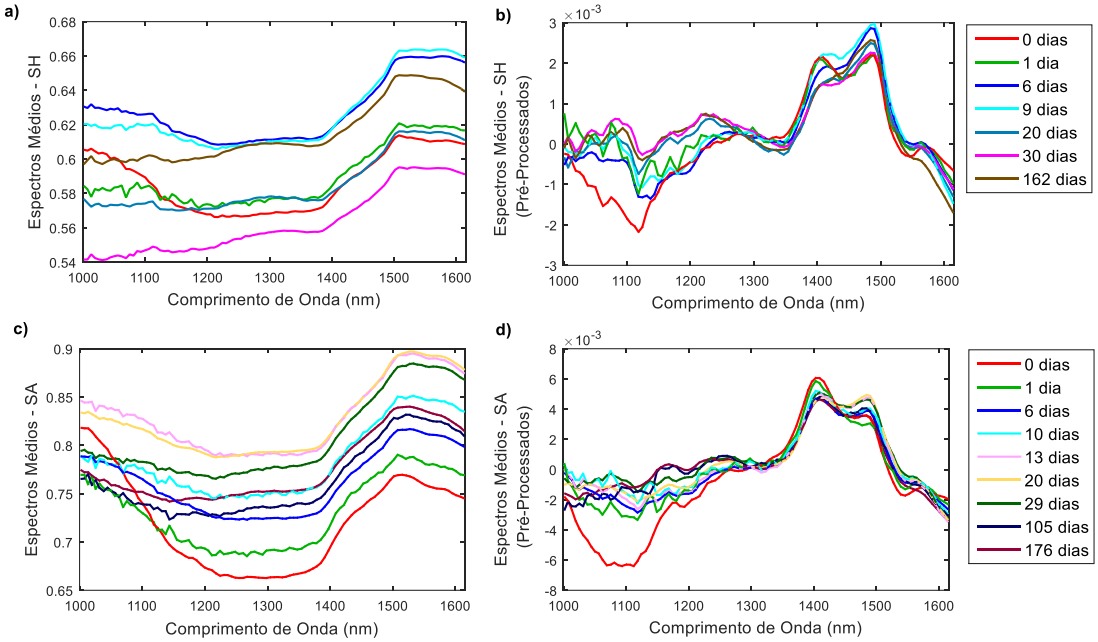
## 3.3 RESULTADOS E DISCUSSÃO

### 3.3.1 Análise Espectral

Para avaliar se as amostras de sangue apresentavam diferenças espectrais ao longo do envelhecimento, todas as amostras foram analisadas usando a metodologia de Fonseca et *al*. (2022). A ferramenta de pré-processamento 1ª derivada de Savitzky-Golay com 15 pontos de suavização e polinômio de 2º grau foi aplicada nos espectros e um modelo de PCA foi construído.

Inicialmente, as amostras de sangue humano e de sangue animal foram avaliadas separadamente. As Figura 6a e a Figura 6b mostram a média dos espectros originais e pré-processados das amostras de sangue humano. Enquanto as Figuras 6c e 6d mostram a média dos espectros original e pré-processado de sangue animal. Como pode ser visto nessas figuras, os espectros NIR são muito suscetíveis a influências físicas devido à superfície dos substratos, o que causa efeitos de espalhamento de radiação e ruído.

**Figura 6 -** Média dos espectros: a) originais de sangue humano; b) pré-processados de sangue humano; c) originais de sangue animal; d) pré-processados de sangue animal.

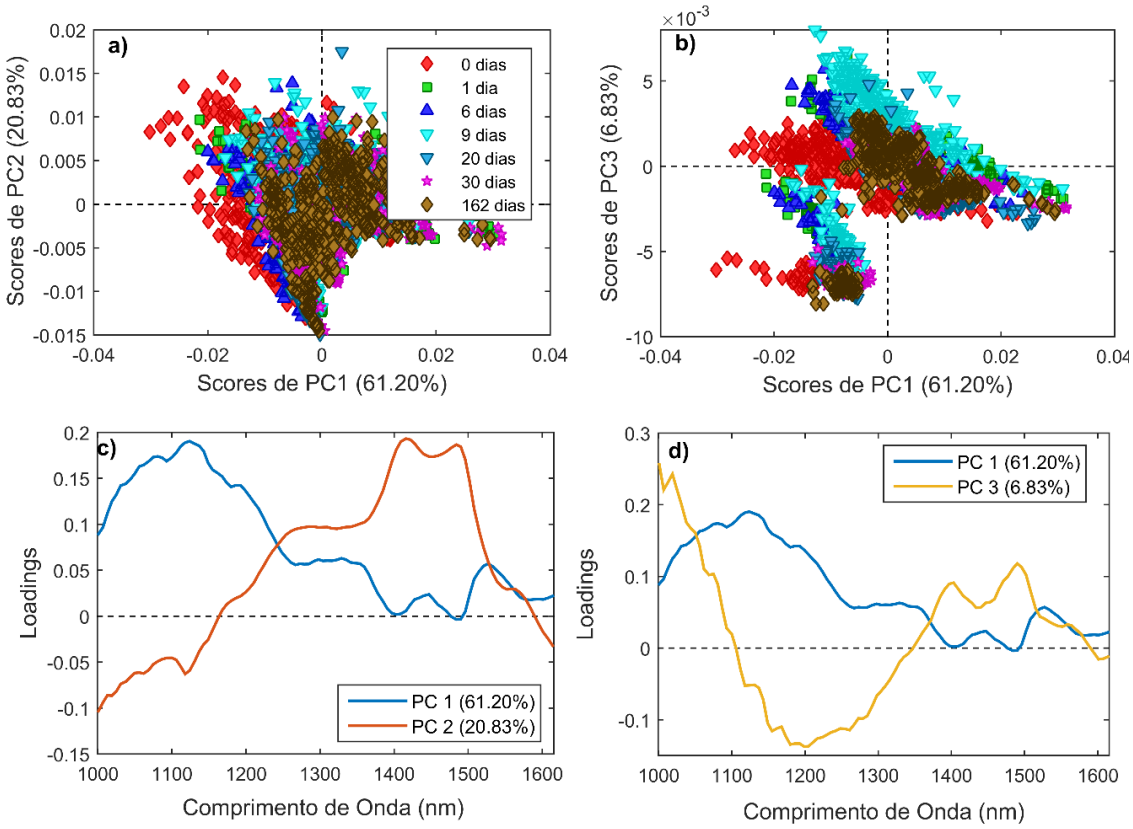



**Fonte:** A autora (2024).

Na Figura 6a, o espectro médio do dia 0 (em vermelho) apresenta uma banda mais intensa na região em torno de 1000 nm. Esse comportamento era esperado, pois essa região ainda apresenta parte da banda de combinação de vibrações de estiramento simétrico e assimétrico da ligação $H-O-$ presente em moléculas de água que se encontra em torno de 985 nm (EDELMAN et al., 2012). Além dessa região, a faixa entre 1440 e 1485 nm também está associada ao primeiro sobretom de estiramento da ligação $-O-H$. Entre 1347 e 1367 nm ocorre a banda de absorção espectral associada a combinações do segundo sobretom da ligação $-C-H$ em grupos metil ($-CH_3$) que pode estar associado à molécula de hemoglobina (Anexo A). Enquanto em torno de 1511 nm o primeiro sobretom de estiramento da ligação $-N-H$ encontrada em proteínas (BURNS; CIURCZAK, 2009; EDELMAN et al., 2012)

A Figura 7 mostra os modelos de PCA construídos utilizando apenas espectros pré-processados de sangue humano. As Figuras 7a e 7b mostram os *scores* de PC1xPC2, e de PC1xPC3, respectivamente. As cores estão relacionadas com os diferentes tempos de envelhecimento. Pode-se constatar que não há formação de agrupamentos significativos em relação ao tempo de envelhecimento do sangue.

**Figura 7 -** Modelo de PCA construído a partir dos espectros de sangue humano pré-processados com 1ª Derivada de Savitzky-Golay; a) *Scores* de PC1xPC2; b) *Scores* de PC1xPC3; c) *Loadings* de PC1 e PC2; d) *Loadings* de PC1 e PC3.

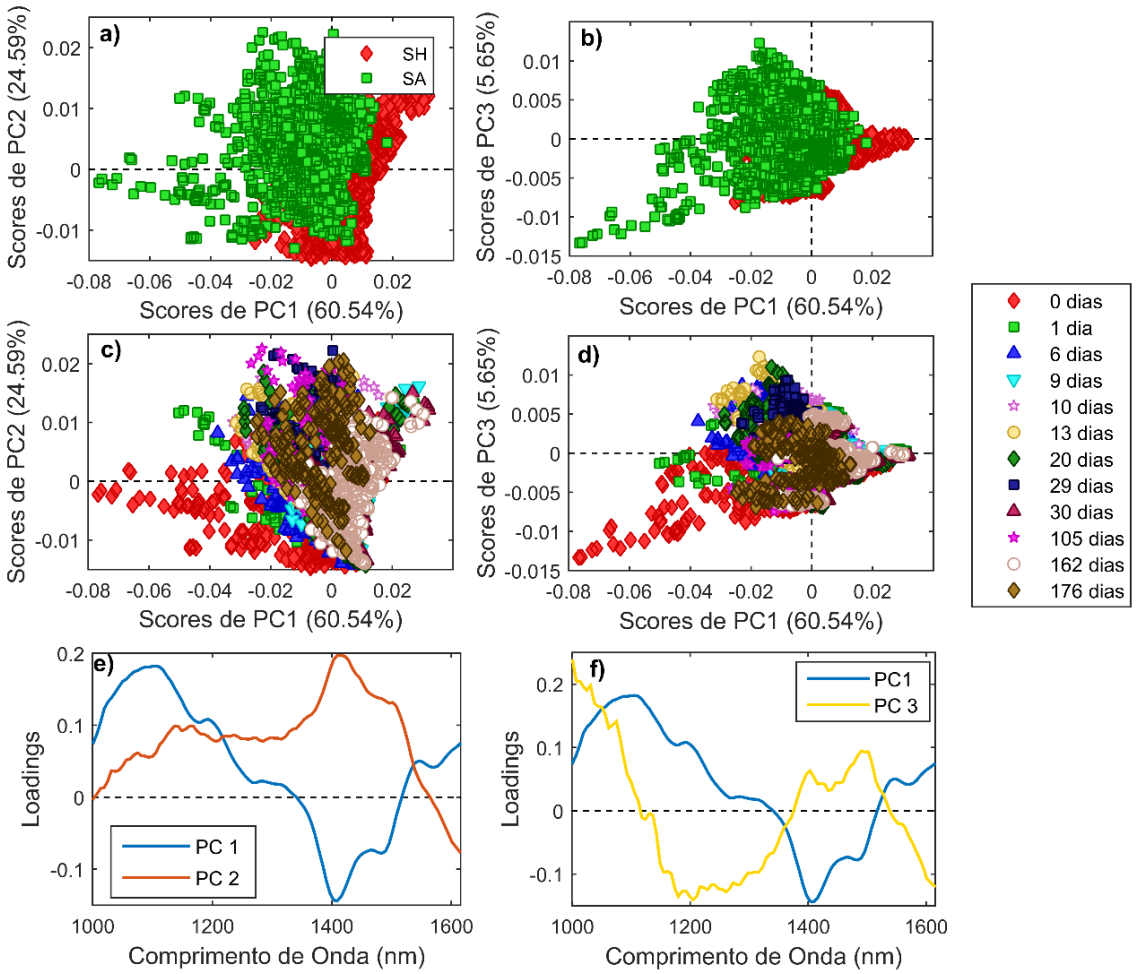



**Fonte:** A autora (2024).

Como era esperado, a influência do substrato no qual as amostras estão depositadas é um fator relevante que influencia no comportamento espectral. Na Figura 7b, pode ser observada a formação de um agrupamento com *scores* mais negativos em PC3. Todos os espectros referentes a esse agrupamento estão depositados em um mesmo piso, CE5, cujo comportamento é visivelmente diferente dos demais. Esse padrão é semelhante ao observado no estudo realizado com as amostras de seis dias (FONSECA et al., 2022; FONSECA, 2020), o que indica que o modelo original pode ser utilizado para identificar amostras de sangue com diferentes tempos de deposição.

Um modelo de PCA também foi construído utilizando todas as amostras de sangue (humano e animal), conforme visto na Figura 8.

**Figura 8 -** Modelo de PCA construído com os espectros pré-processados de sangue humano e sangue animal. a) *Scores* de PC1xPC2 (coloridas de acordo com as classes de doadores); b) *Scores* de PC1xPC3 (coloridas de acordo com as classes de doadores); c) *Scores* de PC1xPC2 (coloridas de acordo com o tempo de deposição) d) *Scores* de PC1xPC3 (coloridas de acordo com o tempo de deposição); e) Gráfico de *Loadings* de PC1 e PC2; f) Gráfico de *Loadings* de PC1 e PC3.



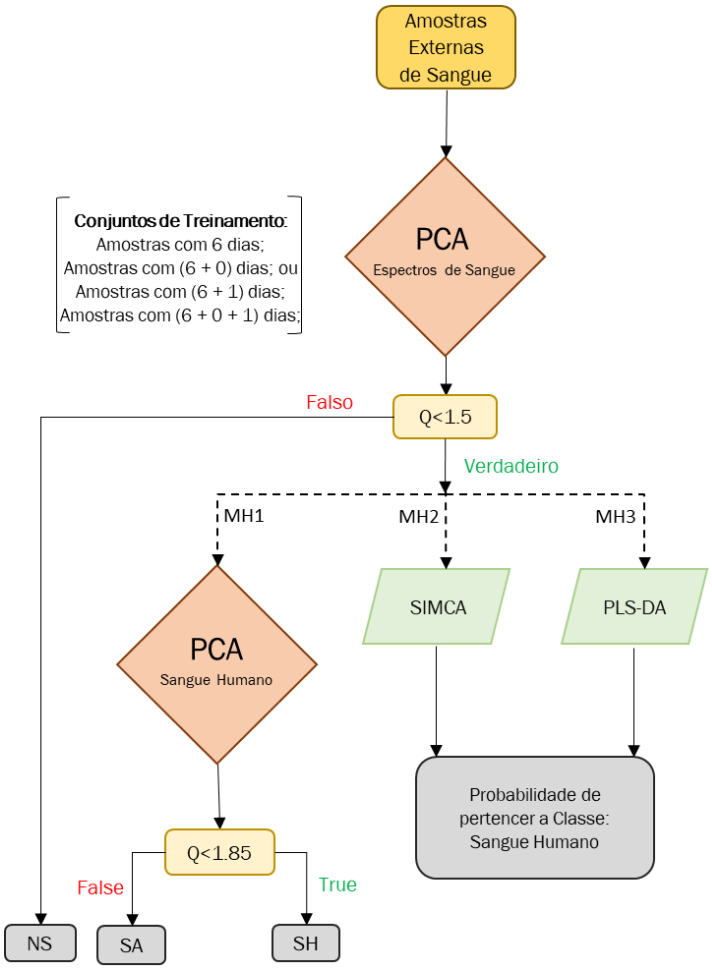

**Fonte:** A autora (2024).

Na Figura 8c, os *scores* de amostras frescas de sangue animal apresentam coordenadas mais negativas em PC1, indicando uma contribuição mais forte de variáveis na faixa entre 1300 e 1500 nm, ainda vista no gráfico de *loadings* (Figuras 8e e 8f). Como mencionado anteriormente, a região em torno de 1300 nm está associada a combinações do segundo tom da ligação $-C-H$ em grupos metil ($-CH_3$) que pode estar associada a presença de hemoglobina, e a região em torno de 1500 nm está relacionada a presença de proteínas. Essa diferença de comportamento entre as amostras mais recentes e as amostras mais antigas pode estar relacionada à oxidação e degradação da hemoglobina que acontece mais rapidamente nas primeiras horas após o contato com o ar atmosférico.

### 3.3.2 Avaliação dos Modelos Hierárquicos

Os espectros de sangue humano e animal com os diferentes tempos de deposição (0, 1, 9, 10, 13, 20, 29, 30, 105, 162 e 176 dias) foram inseridos como um conjunto de predição nos três modelos de classificação hierárquica originais construídos com as amostras de seis dias do Conjunto de Treinamento 1 (CT1) (FONSECA, 2020). Esse conjunto de treinamento possuía espectros de amostras de sangue depositadas em cinco diferentes pisos: CE1, CE3, CE5, PO1, PO3 e PO4. Um esquema que mostra o fluxograma do funcionamento dos modelos hierárquicos de classificação construídos pode ser observado na Figura 9.

**Figura 9** - Esquema representativo dos modelos hierárquicos construídos.




**Fonte:** A autora (2024).

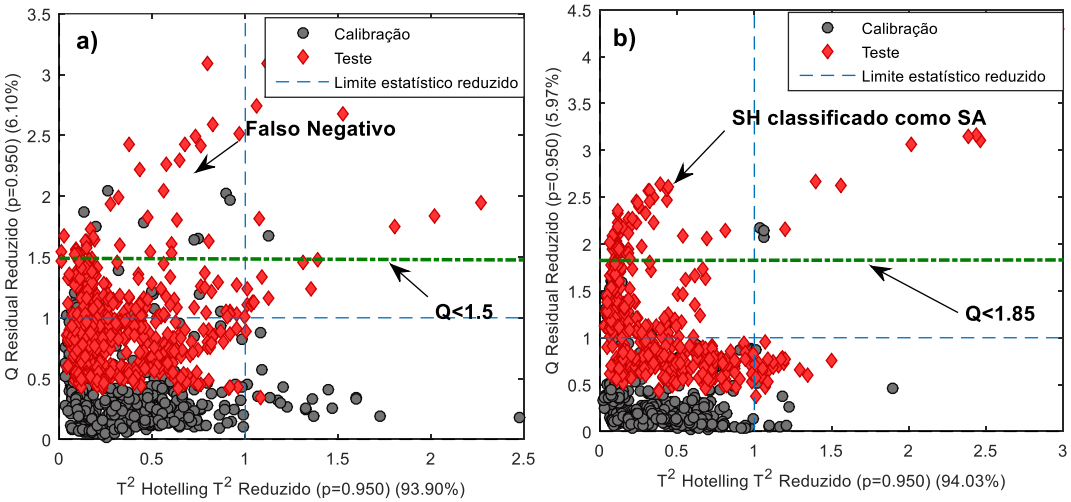
#### 3.3.2.1 Modelo Hierárquico 1

A primeira regra de decisão foi baseada em um modelo de PCA construído para separar amostras de sangue das amostras de falsos positivos comuns. Esse modelo

utilizou o critério Q-Residual (Q<1,5). A segunda regra de decisão também utilizou um modelo de PCA construído apenas com as amostras de sangue humano, com o intuito de separar as amostras de SH das amostras de SA, o critério de decisão também foi o Q-Residual: Q <1,85. Esses valores de Q-Residual foram estabelecidos no trabalho desenvolvido pela autora durante o mestrado (FONSECA et al., 2022; FONSECA, 2020). Essa estratégia é semelhante a utilizada por modelos de classificação construídos por SIMCA, mas, ao contrário do SIMCA tradicional, que se baseia no Q-Residual e no T² de Hotteling, ela é baseada em apenas um único critério, o que a torna mais flexível.

Para uma melhor compreensão dos critérios de decisão utilizados para classificar as amostras, a Figura 10 mostra a projeção dos *scores* de sangue humano fresco (0 dias de idade) no gráfico de influência para ambos os modelos PCA originais (construídos com amostras de seis dias). A Figura 10a mostra que algumas amostras de sangue foram classificadas erroneamente como Não-Sangue (NS). A maioria dessas amostras foi depositada sobre os mesmos pisos (CE3), o que confirma a influência da variabilidade da superfície do substrato, conforme discutido anteriormente (FONSECA et al., 2022; FONSECA, 2020). Na Figura 10b, algumas das amostras de SH foram classificadas como SA.

**Figura 10** - Gráficos de influência; a) Soma quadrada dos resíduos versus gráfico T² de Hotelling do primeiro modelo de PCA construído com amostras de sangue com 6 dias (Modelo Original); b) Gráfico de resíduos do segundo modelo de PCA construído com amostras de sangue humano com 6 dias (Modelo Original).



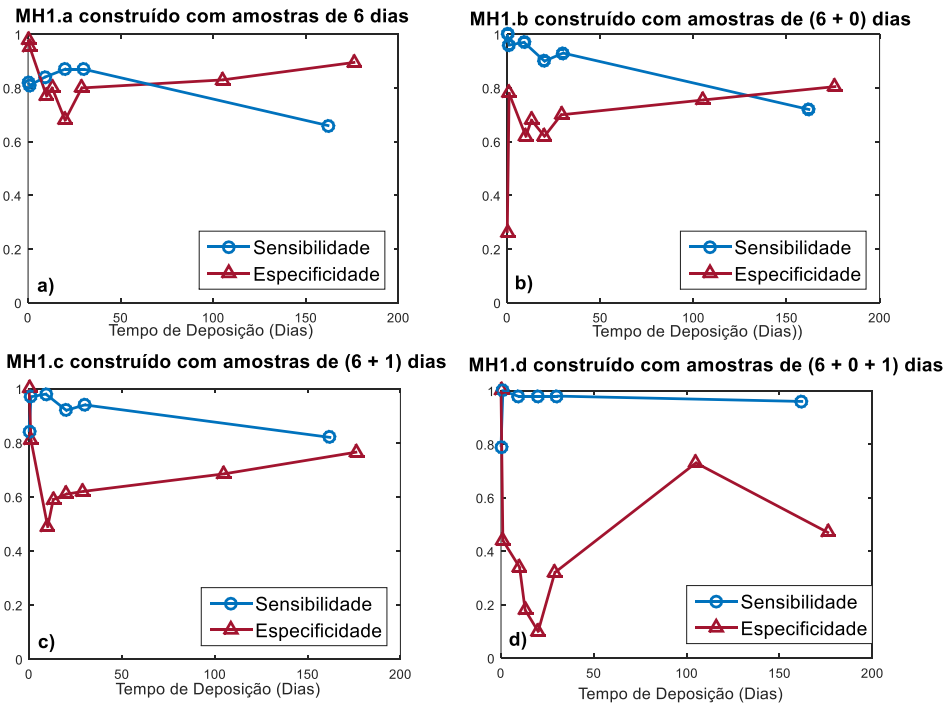

**Fonte:** A autora (2024).

A Figura 11a mostra a sensibilidade e especificidade do modelo original. Os resultados de sensibilidade para amostras de sangue humano foram superiores a 0,80 para

amostras depositadas entre zero e trinta dias, e superiores a 0,85 para amostras entre nove e trinta dias, indicando bom desempenho na identificação das amostras neste intervalo de tempo. Contudo, a sensibilidade para amostras com 162 dias foi de 0,66. Este resultado de sensibilidade, relativamente baixa, era esperado devido ao processo natural de degradação das manchas de sangue ao longo do tempo.

É importante ressaltar que algumas amostras de SH foram classificadas como SA e apenas 6% das amostras foram classificadas erroneamente como NS. Em relação à especificidade, os resultados obtidos foram melhores para amostras mais frescas, o que indica que os resultados de especificidade estavam intimamente relacionados com o tempo de envelhecimento das manchas de sangue, uma vez que amostras antigas de sangue animal e humano foram ambas classificadas como NS. Embora o modelo original não tenha conseguido classificar algumas das amostras mais antigas de sangue humano, o modelo ainda classificou como sangue humano 94% das amostras de 162 dias usando a primeira regra de decisão (6% foram rejeitadas como NS).

**Figura 11 -** Resultados de Sensibilidade e Especificidade do Modelo Hierárquico 1 (MH1) construído com a) amostras de 6 dias; b) amostras de 6 + 0 dias; c) amostras de 6 + 1 dias e d) amostras de 6 + 0 +1 dias.




**Fonte:** A autora (2024).

Para melhorar a robustez dos modelos, amostras de manchas de sangue com zero e um dia de depósito foram adicionadas separadamente ao conjunto de treinamento e

outros dois modelos hierárquicos foram construídos usando a mesma metodologia. Os resultados para cada um dos modelos são mostrados nas Figuras 11b e 11c, respectivamente. Uma melhoria foi observada em ambos os modelos para amostras frescas. Para o modelo construído com amostras de seis e zero dias de depósito, a sensibilidade foi de 1,0 para amostras de SH fresco, o que significa que 100% das amostras de SH com zero dias foram classificadas corretamente.

As amostras de sangue com um dia obtiveram um aumento de sensibilidade que saiu de 0,82 para 0,97. Para o modelo construído com amostras de seis e um dia, a sensibilidade para amostras de SH com zero dias aumentou para 0,84 e para amostras com um dia aumentou para 0,97. Em relação ao modelo MH1.d, que foi construído com amostras de seis, zero e um dia, também houve melhora na sensibilidade em relação a MH1.a, no entanto, os resultados de especificidade foram baixos, levando à confusão entre amostras de sangue animal e amostras de sangue humano. Isto indica que o MH1.d pode não diferenciar com confiabilidade os dois tipos de sangue.

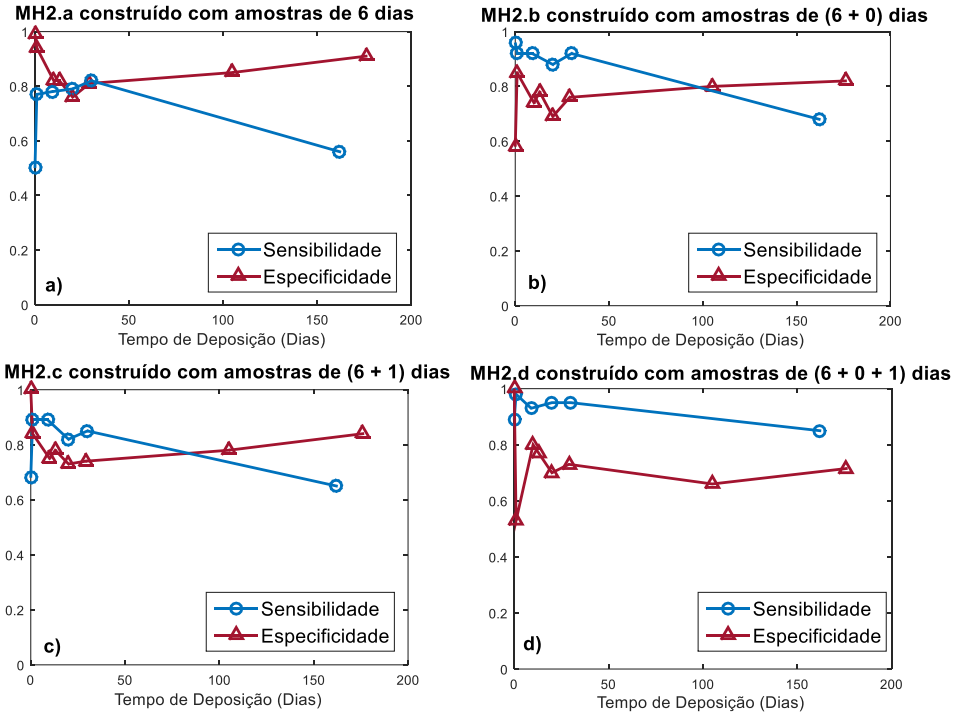
A adição de amostras frescas ao conjunto de treinamento aumentou a taxa de sensibilidade para valores acima de 0,9 para amostras depositadas entre zero e trinta dias. As amostras com 162 dias também melhoraram na classificação, porém ainda houve algumas amostras classificadas como NS. Os melhores resultados de sensibilidade foram alcançados com o modelo construído com amostras de seis e um dia que saíram de 0,65 para 0,82. A classificação correta aumentou significativamente quando amostras do dia 0 dia de depósito foram introduzidas no conjunto de treinamento, especialmente aquelas dos primeiros dias após o depósito do sangue (e 100% nas primeiras seis horas).

#### 3.3.2.2 Modelo Hierárquico 2

O segundo modelo hierárquico deste estudo incorporou um SIMCA tradicional construído modelando apenas uma única classe (Sangue Humano), como segunda regra de decisão. Diferentemente da estratégia do primeiro modelo hierárquico que utilizou um modelo de PCA, o método SIMCA considera tanto os valores do Q-Residual quanto o $T^2$ de Hotelling com nível de confiança de 95% para calcular as distâncias e os limites da classe. Neste caso a especificidade do modelo MH2 aumentou, conforme mostra a Figura 12, pode-se observar que para amostras com zero dias de depósito, o modelo original classificou corretamente apenas 50% das amostras de SH. O modelo apresentou dificuldade em diferenciar amostras frescas de SH das amostras de SA, no entanto, apenas

4% das amostras frescas de SH foram classificadas como falso negativo. Ou seja, o modelo reconhece as amostras de sangue humano, como sangue, mas apresenta dificuldade para diferenciá-las do sangue animal.

**Figura 12 -** Resultados de Sensibilidade e Especificidade do Modelo Hierárquico 2 (MH2) construído com a) amostras de 6 dias; b) amostras de 6 + 0 dias; c) amostras de 6 + 1 dias e d) amostras de 6 + 0 +1 dias.




**Fonte:** A autora (2024).

Quando amostras de sangue com zero dias foram adicionadas ao modelo (Figura 12b), a sensibilidade para amostras de sangue humano aumentou de 0,50 para 0,96 para amostras de dia 0 e de 0,77 para 0,92 para amostras de um dia. O baixo valor de especificidade de 0,58 obtido para amostras de dia 0 ocorreu porque algumas amostras frescas de SA foram classificadas como SH, indicando que amostras frescas de SA são mais facilmente confundidas com SH.

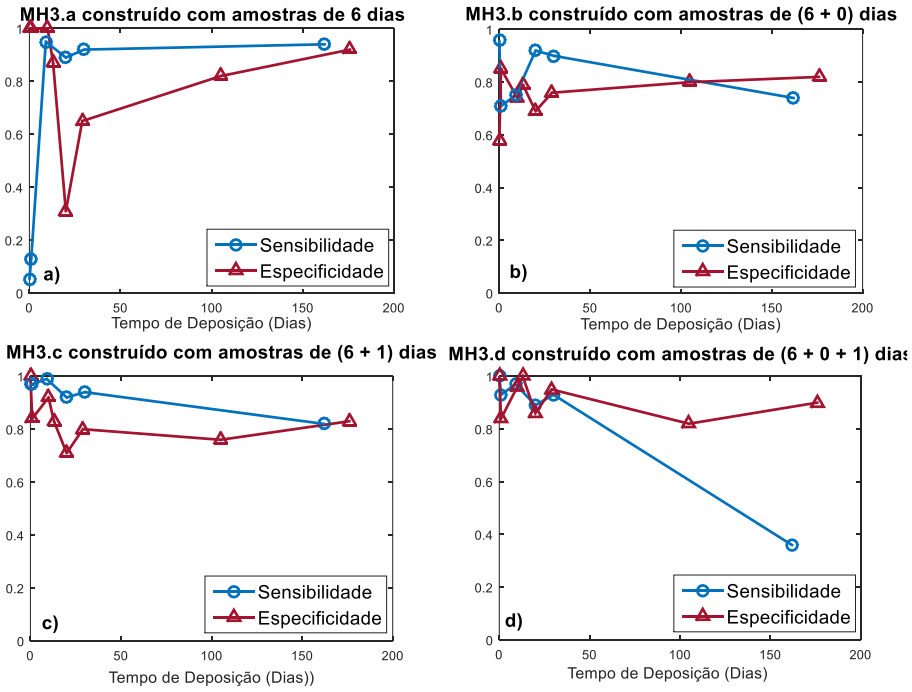
A Figura 12c indica que quando amostras de um dia de depósito foram adicionadas ao conjunto de treinamento, houve também uma melhoria nos resultados de sensibilidade, especialmente para amostras de um dia, nas quais sensibilidade aumentou para 0,89. A adição de amostras de zero e um dia ao conjunto de treinamento resultou em uma melhoria nos valores de sensibilidade (Figura 12d). Além disso, ao considerar amostras mais antigas com tempo de deposição de 162 dias, os melhores resultados de

sensibilidade (0,85) foram alcançados com o modelo MH2.d. Isto indica que a incorporação destes intervalos de tempo específicos no conjunto de treinamento também teve um impacto positivo na capacidade do modelo de detectar amostras de sangue humano mais antigas com exatidão. Porém, houve uma diminuição na especificidade, principalmente para as amostras de um dia, que tiveram um valor de especificidade de 0,53 neste modelo. Esta observação reforça a preocupação quanto à classificação incorreta de amostras recentes de SA como amostras de SH.

### 3.3.2.3 Modelo Hierárquico 3

O terceiro modelo hierárquico utiliza um modelo de PLS-DA como segunda regra de decisão. O PLS-DA é uma ferramenta de análise discriminante que utiliza informações das diferenças entre classes para identificar a direção mais influente para realizar a separação das classes através da criação de uma função discriminante. Embora esta abordagem melhore a exatidão da classificação dentro do conjunto de treinamento, ela pode representar outros desafios na classificação correta de amostras com tempos de deposição diferentes daqueles incluídos no conjunto de treinamento. Por esse motivo, a especificidade do modelo aumentou, como pode ser observado na Figura 13.

**Figura 13** - Resultados de Sensibilidade e Especificidade do Modelo Hierárquico 3 (MH3) construído com a) amostras de 6 dias; b) amostras de 6 + 0 dias; c) amostras de 6 + 1 dias e d) amostras de 6 + 0 +1 dias.



**Fonte:** A autora (2024).

Na Figura 13a, os resultados de sensibilidade (0,05 e 0,13) mostram que amostras de sangue SH com zero dias e um dia de depósito não puderam ser identificadas como SH usando o modelo original. Na verdade, embora estas amostras tenham sido identificadas como sangue, a probabilidade de pertencerem à classe SH foi inferior a 50% e foram classificadas como SA. Apesar disso, este modelo foi o que apresentou maior sensibilidade para amostras de sangue humano com 162 dias: 0,93. Ao adicionar as amostras do dia 0 ao conjunto de treinamento houve uma melhoria nos valores de sensibilidade para as amostras dos dias zero e um, conforme mostrado na Figura 13b. A sensibilidade para amostras do dia zero foi igual a 1,0 e para amostras com um dia foi de 0,71.

Na Figura 13c, a inclusão de amostras de sangue com um dia de depósito no conjunto de treinamento (MH3.c) resultou em melhorias significativas na sensibilidade do modelo. Valores de sensibilidade superiores a 0,92 foram alcançados para amostras com tempos de deposição de até trinta dias. Para amostras com tempo de deposição de 162 dias, a sensibilidade permaneceu elevada em 0,82, e todos os valores de especificidade ficaram acima de 0,71. Ao incluir amostras de sangue de um dia de depósito no conjunto de treinamento, houve um impacto positivo na identificação de amostras de sangue humano fresco e antigo. Essa melhoria foi maior do que a obtida quando foram incluídas amostras do dia zero, provavelmente porque estas amostras sofreram menos processos de degradação e oxidação do que as amostras mais antigas.

Na Figura 13d, quando as amostras de zero e um dia foram incluídas no conjunto de treinamento, não houve alterações significativas na maioria dos resultados de sensibilidade. Porém, comparado ao modelo MH3.c (Figura 13c), houve diminuição da sensibilidade para amostras com tempo de deposição de 162 dias. Da mesma forma, a especificidade permaneceu elevada, mas é importante notar que este modelo rejeitou todas as amostras de sangue de animais com tempo de deposição de um dia, classificando-as como NS (Não-Sangue).

Uma observação importante é o melhor desempenho do MH construído com um modelo PLS-DA para a discriminação do sangue animal em comparação com outros tipos. Isso era esperado, uma vez que a ferramenta dependia da utilização de classes sanguíneas humanas e animais para estabelecer limites distintos entre elas. No entanto, a categoria sangue animal não consistia apenas num número limitado de espectros, mas

também incluía amostras provenientes de apenas duas espécies animais. O impacto desta limitação foi menos pronunciado nos modelos que empregam a ferramenta SIMCA, principalmente porque a única classe modelada foi o sangue humano. Esta distinção torna os modelos hierárquicos construídos com SIMCA mais flexíveis do que aqueles construídos com a ferramenta PLS-DA.

## 3.4 CONCLUSÃO

Os modelos de classificação hierárquica representam uma estratégia promissora para identificar amostras de sangue em situações diversas possíveis de serem encontradas na rotina do perito criminal. Os modelos originalmente construídos com amostras de seis dias (FONSECA, 2020; FONSECA et al., 2022) funcionaram bem para identificar amostras de sangue humano com idades entre nove e trinta dias, com os modelos hierárquicos apresentando valores de sensibilidade acima de 0,77. Porém, amostras de sangue humano mais antigas, com 162 dias, foram mais difíceis de serem identificadas, com valores de sensibilidade de até 0,55. Este resultado era esperado, uma vez que amostras envelhecidas frequentemente sofrem degradação que pode introduzir interferência nas medições espectrais.

Manchas de sangue mais recentes, com zero e um dia de depósito, também apresentaram alguma dificuldade para serem identificadas pelos modelos originais. O modelo hierárquico (MH1.a) que utilizou um modelo de PCA como segunda regra de decisão foi o que apresentou melhor resultado para essas amostras, com sensibilidade de 0,82 e 0,81.

Em um contexto forense, é comum encontrar amostras de sangue mais recentes, num período de zero a trinta dias. Nesse sentido, fez-se necessário melhorar os resultados do modelo original, adicionando ao conjunto de treinamento amostras com zero e um dia. Os melhores resultados foram obtidos utilizando o modelo MH3.c, onde a segunda regra de decisão foi um modelo PLS-DA, e amostras de um dia de depósito adicionadas ao conjunto de treinamento. Para este modelo observou-se um crescimento dos valores de sensibilidade, acima de 0,92 para amostras entre zero e trinta dias de depósito, e para amostras de 162 dias de depósito, a sensibilidade foi 0,88.

Aumentar a variabilidade do conjunto de treinamento, incluindo amostras com diferentes tempos de depósito, foi essencial para melhorar o desempenho dos modelos de classificação hierárquica. Em um contexto criminal, o mais importante é evitar falhas na

identificação de uma amostra de sangue e, consequentemente, evitar a perda de um vestígio. Assim, a metodologia desenvolvida mostrou-se eficaz para a identificação de manchas de sangue depositadas em diferentes pisos e tempos de deposição.

# 4 MODELOS QUIMIOMÉTRICOS PARA ESTIMATIVA DO TEMPO DE DEPOSIÇÃO DE AMOSTRAS DE SANGUE EM DIFERENTES PISOS

## 4.1 INTRODUÇÃO

Estimar o tempo de deposição de manchas de sangue é uma informação muito importante em uma investigação criminal e é, por esse motivo, um tema recorrente em inúmeros estudos na área forense (AGUDELO et al., 2016; BREMMER et al., 2011b; CHENG et al., 2023; GARIGLIO et al., 2024; MC SHINE et al., 2017; MENŻYK et al., 2023; ZADORA; MENZYK, 2018). Isso ocorre porque essa informação pode ser essencial para a determinação da ordem dos acontecimentos de um crime violento e sua elucidação, ou ainda, para a determinação se um vestígio encontrado tem realmente ligação com o crime (WEYERMANN; RIBAUX, 2012). Muitos dos estudos focados em estimar a idade das manchas de sangue se baseiam em técnicas destrutivas ou no uso de equipamentos de bancada (AGUDELO et al., 2016; ANDERSON; HOBBS; BISHOP, 2011; BREMMER et al., 2011b; CHENG et al., 2023; DAS et al., 2020; DOTY; MURO; LEDNEV, 2017; GARIGLIO et al., 2024; KUMAR; SHARMA; SHARMA, 2020; MENŻYK et al., 2023; ZADORA; MENZYK, 2018). Consequentemente, muitas vezes, isso implica na necessidade de extrair amostras de manchas de sangue da cena do crime e transportá-las para laboratórios forenses para subsequente análise, o que pode levar à contaminação, perda de amostra e lentidão na resposta.

Além destes desafios, um estudo recente realizado por Menzyk e colaboradores (2023) destaca a dificuldade de desenvolver um método de datação absoluta para estimar o tempo desde a deposição de manchas de sangue em cenas de crimes reais. Isto se deve à variedade de condições ambientais às quais as manchas de sangue podem ser expostas, o que pode interferir na cinética do processo de degradação dos diferentes componentes do sangue (MENŻYK et al., 2023). Nesse sentido, os métodos analíticos desenvolvidos em ambientes laboratoriais sob condições controladas não reproduziriam com exatidão as condições do mundo real, as quais às manchas de sangue estão sujeitas e uma abordagem caso-específica seria mais apropriada (MENŻYK et al., 2023).

Naturalmente, o cenário ideal envolve a identificação de tais vestígios diretamente na cena do crime, idealmente utilizando um método analítico específico para as condições a que as manchas foram submetidas. Para enfrentar esse desafio, a utilização de equipamentos portáteis operando na região do infravermelho próximo (NIR) pode ser uma alternativa promissora para a identificação das manchas de sangue de maneira rápida

e *in situ* (FONSECA et al., 2022; MORILLAS; GOOCH; FRASCIONE, 2018; PEREIRA et al., 2017).

Como já mencionado, em um estudo recente conduzido por nosso grupo de pesquisa (FONSECA et al., 2022), foi explorado o potencial de um equipamento NIR portátil abrangendo a faixa de comprimento de onda de 900 nm a 1650 nm. O objetivo era diferenciar efetivamente vários tipos de manchas: sangue humano, sangue animal e substâncias que se assemelham ao sangue. De maneira encorajadora, essa abordagem produziu resultados bem-sucedidos para distinguir as diferentes substâncias depositadas em nove pisos distintos.

Um estudo desenvolvido por Dos Santos (2020) utilizou dois equipamentos portáteis, um de espectroscopia no infravermelho próximo (na faixa espectral de 900 nm a 1650 nm) e um de espectroscopia Raman. Foram avaliados os efeitos da temperatura (de 9°C a 43°C) nos espectros obtidos ao longo do tempo (0 a 49 dias), e as manchas de sangue foram depositadas em dois tipos de pisos (porcelanato e cerâmica). Os modelos de regressão foram construídos utilizando PLS e Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machine*). O autor concluiu que o equipamento NIR utilizado teve uma boa performance para avaliar qualitativamente a discriminação entre manchas recentes e antigas. Os modelos PLS construídos usando espectros NIR obtiveram valores de RMSECV de 33,6 horas para amostras de 24h a 15 dias, mas o autor não informou os valores de RMSEP desse modelo. O melhor modelo construído com SVM apresentou RMSEP de 65,6 horas. Percebe-se, portanto, que esse equipamento portátil que opera na região entre 900 nm a 1650 nm não apresentou um desempenho tão satisfatório quando aplicado à estimativa do tempo decorrido desde a deposição de manchas de sangue depositadas nos mesmos pisos (DOS SANTOS, 2020). Uma razão atribuída foi a faixa espectral reduzida do equipamento utilizado.

Dessa forma, o objetivo principal do estudo discutido neste capítulo é avaliar o desempenho de um outro equipamento portátil cuja faixa espectral opera em outra região do infravermelho próximo, entre 1350 nm a 2500 nm. Essa região apresenta bandas observáveis que correspondem às bandas de combinação de grupos hidroxila (-OH) em torno de 1950 nm, bem como bandas de combinação de grupos amida (~2000 - 2100 nm) que podem estar relacionadas a substâncias presentes no sangue, como proteínas, albumina, globulina e hemoglobina. A hipótese aqui levantada é que essa região espectral pode melhorar a qualidade das informações referentes ao envelhecimento do sangue em

comparação ao estudo anterior, uma vez que essa região apresenta maior variabilidade nas bandas relacionadas a demais componentes presentes no sangue.

Considerando os desafios associados às diversas condições em que as manchas de sangue podem ser encontradas, duas estratégias diferentes foram investigadas: (i) Global - um modelo hierárquico foi construído incorporando tanto a Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA) quanto a Regressão de Mínimos Quadrados Parciais (PLSR) usando espectros coletados de manchas de sangue de diferentes doadores depositadas em nove tipos diferentes de piso; (ii) Caso-específica - um outro modelo foi desenvolvido a partir de espectros de manchas de sangue coletadas de diferentes doadores depositados em um único tipo de piso, com o objetivo de prever o tempo decorrido desde a deposição de uma mancha de sangue de um doador desconhecido depositada em um suposto piso encontrado na cena do crime (abordagem caso-específica).

### 4.1.1 Objetivos

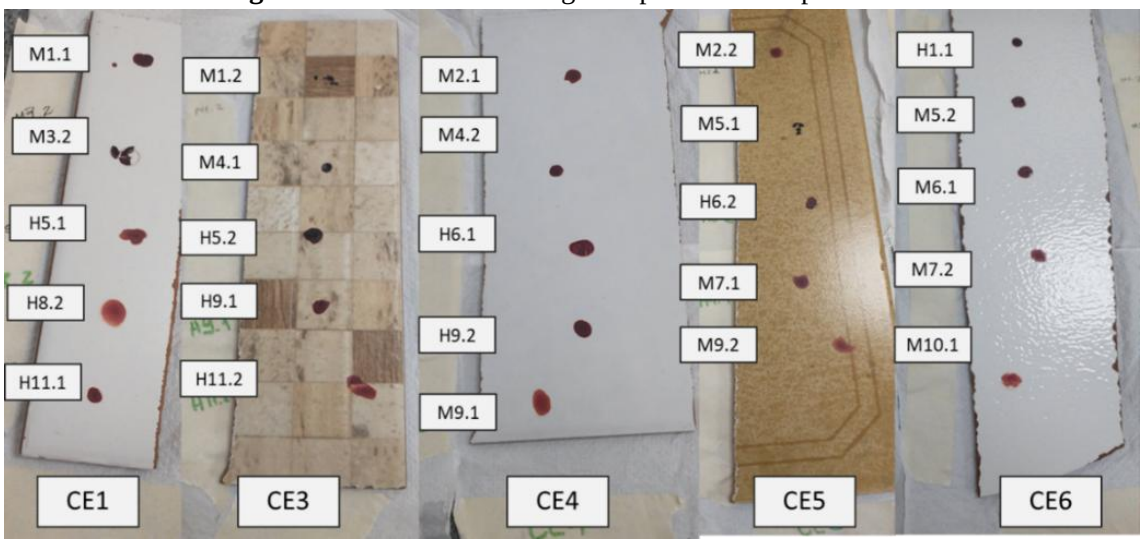
O presente estudo teve o objetivo de avaliar uma nova estratégia para estimar o tempo de deposição de manchas de sangue depositadas em nove diferentes substratos (cinco cerâmicas e quatro porcelanatos) usando espectros coletados através de um equipamento portátil operando na região do infravermelho próximo (cobrindo a faixa espectral de 1350 e 2500 nm). Para isso, os seguintes objetivos específicos foram determinados:

- Avaliar a capacidade de equipamento portátil, operando no infravermelho próximo, associado ao uso de ferramentas quimiométricas para identificar tendências de separação das amostras ao longo do tempo de deposição;
- Avaliar o potencial de um modelo hierárquico incorporando as ferramentas de Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e Regressão por Mínimos Quadrados Parciais (PLSR) para facilitar a segregação das amostras de acordo com intervalos distintos de tempo de deposição, e a estimativa da idade das manchas de sangue;
- Avaliar o desempenho de uma abordagem caso-específica, em que um modelo é construído adequando às condições de um vestígio encontrado em uma suposta cena de crime.

## 4.2 MATERIAIS E MÉTODOS

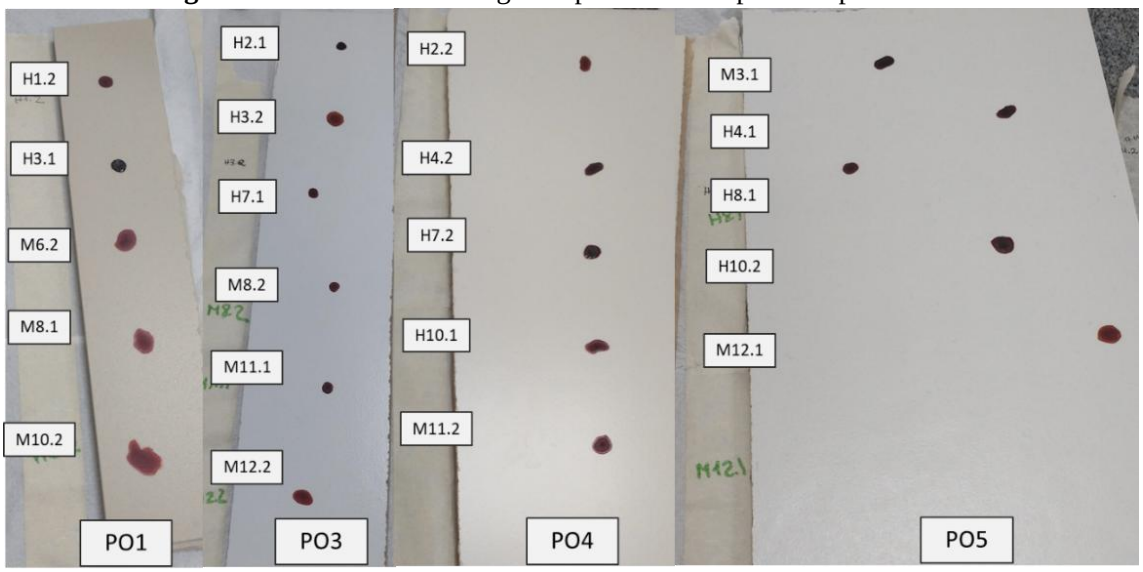
Um colaborador qualificado utilizou lancetas descartáveis para obter amostras de sangue por punção da falange distal do dedo anelar direito de cada um dos 23 voluntários (12 homens e 11 mulheres). O volume de cada mancha de sangue não foi padronizado, e cada doador forneceu duas manchas de sangue, totalizando 46 manchas, que foram depositadas na superfície de nove diferentes substratos (pisos), sendo cinco tipos de cerâmica e quatro tipos de porcelanato, conforme pode ser observado nas Figuras 14 e 15, respectivamente.

**Figura 14 -** Manchas de sangue depositadas em pisos de cerâmica.




**Fonte:** A autora (2024).

**Figura 15 -** Manchas de sangue depositadas em pisos de porcelanato.




**Fonte:** A autora (2024).

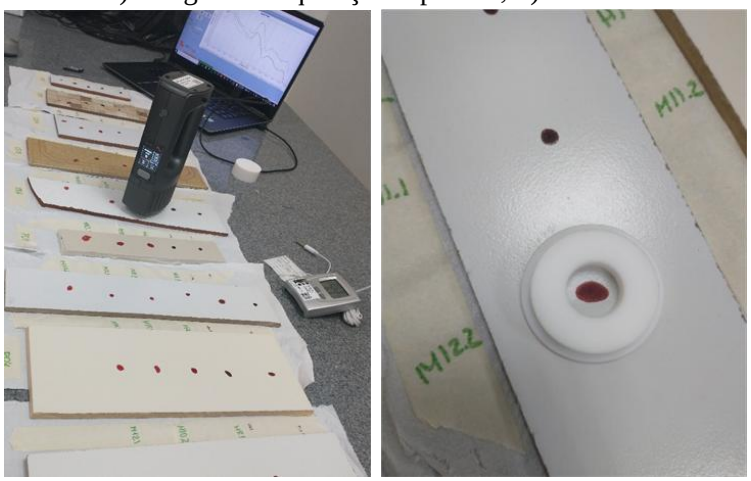
Todos os voluntários assinaram o Termo de Livre Consentimento Esclarecido (TCLE) – disponível no Apêndice B – que foi aprovado pelo Comitê de Ética em Pesquisa envolvendo Seres Humanos da Universidade Federal de Pernambuco (CEP-UFPE), vinculado ao Projeto cujo Certificado de Apresentação de Apreciação Ética (CAAE) é 50883721.2.0000.5208.

### 4.2.1 Aquisição dos Espectros

Os espectros de infravermelho próximo foram adquiridos usando o espectrômetro portátil NeoScanner desenvolvido pela NeoSpectra (Si-Ware) cuja faixa espectral abrange a região de 1350 a 2500 nm, com resolução de 16 nm. O equipamento pesa aproximadamente 1 kg e tem dimensões de 17,8 cm de comprimento, 9,1 cm de largura e 6,2 cm de altura. O diâmetro de abertura do feixe é aproximadamente 1 cm e o tempo de varredura foi de 2 segundos. O equipamento foi conectado a um computador por meio de um cabo USB e os espectros de absorbância foram adquiridos por meio do software fornecido pelo fabricante.

Para reduzir a interferência do substrato e minimizar a dispersão da radiação, um acessório de anel de Teflon (Figura 16b) com uma abertura de 1,5 cm de diâmetro foi anexado à saída do feixe. Este acessório também foi usado para centralizar o feixe na mancha de sangue para aquisição espectral ideal. O material Teflon do acessório foi escolhido por sua baixa absorção na faixa espectral do infravermelho próximo e contribuição mínima para o sinal espectral geral. A Figura 16a mostra um exemplo representativo da aquisição espectral.

**Figura 16 -** a) Imagem da aquisição espectral; b) Acessório de Teflon.




**Fonte:** A autora (2024).

Durante o primeiro dia de acompanhamento das amostras de manchas de sangue, os espectros foram coletados após um intervalo de 30 minutos da deposição, e em seguida, foram adquiridos a cada hora durante 10 horas, buscando avaliar o processo de secagem do sangue. Após esse período inicial de análise, os espectros foram coletados em intervalos de aproximadamente 24 horas durante 15 dias para avaliar as mudanças espectroscópicas ao longo do tempo. Todos os espectros de amostra foram adquiridos em triplicata, totalizando 138 espectros em cada intervalo e resultando em um conjunto total de 3.588 espectros.

Para simular as condições comumente encontradas em uma cena de crime, a temperatura e a umidade da sala foram deixadas sem controle, mas monitoradas diariamente. A temperatura média da sala foi de aproximadamente 25,7°C (±2°C), e a umidade foi, aproximadamente, 55% (±10%).

### 4.2.2 Análise de Dados

Para realizar a análise quimiométrica dos dados espectrais, foi utilizado o software Matlab (MATLAB®, MathWorks) com o PLS Toolbox versão 8.6.2 (Eigenvector Research, Inc). Os espectros médios foram calculados pela média do perfil espectral de triplicatas de cada amostra, e diferentes ferramentas de pré-processamento foram avaliadas para atenuar efeitos espectrais indesejados, como ruído e espalhamento. Essas ferramentas incluíram o filtro de suavização de Savitzky-Golay (usando janelas com 3, 5 e 7 pontos), Variação Normal Padrão (SNV, do inglês *Standard Normal Variate*), 1ª Derivada de Savitzky-Golay (usando janelas com 3, 5 e 7 pontos) e Centragem na Média (MC, do inglês *Mean Centering*).

A Análise de Componentes Principais (PCA) foi realizada para avaliar se havia mudança no comportamento dos espectros para os diferentes intervalos de tempo. Baseado na análise exploratória inicial, o conjunto de dados foi dividido em dois subconjuntos de acordo como os intervalo de tempo: um subconjunto composto por amostras com tempo de deposição variando de 0,5 h a 10 h (amostras frescas) e o outro subconjunto composto por amostras com tempo de deposição entre 24 h e 360 h (1 dia a 15 dias). Para realizar essa separação dos dados em dois conjuntos, foi construído um modelo de classificação usando PLS-DA. A PLSR foi utilizada para construir os modelos de regressão. Além disso, o filtro de Mínimos Quadrados Ponderados Generalizados

(GLSW) foi aplicado como uma etapa de pré-processamento para reduzir os efeitos do substrato (ou seja, pisos) nos dados espectrais.

#### 4.2.2.1 Abordagem Global

A primeira abordagem teve como objetivo criar um modelo único capaz de ser utilizado para estimar o tempo desde a deposição de manchas de sangue em diferentes pisos. Como o conjunto de dados foi dividido em dois subconjuntos, um modelo hierárquico foi construído usando como primeira regra de decisão um modelo de Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA) para classificar as amostras em amostras "frescas" e amostras "antigas". Nesta etapa prévia de classificação, as amostras são separadas e previstas em dois modelos de regressão diferentes, dependendo se são consideradas "frescas" ou "antigas". Amostras classificadas como antigas foram submetidas a um modelo de regressão construído com amostras com tempo de deposição de 24 h a 360 h (1 dia a 15 dias). Amostras classificadas como frescas, são submetidas a um modelo de regressão construído com amostras com tempos de deposição entre 2 h e 10 h.

Como já discutido nos Capítulo 2 e 3 desta tese, os modelos hierárquicos são uma estratégia utilizada para resolver problemas que envolvem mais de uma ferramenta quimiométrica e podem ser combinados em formato de árvore de decisão. Foram avaliados quatro conjuntos de treinamento formados pelos espectros médios das manchas de sangue depositadas em pisos diferentes das manchas de sangue depositadas em quatro conjuntos de predição. Cada conjunto de treinamento foi formado por um total de 952 espectros médios, enquanto cada conjunto de predição foram construídos com 264 espectros médios.

#### 4.2.2.2 Abordagem Caso-Específica

A segunda abordagem foi desenvolvida considerando um cenário em que o perito criminal identifica uma mancha de sangue de um indivíduo desconhecido depositada em um piso encontrado em um suposto local de crime. Os modelos de regressão foram desenvolvidos com espectros de manchas de sangue de quatro doadores diferentes depositadas no mesmo piso, estimando-se o tempo decorrido desde a deposição de uma mancha de sangue de um doador externo. O objetivo foi construir modelos de regressão

com manchas de sangue submetidas às mesmas condições ambientais da mancha de sangue encontrada na suposta cena de crime.

Foram utilizados os mesmos passos quimiométricos da primeira abordagem: filtro de suavização Savitzky-Golay (usando uma janela de 7 pontos), SNV, GLSW e centragem na média foram aplicados como métodos de pré-processamento. De modo semelhante, foram construídos os modelos PCA, PLS-DA e PLSR.

O conjunto de treinamento foi constituído pelas triplicatas dos espectros das manchas de sangue de quatro doadores que foram depositadas em um mesmo piso, totalizando 396 espectros, enquanto o conjunto de predição foi formado pelas triplicatas dos espectros de mancha de sangue de um quinto doador (externo ao conjunto de treinamento, mas depositados no mesmo piso), totalizando 81 espectros. Para realizar a construção dos modelos dessa abordagem optou-se por usar as triplicatas dos espectros ao invés dos espectros médios buscando aumentar a variabilidade das amostras.

## 4.3 RESULTADOS E DISCUSSÃO

### 4.3.1 Abordagem Global

#### 4.3.1.1 Análise Exploratória

O primeiro passo antes de construir os modelos de regressão foi avaliar os espectros e buscar identificar efeitos indesejados na aquisição. Como os espectros foram coletados *in situ*, diretamente das manchas de sangue depositadas nas placas de piso, foram observados alguns efeitos de espalhamento da radiação e ruído espectral devido à superfície do substrato e à natureza da medida espectral (reflectância).

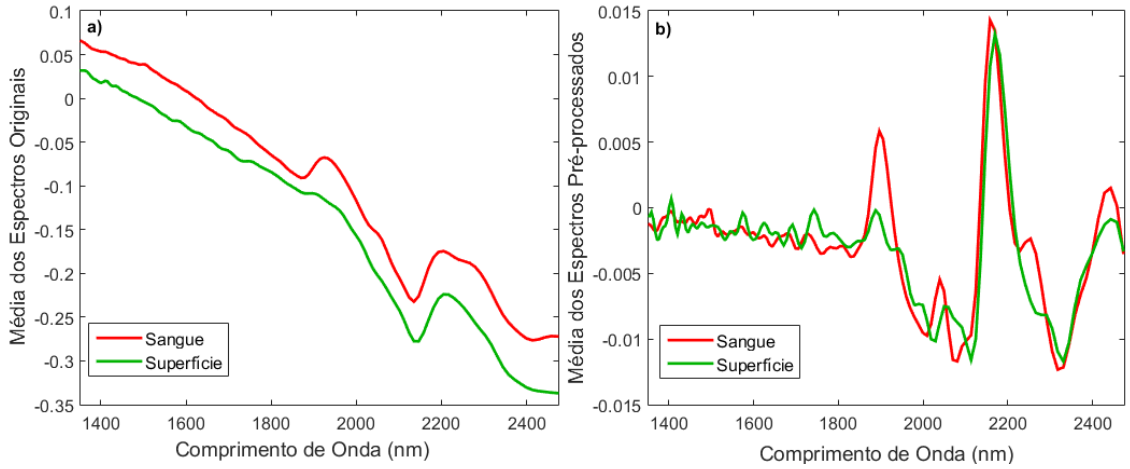
A Figura 17a mostra uma comparação entre os espectros médios dos diferentes pisos e os espectros médios de todas as amostras de sangue. A semelhança do formato entre os dois espectros destaca a influência significativa da superfície dos pisos e destaca a necessidade de considerar os efeitos do substrato nos dados espectrais. Essa observação está em consonância com um estudo anterior realizado por nosso grupo de pesquisa que avaliou a influência de diferentes substratos no comportamento espectral de manchas de sangue humano, animal e falsos positivos comuns (FONSECA et al., 2022).

Na Figura 17b, os espectros médios pré-processados com 1ª Deriva de Savitzky-Golay (5 pontos e polinômio de 2ª ordem) podem ser observados, revelando diferenças espectrais mais claras - especificamente, a banda em torno de 1940 nm, que pode estar

associada à banda de combinação de vibrações de estiramento simétricas e assimétricas das ligações hidroxila ($-OH$) presente nas moléculas de água.

Algumas regiões do espectro também podem estar associadas à presença de moléculas de hemoglobina e outras proteínas presentes no sangue: entre 2000 nm e 2100 nm encontra-se uma região que pode estar associada às bandas de combinação de amida A e amida II presentes em substâncias encontradas no sangue como proteínas, hemoglobina, albumina e globulina. A região em torno de 2290 nm também pode estar relacionada com essas macromoléculas devido à banda de combinações de vibrações de estiramento e deformação da ligação $-C-H$ em grupos metil ($-CH_3$) (EDELMAN et al., 2012).

**Figura 17**- a) Média dos espectros originais; b) Médio dos espectros pré-processados (1ª Derivada, 5 pontos e polinômio do 2º grau).




**Fonte:** A autora (2024).

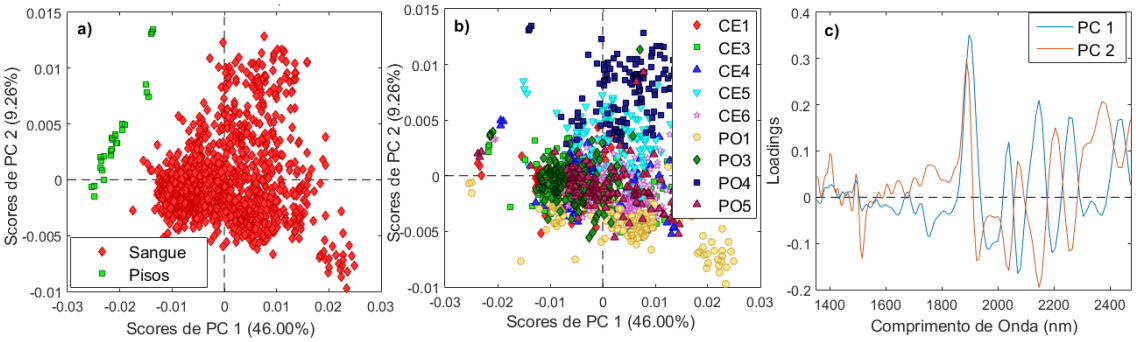
A Figura 18 mostra o modelo de PCA construído com os espectros dos pisos limpos e das manchas de sangue. No gráfico de *scores*, Figura 18a, a primeira componente principal (PC1) explica 46,08% da variância total e separa claramente os espectros das amostras de sangue (em vermelho) dos espectros dos diferentes pisos (em verde).

Os valores de *scores* positivos para as amostras de sangue em PC1 indicam que elas têm características espectrais distintas em comparação com os pisos, enquanto os valores de coordenadas mais negativos para todos os pisos nessa componente sugerem que eles têm características espectrais mais parecidas. No entanto, apesar da distinção entre as manchas de sangue e os pisos, os espectros dos pisos ainda têm um impacto considerável nos espectros das manchas de sangue, o que é evidente nos *scores* de PC2, que podem ser observados na Figura 18b (que mostra o mesmo gráfico de PC1xPC2, mas

colorido de acordo com os tipos de pisos): os *scores* das manchas de sangue têm coordenadas em PC2 semelhantes aos respectivos pisos nos quais foram depositados. Esse fato sugere que os pisos têm um efeito significativo nos espectros de manchas de sangue.

No gráfico de *loadings* (Figura 18c), os espectros indicam que a diferença entre as amostras de sangue e dos pisos em PC1 está correlacionada com a região em torno de 1940 nm, associada à presença de hidroxila ($-OH$) nas amostras de sangue, pois obtiveram *loadings* positivos e as bandas em torno de 2056 nm, 2170 nm relacionadas à presença de hemoglobina, globulina e albumina.

**Figura 18 -** Modelo de Análise de Componentes Principais construído com todas as amostras de sangue e de pisos (1ª Derivada de Savitzky-Golay - janela de 5 pontos e polinômio do 2° grau - e centragem na média): a) *Scores* de PC1xPC2 coloridos de acordo com o tipo de amostra; b) *Scores* de PC1xPC2 colorido de acordo com o tipo de piso; c) *Loadings* de PC1e PC2.

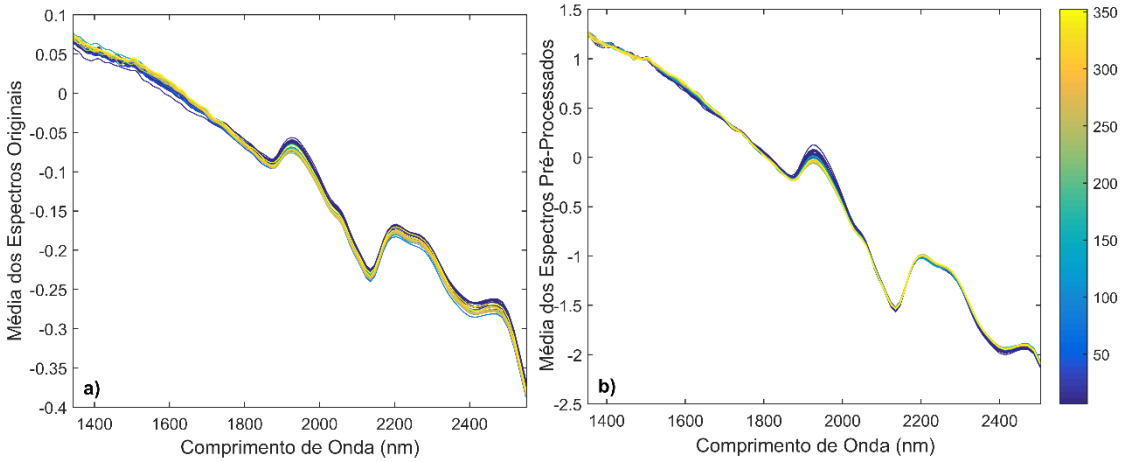



**Fonte:** A autora (2024).

Depois de analisar a influência dos pisos nas amostras de manchas de sangue, explorou-se a influência do tempo de deposição nos espectros. Na Figura 19 observam-se os gráficos dos espectros originais sem pré-processamento (Figura 19a) e os espectros pré-processados com filtro de suavização Savitzky-Golay com uma janela de 5 pontos e polinômio de 2ª ordem e SNV (Figura 19b). Os espectros estão coloridos de acordo com um gradiente de tempo de deposição variando de 0,5 h (azul) a 360 h (amarelo).

Na Figura 19b é visível que os espectros das amostras de sangue mais fresco (azul) apresentam uma maior intensidade na região em torno de 1940 nm, o que já era esperado devido à presença de água. Percebe-se que essa banda diminui ao longo do tempo devido ao processo de evaporação da água e secagem das manchas de sangue.

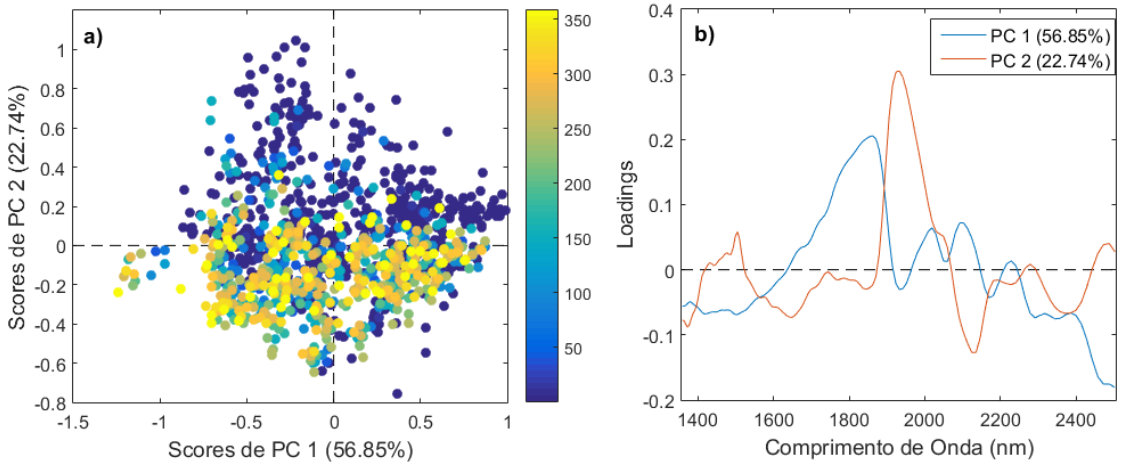
**Figura 19 -** Gráfico da Absorbância. a) Média dos espectros de sangue originais; b) Média dos espectros de sangue pré-processados (Suavização de Savitzky-Golay com janela de 5 pontos e polinômio de 2°grau e SNV).




**Fonte:** A autora (2024).

A Figura 20 mostra o modelo de PCA construído com os espectros das manchas de sangue envelhecidas em diferentes tempos de deposição, o mesmo gradiente de coloração foi usado - de 0,5 h (azul) a 360 h (amarelo) -, e é possível observar que os *scores* das amostras mais antigas possuem coordenadas mais negativas em PC2. Nos loadings (Figura 20b), a região positiva está mais relacionada à banda em torno de 1940 nm e uma banda menor em torno de 2056 nm, sendo que a primeira está correlacionada com a presença de água e a segunda com a presença de hemoglobina e outras proteínas, de modo que é possível constatar uma tendência de separação entre as amostras recentes, em azul, e as amostras mais antigas, em amarelo.

**Figura 20** - Modelo de PCA construído com todas as amostras de sangue: a) *Scores* de PC1xPC2; b) *Loadings* de PC1 e PC2.



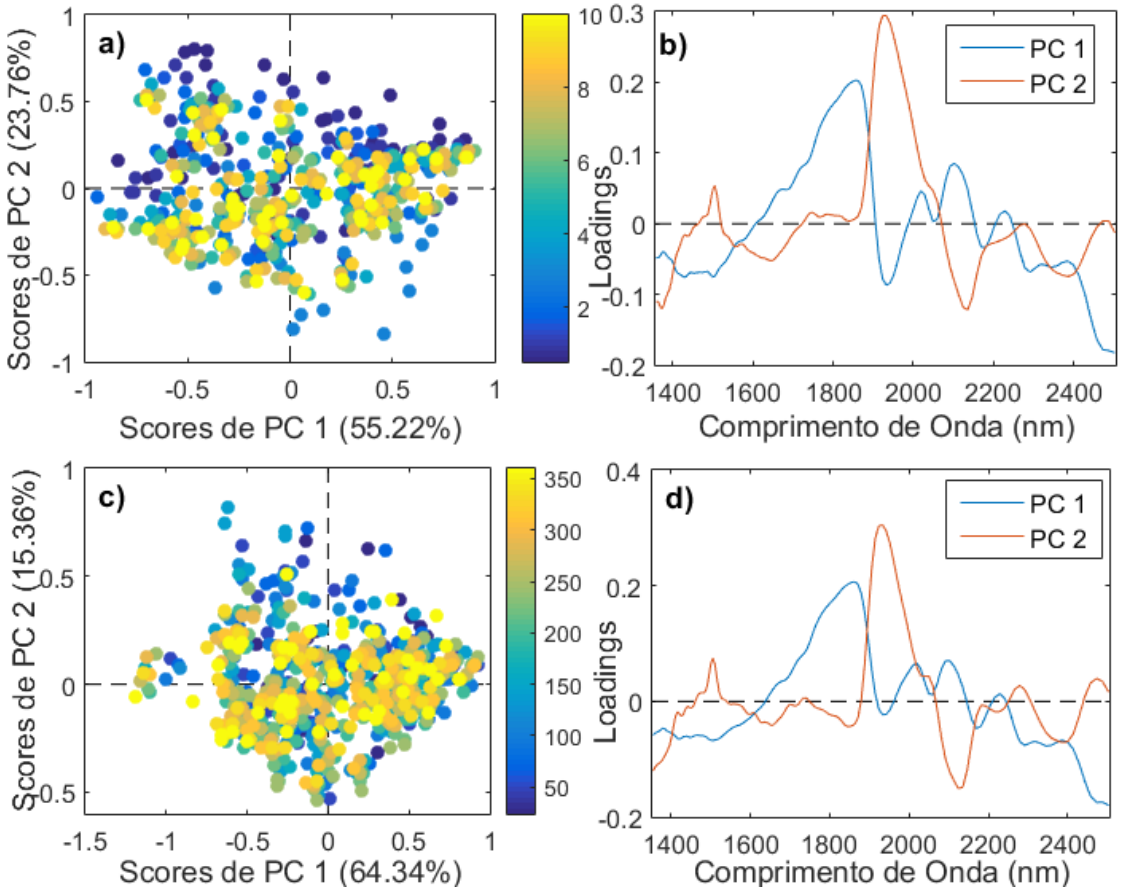

**Fonte:** A autora (2024).

Durante o envelhecimento do sangue, ocorre o processo de degradação da hemoglobina em desoxihemoglobina, metemoglobina e, por fim, hemicromo (Das et al., 2020) e tanto o processo de evaporação da água quanto o processo de oxidação da hemoglobina são mais rápidos nas primeiras horas que o sangue sai do corpo humano (SOBAC; BRUTIN, 2014). Como a dinâmica do envelhecimento do sangue é diferente nas primeiras horas de deposição (BREMMER et al., 2011a, 2011b, 2012), os espectros foram separados em dois subconjuntos: o primeiro deles, com amostras mais recentes, depositadas entre 0,5 h e 10 h, e o segundo com amostras mais antigas, com tempo de deposição entre 24h e 360h (15 dias).

Na Figura 21 observam-se o modelo de PCA construído com amostras mais recentes (Figura 21a e 21b) e o modelo de PCA construído com amostras mais antigas (Figura 21c e 21d). A tendência de separação entre amostras recentes e antigas, observada na Figura 20, não é mais encontrada nesses modelos, mas ambos os modelos têm comportamentos semelhantes. O gráfico de *scores* das duas primeiras componentes, assim como das demais componentes não mostram nenhuma tendência de separação relacionada ao processo de envelhecimento.

Uma possível razão se deve à menor taxa cinética da reação de degradação da hemoglobina após as primeiras horas de deposição (BREMMER et al., 2011a; TSURUGA et al., 1998). No entanto, separar os espectros em dois subconjuntos é importante não apenas pelos comportamentos cinéticos de degradação da hemoglobina, mas também porque os intervalos de tempo em que os espectros foram coletados em ambos os conjuntos são bastante diferentes: o primeiro teve um intervalo de 1 h, enquanto a segunda teve intervalo de 24 horas e o intervalo influenciará no erro estimado para a construção dos modelos de regressão.

**Figura 21 -** Modelo de PCA a) *Scores* de PC1XPC2 e b) *Loadings* de PC1xPC2 das amostras de sangue com tempo de deposição entre 0.5 h e 10 h; c) *Scores* de PC1xPC2 e d) *Loadings* de PC1xPC2 das amostras de sangue com tempo de deposição entre 24 h e 360 h;




**Fonte:** A autora (2024).

É importante esclarecer que a dinâmica do processo de secagem de uma gota de sangue é muito complexa. Estudos anteriores mostraram que a evolução morfológica da secagem de uma gota de sangue se comporta de maneira semelhante à de uma suspensão coloidal: à medida que a água evapora, o fluido começa a assumir uma consistência de gel que influenciará o processo de evaporação e a formação de rachaduras na superfície da mancha de sangue (ATTINGER et al., 2013; SOBAC; BRUTIN, 2011).

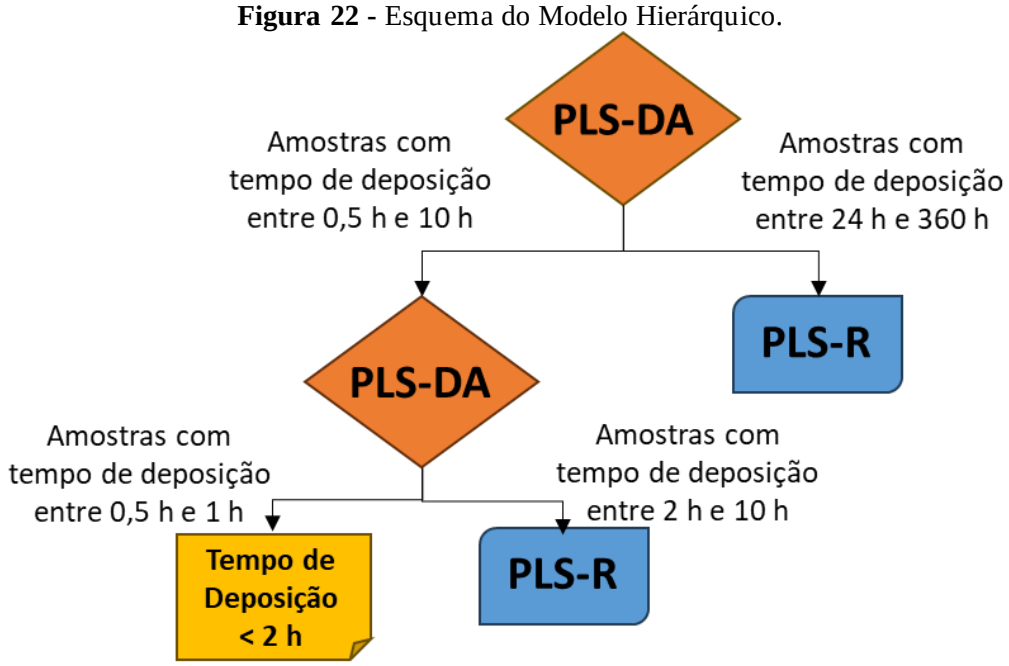
### 4.3.1.2 Classificação das Amostras e Construção do Modelo Hierárquico

Como já detalhado, durante a análise exploratória, foi observada a necessidade de separar as amostras mais recentes e as amostras mais antigas antes do desenvolvimento dos respectivos modelos de regressão. Para combinar os modelos de classificação e os modelos de regressão, a estratégia utilizada para desenvolver um modelo único, foi a construção de um modelo hierárquico.

Inicialmente foi construído um modelo de Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA) que englobou duas classes: amostras recentes (tempo de deposição de 0,5 h e 10 h) e amostras envelhecidas (tempo de deposição de 24 h a 360 h – ou 15 dias). No entanto, constatou-se que o comportamento das amostras mais frescas, com tempo de deposição entre 0,5 h e 1 h, era diferente das demais devido ao alto teor de água. Por esse motivo, foi construído um outro modelo PLS-DA para separar as amostras mais frescas, com 0,5 h e 1 h, das demais amostras (com tempo de deposição entre 2 h e 10 h). O primeiro modelo obteve um resultado de sensibilidade e especificidade acima de 0,95, enquanto o segundo modelo PLS-DA obteve um valor de sensibilidade de 0,92 e especificidade de 0,99.

Após as etapas de classificação das amostras, foram construídos dois modelos de regressão: um atendendo aos intervalos de 2 h e 10 h, e outro referente ao intervalo de 24 h e 360 h (15 dias). Nesse sentido, cada modelo PLS-DA funcionou como um critério de decisão baseado nas classes a qual cada amostra pertencia, determinando para qual modelo de regressão as amostras iriam na sequência do modelo hierárquico. Enquanto os modelos PLSR funcionaram como ponto final, fornecendo o tempo estimado de deposição para cada amostra. Um esquema referente ao modelo hierárquico implementado pode ser visto na Figura 22.

**Figura 22 -** Esquema do Modelo Hierárquico.




**Fonte:** A autora (2024).

4.3.1.3 Conjuntos de Treinamento e Predição

Como já mencionado, a dinâmica da secagem do sangue também está sujeita à influência de diversos fatores, como condições ambientais de temperatura, umidade, fluxo de ar, exposição solar, porosidade e molhabilidade do substrato, volume de sangue depositado e concentração de sangue (BRUTIN et al., 2011; SMITH; NICLOUX; BRUTIN, 2020; SOBAC; BRUTIN, 2011, 2014). Em um cenário forense, nenhum destes fatores pode ser controlado e as manchas de sangue podem assumir uma variedade de formas e padrões de formato e fissuras.

É importante ressaltar que uma das dificuldades de construir modelos com espectros coletados *in situ* é garantir que a influência da superfície (neste caso, piso) sobre a qual os espectros foram coletados seja a menor possível, ou que haja uma grande variabilidade de substratos no conjunto de treinamento para que o modelo possa ser representativo e capaz de prever a variabilidade de um espectro coletado em uma superfície diferente. Portanto, uma das etapas mais importantes na construção do modelo é a validação e a escolha do conjunto de treinamento para garantir a robustez do modelo.

Para realizar a definição dos conjuntos de treinamento e predição, utilizou-se como critério a retirada, para compor o conjunto de validação externa, de amostras depositadas em um piso de porcelanato e um piso de cerâmica, conforme a Tabela 4, preocupando-se sempre em utilizar no conjunto de predição amostras depositadas em pisos totalmente externos ao conjunto de treinamento. Essa mesma estratégia, já foi abordada em um estudo realizado por Fonseca *et al.* (2022) para identificar e diferenciar amostras de sangue humano, sangue animal e outras substâncias (falsos positivos comuns) depositadas em diferentes pisos, no qual foi enfatizada a importância da variabilidade dos substratos do conjunto de treinamento para garantir que a construção dos modelos hierárquicos seja representativa (FONSECA et al., 2022). Com base nesta perspectiva, os conjuntos de treinamento foram escolhidos de modo a garantir uma maior variabilidade dos pisos, utilizando a estratégia descrita pelo estudo mencionado.

**Tabela 4**. Resumo dos diferentes pisos usados nos diferentes conjuntos de treinamento e predição.

| N° | Conjunto de Treinamento | Conjunto de Predição |
|---|---|---|
| 1 | CE3,CE4,CE5,CE6,PO1,PO4,PO5 | CE1; PO3 |

| | | |
|---|---|---|
| 2 | CE3,CE4,CE5,CE6,PO1,PO3,PO4 | CE1; PO5 |
| 3 | CE1,CE3,CE4,CE5,PO1,PO4,PO5 | CE6; PO3 |
| 4 | CE1,CE3,CE4,CE5,PO1,PO3,PO4 | CE6; PO5 |

4.3.1.4 Modelos de Regressão

Foram construídos três modelos de regressão usando PLS para cada conjunto de treinamento seguindo diferentes métodos de pré-processamento: a) Suavização + SNV; b) Suavização + SNV + GSLW ($\alpha = 0,2$); c) Suavização + SNV + GLSW ($\alpha = 0,03$).

A ferramenta de Mínimos Quadrados Generalizados Ponderados (GLSW) funciona como um filtro para atenuar interferências através da estimativa das diferenças entre amostras de um grupo que deveriam ser semelhantes (nesse caso, amostras com o mesmo tempo de deposição). O parâmetro $\alpha$ está relacionado ao peso do efeito do filtro que será aplicado ao conjunto de dados, quanto menor o valor de $\alpha$, maior será o peso do filtro aplicado. Para averiguar os valores deste parâmetro, utilizou-se a estratégia descrita por Silva e colaboradores (2022). Naquele estudo, foram construídas superfícies de resposta para definir o parâmetro $\alpha$ para o filtro de GLSW aplicado a espectros de infravermelho médio coletados a partir de documentos antigos, com o objetivo de construir modelos de regressão para datar esses papéis. A construção destas superfícies de resposta baseou-se nos valores de $\alpha$ empregados dentro do filtro GLSW, juntamente com os valores RMSEC e RMSECV obtidos a partir de modelos de regressão PLS (SILVA G.; SILVA S.; PIMENTEL, 2022).

Os gráficos de superfície de resposta podem ser visualizados no Apêndice C. A partir dos valores obtidos através das superfícies foram escolhidos diferentes valores de $\alpha$ para avaliar o desempenho dos modelos de regressão: $\alpha = 0$; $\alpha = 0,2$ e $\alpha = 0,03$, esses valores foram escolhidos porque proporcionaram um RMSECV baixo utilizando poucas variáveis latentes, de modo a otimizar o modelo.

Inicialmente, foram construídos modelos PLS utilizando amostras com tempo de deposição entre meia hora a dez horas. Os modelos foram avaliados e constatou-se que as amostras mais frescas, com tempo de depósito de meia hora e uma hora, tiveram comportamento diferente e deslocaram a reta, causando aumento de erro e do *bias*. Como já mencionado, esse fato era esperado, devido à maior presença de água nas amostras de 0,5 h e 1 h. Dessa forma, essas amostras foram retiradas, a partir do modelo de classificação mencionado anteriormente, e novos modelos PLS foram construídos utilizando as amostras com tempo de deposição entre 2 h e 10 h. É importante ressaltar que do ponto de vista prático, um perito raramente chega ao local antes de 1 ou 2 h após

a ocorrência de um crime violento, portanto, é pouco provável que amostras com esse tempo de deposição sejam encontradas em um intervalo de tempo tão curto. A Tabela 5 apresenta as figuras de mérito de cada modelo de regressão construído com os diferentes conjuntos de treinamento/predição.

**Tabela 5 -** Resumo das figuras de mérito dos modelos PLS construídos com amostras com tempo de depósito entre 2h e 10h e os diferentes pré-processamentos.

| Conjunto de Treinamento | Intervalo de Tempo | Pré-Processamento | LV | $R^2_{CV}$ | $R^2_{Pred}$ | RMSECV (h) | RMSEP (h) | $Bias_{CV}$ (h) | $Bias_{Pred}$ (h) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2h - 10 h | SNV | 9 | 0,77 | 0,77 | 1,23 | 1,27 | -0,03 | 0,28 |
| | | SNV; GLSW (α = 0,2) | 6 | 0,76 | 0,73 | 1,27 | 1,34 | -0,03 | -0,03 |
| | | SNV; GLSW (α = 0,03) | 4 | 0,83 | 0,86 | 1,06 | 0,99 | -0,06 | 0,18 |
| 2 | 2h – 10 h | SNV | 9 | 0,76 | 0,81 | 1,26 | 1,15 | -0,02 | 0,25 |
| | | SNV; GLSW (α = 0,2) | 9 | 0,82 | 0,81 | 1,11 | 1,15 | -0,01 | 0,11 |
| | | SNV; GLSW (α = 0,03) | 5 | 0,85 | 0,83 | 1,02 | 1,06 | -0,01 | 0,14 |
| 3 | 2h - 10 h | SNV | 9 | 0,77 | 0,75 | 1,24 | 1,33 | 0,01 | -0,22 |
| | | SNV; GLSW (α = 0,2) | 6 | 0,77 | 0,74 | 1,25 | 1,34 | 0,01 | -0,21 |
| | | SNV; GLSW (α = 0,03) | 5 | 0,84 | 0,83 | 1,03 | 1,09 | 0,01 | -0,01 |
| 4 | 2h - 10 h | SNV | 9 | 0,75 | 0,82 | 1,30 | 1,13 | 0,03 | 0,12 |
| | | SNV; GLSW (α = 0,2) | 7 | 0,76 | 0,82 | 1,16 | 1,13 | 0,02 | 0,24 |
| | | SNV; GLSW (α = 0,03) | 4 | 0,81 | 0,88 | 1,13 | 0,90 | 0,02 | 0,01 |

Em relação aos diferentes tipos de pré-processamentos, pôde-se observar que houve, de modo geral, uma redução do *bias* de predição quando o filtro de GLSW foi aplicado para todos os conjuntos de treinamento. O modelo construído com o Conjunto de Treinamento 1, por exemplo, que utilizou o α = 0,2 teve um valor de *bias* de -0,035h (- 2,1 min) e para α = 0,03, o *bias* de predição foi de 0,185h (11,1 min), enquanto o *bias* do modelo que não utilizou o GLSW foi de 0,286 h (17,16 min).

Em relação aos valores de RMSEP, também foi observada uma redução desse parâmetro à medida que o filtro GLSW foi aplicado. Para o modelo construído com o CT1, sem a aplicação do filtro, o valor de RMSEP foi 1,267h (76,02 min), com α = 0,2, esse parâmetro foi igual a 1,343h (80,58 min) e para α = 0,03 foi 0,992h (59,52 min). Embora os dois primeiros valores de RMSEP não tenham apresentado diferenças estatísticas com base em um teste F a um nível de 95% de confiança, eles não foram

estatisticamente semelhantes ao último (0,992h), com base no mesmo critério. Uma diminuição no valor α (ou seja, um aumento no peso do filtro) teve um impacto maior no RMSEP, e o valor α de 0,2 teve um impacto maior na redução do *bias* de predição. Os modelos construídos com o Conjunto de Treinamento 2 (CT2) tiveram um comportamento semelhante, o *bias* de predição com α = 0,2 foi de 0,108 h (6,48 min) e para α = 0,03 foi de 0,142 h (8,52 min), e o RMSEP com α = 0,2 foi de 1,146 h (68,76 min) e para α = 0,03 foi 1,015 h (60,9 min). Do mesmo modo, ambos os valores foram comparados utilizando teste F de 95% de confiança e, novamente, não foram estatisticamente equivalentes.

Os valores de RMSEP dos modelos construídos com o Conjunto de Treinamento 4 (CT4) sem o GLSW e com o GLSW (α = 0,2) foram estatisticamente equivalentes quando comparados utilizando teste F com 95% de confiança. No entanto, diferentemente dos demais conjuntos de treinamento, o modelo construído usando CT4 e α = 0,2, teve um aumento no valor do *bias* que foi igual a 0,244h (14,64 min), esse valor foi maior do que o obtido pelo modelo que não utilizou o filtro que teve um valor de *bias* igual a 0,123h (7,38 min).

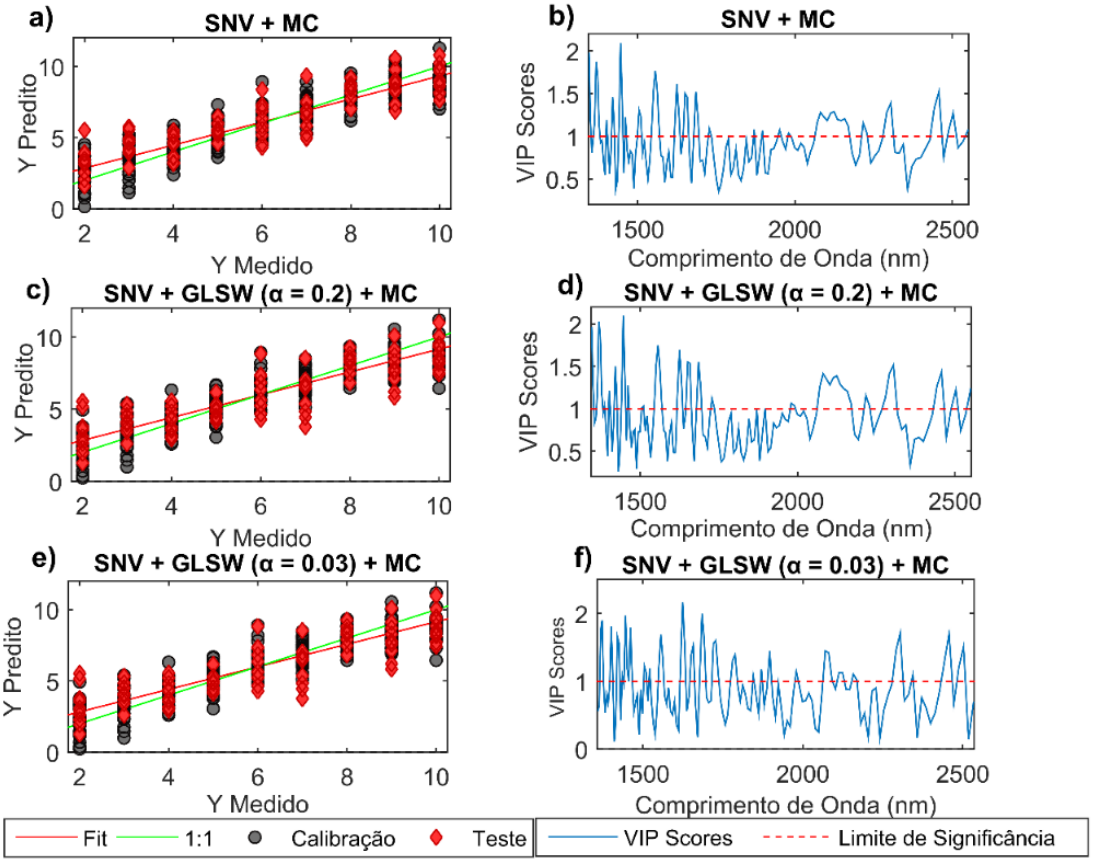
Comparando todos os modelos de regressão construídos para esse intervalo de tempo (2h e 10h), o que obteve melhor desempenho foi o que utilizou o CT4 e o filtro GLSW com α = 0,03. Esse modelo obteve um menor valor de *bias* de predição, sendo igual a 0,002 h (0,12 min) e um RMSEP de 0,906 h (54,46 min). Sendo esse valor de RMSEP comparado com os menores valores de RMSEP dos demais conjuntos de treinamento utilizando teste F a 95% de confiança e foi estatisticamente diferente de todos eles.

O principal efeito do uso do GLSW está na diminuição das diferenças ou ruídos dentro de uma mesma classe de conjunto de dados, no caso em estudo, dentro de um mesmo grupo de amostras com o mesmo intervalo de dentro, podendo diminuir, por exemplo, a diferença/ruídos relacionados aos tipos de pisos. Nesse sentido, quanto menor o valor de alfa, maior o filtro aplicado, e menor a quantidade de ruídos e interferentes incluídos nos modelos.

A Figura 23 mostra os modelos de regressão PLS construídos com CT1 (com todos os diferentes métodos de pré-processamento) e os *VIP Scores*. Visualmente, as diferenças entre os modelos são imperceptíveis e os *VIP Scores* são muito difíceis de interpretar, o que indica a dificuldade do modelo em correlacionar as variáveis aos

diferentes tempos de deposição - provavelmente devido à interferência da superfície do piso.

**Figura 23** - Modelo de regressão PLS construído com CT4 (amostras com tempo de deposição entre 2h e 10h); a) Referência versus previsto; b) Gráfico de VIP *Scores* pré-processado com Suavização, SNV e MC; c) Referência versus previsto; d) Gráfico VIP *Scores pré*-processado com Suavização, SNV, GLSW (α = 0,2) e MC; e) Referência versus previsto e f) Gráfico de VIP *Scores* pré-processado com Suavização, SNV, GLSW (α = 0,03) e MC.




**Fonte:** A autora (2024).

Em relação aos modelos construídos utilizando amostras com tempos de deposição maiores - entre 24 h (um dia) e 360 h (15 dias) - os conjuntos de treinamento tiveram desempenho semelhante ao explicado para os modelos construídos com amostras com 2 h e 10 h.

Os resultados das figuras de mérito podem ser vistos na Tabela 6. Nessa tabela, pode-se constatar o efeito do GLSW na melhoria dos valores de figura de mérito de cada modelo, todos os coeficientes de correlação ($R^2$cal, $R^2_{CV}$ e $R^2_{Pred)}$ aumentaram quando o parâmetro α diminuiu, e o GLSW também afetou os valores de *bias* e os demais parâmetros de erro do modelo.

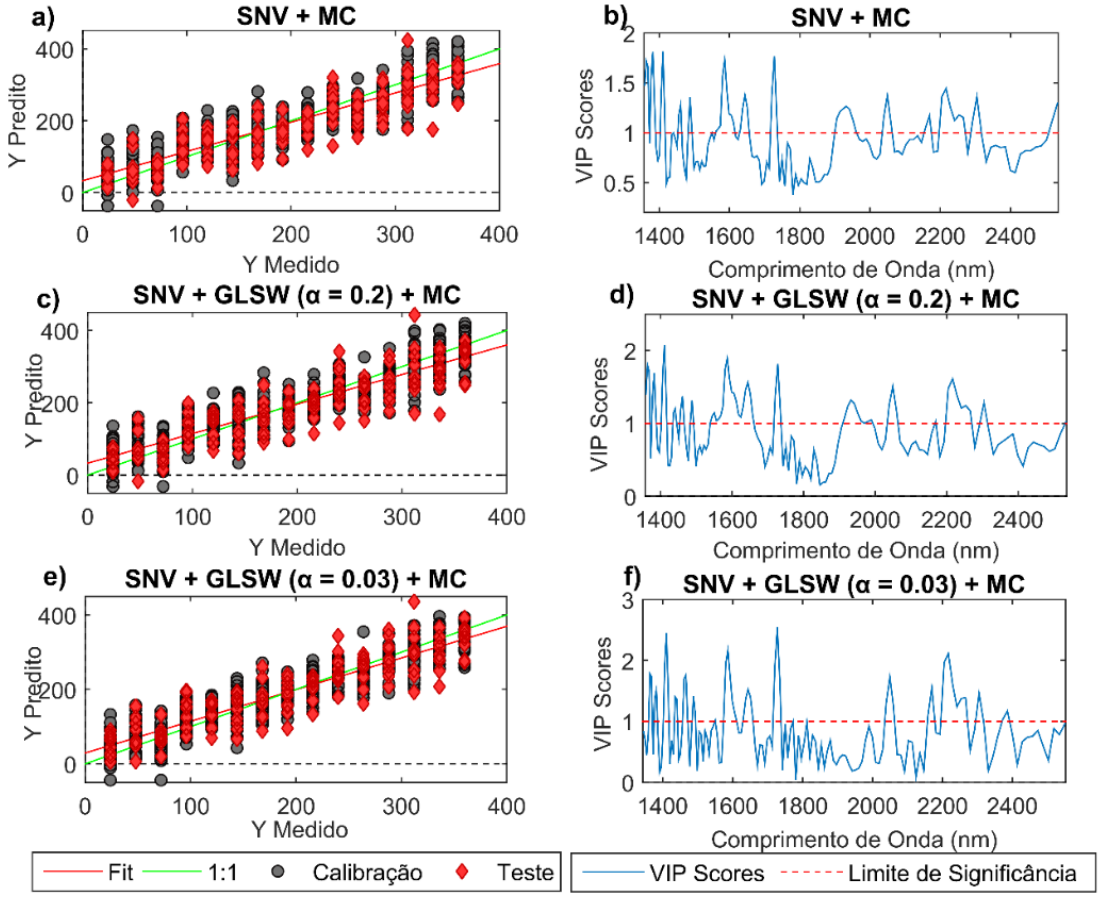
**Tabela 6 -** Resumo das figuras de mérito dos modelos PLS construídos com amostras com tempo de depósito entre 24h e 360h e os diferentes pré-processamentos.

| Conjunto de Treinamento | Intervalo de Tempo | Pré-Processamento | LV | $R^2_{CV}$ | $R^2_{Pred}$ | RMSECV (h) | RMSEP (h) | $Bias_{CV}$ (h) | $Bias_{Pred}$ (h) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 24h - 360 h | SNV | 8 | 0,73 | 0,69 | 54,28 | 58,84 | 0,15 | -12,82 |
| | | SNV; GLSW (α = 0,2) | 6 | 0,77 | 0,75 | 50,59 | 52,17 | 0,46 | -3,37 |
| | | SNV; GLSW (α = 0,03) | 6 | 0,79 | 0,79 | 47,89 | 47,36 | -0,43 | -2,85 |
| 2 | 24h - 360 h | SNV | 8 | 0,76 | 0,76 | 51,03 | 51,26 | 0,23 | 3,12 |
| | | SNV; GLSW (α = 0,2) | 6 | 0,80 | 0,78 | 46,91 | 49,43 | -0,08 | 8,92 |
| | | SNV; GLSW (α = 0,03) | 4 | 0,82 | 0,82 | 43,32 | 47,62 | -0,02 | 17,59 |
| 3 | 24h - 360 h | SNV | 8 | 0,75 | 0,76 | 52,01 | 51,53 | 0,24 | -11,75 |
| | | SNV; GLSW (α = 0,2) | 6 | 0,78 | 0,79 | 48,38 | 48,97 | 0,12 | -10,97 |
| | | SNV; GLSW (α = 0,03) | 4 | 0,81 | 0,85 | 44,77 | 39,98 | 0,11 | 2.67 |
| 4 | 24h - 360 h | SNV | 9 | 0,76 | 0,79 | 50,85 | 49,24 | -0,33 | -11,67 |
| | | SNV; GLSW (α = 0,2) | 6 | 0,77 | 0,78 | 49,42 | 50,18 | -0,38 | -9,99 |
| | | SNV; GLSW (α = 0,03) | 4 | 0,81 | 0,80 | 44,65 | 46,48 | 0,74 | 2,53 |

Os melhores modelos de regressão foram construídos com CT3 e CT4, ambos utilizaram filtro GLSW com α = 0,03). O *bias* de predição foi de 2,667 h para CT3 e 2,528 h para CT4. Apesar do valor de RMSEP obtido no modelo construído com CT3 usando GLSW e α = 0,03 ter sido menor (39,977 h) que o RMSEP obtido com o modelo construído com CT4 usando GLSW e α = 0,03 (46,484 h), quando comparados utilizando teste F (95% de confiança), eles foram estatisticamente iguais. Nesse sentido, não há diferenças estatísticas significativas que justifiquem a escolha de um conjunto em preferência do outro. Contudo, como os resultados obtidos pelo modelo de regressão usando CT4 no intervalo de tempo entre 2h e 10h foram melhores em relação aos demais, optou-se por utilizar o mesmo conjunto para construção do modelo hierárquico.

A Figura 24 apresenta os modelos de regressão PLS construídos com as amostras com tempo de depósito entre 24h e 360h do CT4. Os VIP *Scores* (Figura 24b, 24d e 24e) para esses modelos mostraram influência de diferentes regiões espectrais. As bandas em torno de 1690 nm, 1740 nm, 2056 nm e 2290 nm parecem ter maior importância para diferenciar amostras neste intervalo de tempo, essas bandas podem estar associadas à presença de hemoglobina, albumina e globulina (EDELMAN et al., 2012).

**Figura 24 -** Modelo de regressão PLS construído com CT4 (amostras com tempo de depósito entre 24h e 360h); a) Referência versus previsto e b) Gráfico de *VIP Scores* pré-processado com Suavização, SNV e CM; c) Referência versus previsto e d) Gráfico *VIP Scores* pré-processado com Suavização, SNV, GLSW (α = 0,2) e MC; e) Referência versus previsto e d) Gráfico de pontuações VIP pré-processado com Suavização, SNV, GLSW (α = 0,03).




**Fonte:** A autora (2024).

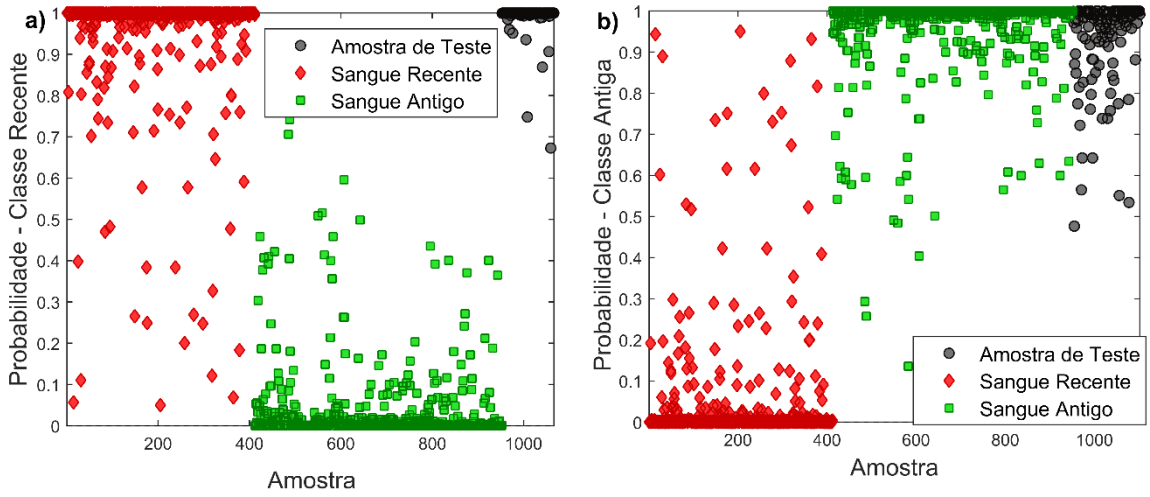
A datação de manchas de sangue é um desafio altamente complexo. Estudos cinéticos realizados sobre o processo de auto-oxidação da hemoglobina mostram que ele pode ser dividido em duas etapas de reação de primeira ordem, uma mais rápida, na qual a cadeia α é oxidada, e outra mais lenta, na qual a cadeia β é oxidada (TSURUGA et al., 1998). Porém, há divergências quanto ao intervalo de tempo que cada uma dessas etapas leva: um desses estudos afirma que a etapa rápida leva horas (BREMMER et al., 2012), enquanto outro estudo afirma que a etapa rápida pode levar alguns dias (BREMMER et al., 2011b). Considerando que a hemoglobina representa 97% do sangue (BREMMER et al., 2012), é possível que as diferentes taxas cinéticas do processo de auto-oxidação da hemoglobina possam ter influenciado o modelo de regressão neste intervalo de tempo, levando a maiores valores de erro do modelo.

#### 4.3.1.5 Resultados do Modelo Hierárquico

O Conjunto de Treinamento 4 foi utilizado para construir o Modelo Hierárquico. Este conjunto de treinamento foi escolhido por apresentar bom desempenho em ambos os modelos de regressão. Assim, foi construído um modelo de classificação utilizando PLS-DA com 3LV e as mesmas ferramentas de pré-processamento utilizadas para construir os modelos de regressão: filtro de suavização Savitzky-Golay com janela de 5 pontos e 2º ordem polinomial, SNV, GLSW ($\alpha = 0{,}03$) e Centragem na Média (CM). Este modelo apresentou valores de sensibilidade e especificidade acima de 0,95.

A Figura 25 mostra a probabilidade de predição de classe de amostras pertencentes à classe de amostra recente (Figura 25a) e à classe de amostra antiga (Figura 25b). A projeção das amostras dos conjuntos de predição pode ser vista em cinza. A Figura 25a mostra a projeção das amostras do Conjunto de Predição 4 (CP4) com tempo de deposição de 2h a 10h e todas as amostras foram classificadas corretamente. A projeção das amostras CP4 com tempo de deposição entre 24 h e 360 h pode ser vista na Figura 25b. Este conjunto de amostras parece ser mais heterogêneo entre si e algumas amostras foram classificadas incorretamente como amostras recentes.

**Figura 25** – Gráficos de Probabilidade de Predição de Classe do modelo PLS-DA com projeção das amostras PS4 (cinza) com a) tempo de deposição entre 2h e 10h (Recente); b) horário de depósito entre 24h e 360h (Antigo).




**Fonte:** A autora (2024).

Após a classificação por PLS-DA, as amostras previstas foram projetadas nos respectivos modelos de regressão de acordo com a classe em que foram classificadas.

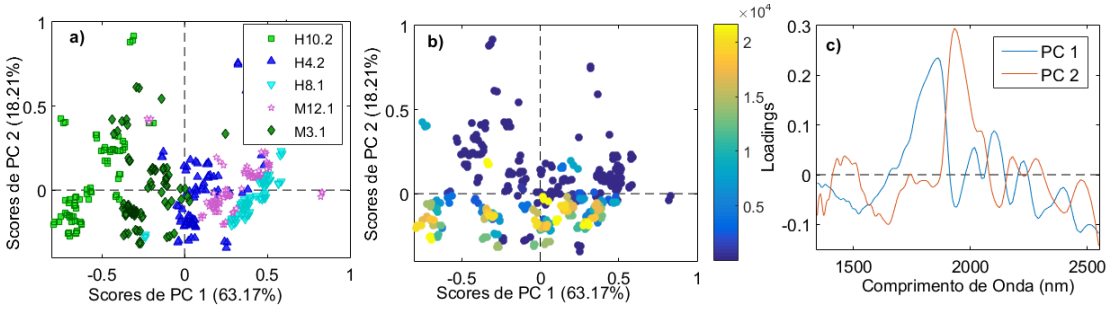
Apenas três amostras de sangue do conjunto de predição foram classificadas incorretamente como amostras recentes e foram projetadas no primeiro modelo de regressão, duas dessas amostras pertenciam ao mesmo doador, uma delas tinha 24 horas e o tempo previsto foi 4,5 horas e a outra era 96 horas e o tempo previsto foi 7,9 h. A terceira amostra pertencia a outro doador e tinha 120 horas e o tempo previsto foi de 4,5 horas.

### 4.3.2 Abordagem Caso-Específico

#### 4.3.2.1 Análise Exploratória

Com base no estudo da primeira estratégia, foram utilizadas as mesmas ferramentas para construir um modelo utilizando um único piso — simulando o que poderia ser encontrado na cena de um crime. A Figura 26 mostra o gráfico do modelo PCA construído com as triplicatas dos espectros de manchas de sangue depositados no piso PO5 para cada tempo. Inicialmente, apenas a Suavização de Savitsky-Golay (Polinômio do 2° grau e janela de 5 pontos), SNV e Centragem na Média foram utilizados como ferramentas de pré-processamento. É possível observar uma tendência de separação entre as amostras de diferentes doadores (Figura 26a) no eixo de PC1. Em relação ao tempo de deposição, ao observar as coordenadas dos *scores* em PC2, pode-se constatar a tendência de separação entre as amostras mais recentes, em azul, e as amostras mais antigas, em amarelo.

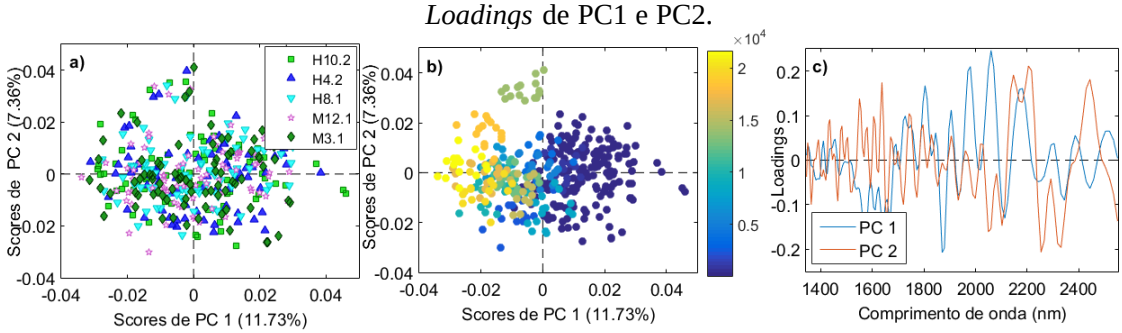
**Figura 26** – Modelo de PCA model construído com os espectros das triplicatas das manchas de sangue depositadas no piso PO5, pré-processados com Suavização, SNV e MC; a) coloridas de acordo com os doadores; b) coloridas de acordo com o tempo de deposição; c) *Loadings* de PC1 e PC2.




**Fonte:** A autora (2024).

Após a aplicação do pré-processamento GLSW (α = 0,03), observou-se que as diferenças dos *scores* observadas no modelo de PCA anterior (Figura 26a) relacionadas aos diferentes doadores foi minimizada, conforme pode ser visto na Figura 27a. Enquanto a separação relacionada ao tempo de deposição foi acentuada (Figura 27b). Os *Loadings* sofrem bastante a influência do GLSW e tornam-se complexos para uma difícil interpretação direta (Figura 27c), contudo é possível verificar uma forte influência da região entre 1800 e 2200 nm, sendo a região em torno de 1940 nm característica de bandas de combinação de vibrações de deformação e de estiramento assimétrico de ligação $H-O-H$ presente em moléculas de água, e as bandas em torno de 2056 nm e 2170 nm podem estar relacionadas com ligações presentes em proteínas como hemoglobina, albumina e globulina (EDELMAN et al., 2012).

**Figura 27** – Modelo de PCA model construído com os espectros das triplicatas das manchas de sangue depositadas no piso PO5, pré-processados com Suavização, SNV, GLSW e MC; a) coloridas de acordo com os doadores; b) coloridas de acordo com o tempo de deposição; c) *Loadings* de PC1 e PC2.




**Fonte:** A autora (2024).

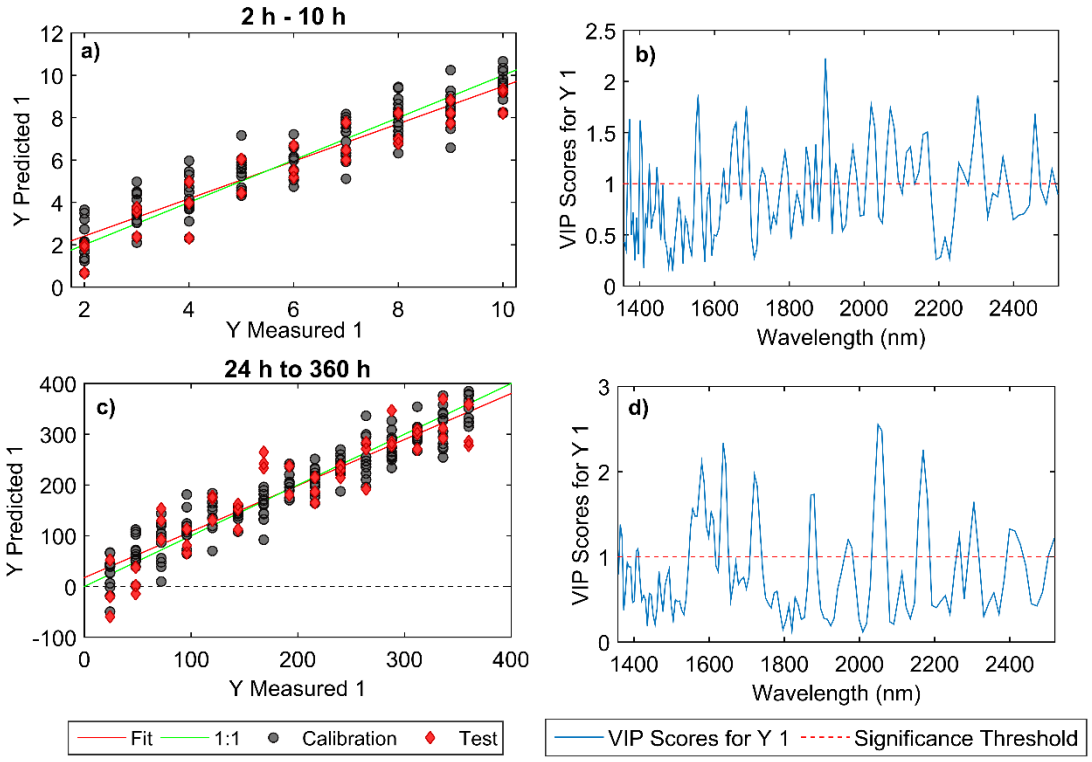
### 4.3.2.2 Modelos de Regressão

Os modelos de regressão foram construídos de maneira semelhante à utilizada na primeira abordagem, discutida no tópico 4.3.1.4, foram construídos dois modelos de regressão, um para cada intervalo de tempo de deposição (o primeiro, para amostras de 2h a 10h, e o segundo para amostras de 1 dia a 15 dias), e utilizou-se as mesmas ferramentas de pré-processamento discutidas. A única diferença desses modelos foi o fato de serem utilizados apenas amostras de um mesmo piso.

A Figura 28 mostra os modelos PLS construídos com os espectros coletados das manchas de sangue depositadas entre 2 horas e 10 horas (Figura 28a), e com espectros coletados das manchas de sangue depositadas entre 24 horas e 360 horas (Figura 28b). Ambos os modelos obtiveram resultados semelhantes aos obtidos com o modelo

hierárquico construído a partir de amostras depositadas em nove pisos diferentes discutidos no tópico 4.3.1.4. Os valores de RMSEP obtidos utilizando a Abordagem Caso-Específico para o primeiro modelo foi de 0,89 h (53,4 min), com *bias* de predição de -0,32 horas (-19,2 minutos) e o RMSEP para o segundo modelo foi de 44,22 h com *bias* de predição de -1,84 horas ( -110,4 minutos). Relembrando que os valores de RMSEP obtidos utilizando o CT4 e a Abordagem Global foram de 0,906h (54,35 min), para o primeiro modelo, e 46,48 h para o segundo modelo.

**Figura 28** – Modelo de regressão PLS construído com amostras depositadas em piso PO5; a) Gráfico de referência versus previsto e b) Gráfico de VIP *Scores* pré-processado com Suavização, SNV, GLSW (α = 0,03) e MC (amostras com tempo de deposição entre 2h e 10h); c) Referência versus previsto e d) VIP *Score*s pré-processado com Suavização, SNV, GLSW (α = 0,03) e MC (amostras com tempo de deposição entre 24 h e 360 h).




**Fonte**: A autora (2024)

Comparando os resultados obtidos do modelo construído com um único piso (Abordagem Caso-Específica) com os dos modelos construídos com diferentes pisos (Abordagem Global), observou-se que não houveram diferenças estatísticas significativas nos valores de RMSEP baseado em um teste F a um nível de 95% de confiança. Os modelos construídos para o intervalo de 2 a 10 horas apresentaram RMSEP próximos de

1 h, enquanto os modelos para o intervalo de 24 horas a 15 dias apresentaram RMSEP de pouco menos de 2 dias. Esses resultados indicam que ambas as estratégias podem ser úteis para estimar o tempo de deposição de manchas de sangue em pisos. Nesse sentido, a escolha da abordagem a ser utilizada pode vir a ser atrelada às vantagens e desvantagens de cada uma delas.

Por exemplo, na Abordagem Global, os modelos são construídos previamente e os resultados são obtidos de forma imediata, agilizando a resposta em um contexto criminal. Contudo, a dificuldade está em garantir que o substrato (piso) e as condições ambientais encontradas no local do crime são semelhantes aos substratos/condições usados na etapa de treinamento. Já na Abordagem Caso-Específico garante-se que o substrato/condições ambientais serão as mesmas encontradas no local, porém haverá morosidade no tempo de resposta, visto que é necessário preparar amostras e acompanhar durante um intervalo de tempo para que se possa construir um modelo nas mesmas condições encontradas no local do crime.

## 4.4 CONCLUSÃO

Os resultados mostram que espectros de infravermelho próximo obtidos com equipamentos portáteis podem ser aplicados *in situ* para estimar o tempo de deposição de manchas de sangue quando associados a modelos hierárquicos construídos a partir da associação de ferramentas quimiométricas como PLS-DA e PLSR. A seleção de métodos de pré-processamento apropriados revela-se crucial para minimizar a interferência do substrato nos espectros, quando a estratégia de modelo global é adotada.

A construção de modelos de regressão focados em diferentes intervalos de tempo mostrou-se uma alternativa para superar os desafios dos diferentes processos cinéticos associados à degradação da hemoglobina e, consequentemente, ao envelhecimento sanguíneo. O modelo desenvolvido utilizando CT4 e GLSW com $\alpha = 0,03$ apresentou desempenho promissor para amostras variando entre 2h e 10h, com *bias* de predição de 0,002 h (0,12 min) e RMSEP de 0,906 h (54,46 min). Quando aplicado a amostras no intervalo de 24h a 360h, o modelo de regressão produziu um *bias* de predição de 2,53 h e um RMSEP de 46,48 h. Os resultados obtidos com as manchas de sangue depositadas em um único piso (abordagem caso-específica) corroboraram os resultados obtidos com a primeira estratégia e não houve diferenças estatísticas significativas nos valores de RMSEP com base em um teste F a um nível de 95% de confiança.

Embora não existam estudos anteriores sobre estimativa do tempo de deposição de manchas de sangue diretamente em pisos utilizando equipamentos portáteis para comparar com os resultados obtidos, um trabalho recente publicado por Gariglio e seu grupo de pesquisa (2024) investigou o uso de espectroscopia UV-Vis e NIR para estimar a deposição tempo de manchas de sangue em lâminas de vidro durante 16 dias, usando uma abordagem caso-específica. O modelo de regressão construído com espectros NIR alcançou um RMSEP de aproximadamente 55 horas, enquanto o modelo construído com espectros UV-Vis alcançou um RMSEP de aproximadamente 76 horas. Eles obtiveram o melhor resultado utilizando fusão de dados de nível médio, usando tanto os espectros NIR quanto os espectros UV-Vis e obtiveram um RMSEP de 35 horas (GARIGLIO et al., 2024). No entanto, é crucial enfatizar não só a disparidade na complexidade dos dados, mas também a praticidade da sua aquisição e da construção dos modelos.

A utilização de equipamentos portáteis torna o processo de aquisição de dados mais prático e adaptável aos contextos da cena do crime. Nesse sentido, os resultados alcançados pelo modelo hierárquico desenvolvido no presente estudo representam um avanço significativo na utilização de equipamentos portáteis de infravermelho próximo em contextos forenses.

# 5 ESTIMATIVA DO TEMPO DE DEPOSIÇÃO DE MANCHAS DE SANGUE EM TECIDOS UTILIZANDO IMAGENS HIPERESPECTRAIS NO INFRAVERMELHO PRÓXIMO E FERRAMENTAS QUIMIOMÉTRICAS

## 5.1 INTRODUÇÃO

Um dos principais desafios das polícias científicas é situar os eventos de um crime e as evidências no tempo durante as investigações. Posto que, ao determinar o tempo de deposição de um vestígio, pode-se garantir que ele realmente seja uma evidência que está relacionada aos fatos, ou descartá-lo para que não confunda o esclarecimento dos acontecimentos (RIBAUX et al., 2010; WEYERMANN; RIBAUX, 2012).

Quando se trata de vestígios como manchas de sangue e demais fluidos biológicos, o desafio de determinar seu tempo de deposição é ainda maior devido à complexidade da composição das amostras, à possível presença de contaminantes, as diferentes condições ambientais às quais podem ter sido submetidas, assim como à interferência de uma infinidade de superfícies sob as quais podem ser encontradas (CANO-TRUJILLO et al., 2023; ZADORA; MENZYK, 2018).

A maioria dos testes utilizados para identificar manchas de sangue em cenas de crime - como o teste de Kastle-Meyer, ou Luminol - utilizam reagentes, estão sujeitos à contaminação e possíveis resultados falsos-positivos (MISTEK; LEDNEV, 2015). Além disso, a maioria dos ensaios são destrutivos, o que limita sua aplicação na investigação forense quando a quantidade de evidências de sangue é pequena. Uma alternativa a estes testes, é o uso de técnicas espectroscópicas, particularmente, as baseadas em espectroscopia no infravermelho próximo (EWING; KAZARIAN, 2017; MARIOTTI; ORTIZ; FERRÃO, 2023; SHARMA; KUMAR, 2018; ZADORA; MENZYK, 2018). Diferentes estudos mostram a aplicação dessas técnicas para identificar, diferenciar e estimar o tempo de deposição de manchas de sangue (HART et al., 2022; KUMAR; SHARMA; SHARMA, 2020; LIN et al., 2018; MORILLAS; GOOCH; FRASCIONE, 2018; SHARMA; KUMAR, 2018; WINNEPENNINCKX et al., 2022; ZHAO et al., 2019).

As evidências de sangue são, muitas vezes, utilizadas para estudar o seu padrão e determinar o tipo de ferimento que causou aquela mancha, a altura da qual o sangue foi depositado, o ângulo de impacto, entre outras informações (ATTINGER et al., 2013; BEVEL; GARDNER, 2008). Tanto que há uma área da ciência forense voltada apenas

para a Análise dos Padrões de Manchas de Sangue (BPA, do inglês *Bloodstain Pattern Analysis*) (ATTINGER et al., 2013; BEVEL; GARDNER, 2008; TAYLOR et al., 2016), contudo, realizar esse estudo pode ser muito difícil quando as manchas estão depositadas em superfícies muito escuras ou estampadas (BEVEL; GARDNER, 2008; MARIOTTI; ORTIZ; FERRÃO, 2023; TAYLOR et al., 2016). Nesse contexto, uma alternativa viável para identificar, discriminar e datar as manchas de sangue é utilizar imagens adquiridas de câmeras hiperespectrais no infravermelho próximo (HSI-NIR), que fornecem tanto informações espaciais quanto informações espectrais (AIRLIE et al., 2020; HART et al., 2022; MARIOTTI; ORTIZ; FERRÃO, 2023).

Edelman e colaboradores (2012) utilizaram uma câmera hiperespectral operando na região do visível e do infravermelho (400-1000 nm) para adquirir imagens de manchas de sangue em tecidos de algodão branco com diferentes tempos (0,5 horas, 2 dias, 15 dias, 40 dias e 200 dias). A partir dessas imagens hiperespectrais, foram construídos modelos de K-Vizinhos mais próximos (KNN, do inglês K-Nearest Neighboor) para estimar o tempo de deposição dessas manchas. Todas as idades estimadas foram comparadas com a idade real das manchas para calcular o erro absoluto e relativo, obtendo-se uma média de erro relativo de 13,4%. Esse estudo foi pioneiro na utilização de imagens hiperespectrais para estimativa do tempo de deposição de manchas de sangue, mas utilizou amostras de apenas um único tipo de tecido e um único doador para construir o modelo. Além disso, não foram construídos modelos de regressão para estimar a idade das manchas de sangue, mas uma ferramenta de agrupamento/classificação, de modo que não se pode classificar manchas de sangue com tempo diferente de cada agrupamento (EDELMAN; VAN LEEUWEN; AALDERS, 2012).

Pereira e colaboradores (2021) também estudaram imagens hiperespectrais no infravermelho próximo para discriminar manchas de sangue humano, sangue animal e possíveis falsos positivos comuns depositadas na superfície de dez diferentes tipos de tecido (algodão e sintético) com diversas colorações e estampas. Foram construídos modelos hierárquicos de classificação combinando diferentes ferramentas quimiométricas com 100% de sensibilidade e 98% de especificidade, provando a eficiência da HSI-NIR para identificar manchas de sangue em diferentes tipos de tecido, incluindo tecidos escuros (PEREIRA et al., 2021).

Malegori e seu grupo de pesquisa estudaram o uso do HSI-NIR tanto para identificar traços invisíveis de fluidos biológicos como sangue, sêmen e urina em tecidos

(MALEGORI et al., 2020), quanto para estimar o tempo de deposição das manchas de fluido biológicos em tecidos (MANIS et al., 2022). No trabalho mais recente, foram utilizados apenas tecidos de algodão e sintético na cor preta e as imagens hiperespectrais foram adquiridas no intervalo de 1 h durante das primeiras 48 h e de 12 h em 12 h durante os 19 dias subsequentes, e a faixa espectral operacional da HSI foi de 1000 nm a 2500 nm. A média espectral dos pixels das manchas foi utilizada para a construção dos modelos de regressão por mínimos quadrados parciais (PLSR) e apresentaram valores de RMSECV de 11,4 h para amostras de sangue em algodão e RMSECV de 15,5 h para amostras de sangue em poliéster (MANIS et al., 2022).

A imagem hiperespectral é particularmente interessante neste contexto, pois é adequada para identificação de evidências sem contato, minimizando o risco de contaminação e destruição de vestígios. Outras vantagens associadas às técnicas espectroscópicas incluem a rapidez de análise, ausência de solventes e pré-tratamentos de amostras, capacidade de discriminar e identificar diversos fluidos biológicos. Além disso, tecidos são um suporte de fácil manuseio, que pode ser levado até os laboratórios periciais, diferente de superfícies como os pisos. Posto que, já é muito comum na rotina desses laboratórios, o encaminhamento de vestimentas para constatação e coleta de material biológico para análises de DNA, por exemplo. Todos estes aspectos tornam a técnica HSI-NIR particularmente apropriada e promissora para perseguir o objetivo de caracterização de vestígios biológicos.

### 5.1.1 Objetivos

O objetivo principal do presente trabalho foi desenvolver uma um método usando imagens hiperespectrais na região do infravermelho próximo para estimar o tempo de deposição de manchas de sangue em diferentes tipos de tecido e de colorações distintas. Nesse sentido, os objetivos específicos deste trabalho foram:

- Avaliar a eficácia da ferramenta de calibração PLS, para construir modelos de regressão capazes de estimar o tempo de deposição das manchas de sangue;
- Avaliar o desempenho de duas diferentes abordagens, global e caso-específico, para a construção de modelos adequando às condições de um vestígio.

## 5.2 MATERIAIS E MÉTODOS

As amostras de sangue foram coletadas através da punção da falange distal do dedo anelar de oito voluntários (duas mulheres e seis homens) pelo uso de lancetas descartáveis manuseadas por um colaborador qualificado. Cada doador forneceu quatro manchas de sangue, totalizando 64 manchas, que foram depositadas na superfície de oito tecidos diferentes, sendo quatro tecidos de algodão (com colorações branca, bege, preto e vermelho) e quatro tecidos sintéticos (com colorações branca, bege, preto e vermelho) como pode ser observado na Figura 29. Para simular as condições comumente encontradas em uma cena de crime, a temperatura e a umidade da sala foram deixadas sem controle, mas monitoradas diariamente durante os 16 dias de análises. A temperatura média da sala foi de aproximadamente 25,7°C (±2°C), e a umidade foi, aproximadamente, 55% (±10%).

A Figura 29 mostra como foi feita a disposição das manchas de sangue depositadas em cada um dos tecidos de algodão e sintético.

**Figura 29 -** Manchas de sangue depositadas em tecidos de algodão e sintético.

**Fonte:** A autora (2024).

Os códigos CWH, CBE, CBL e CRE foram usados para indicar os tecidos de algodão das cores branco, bege, preto e vermelho, respectivamente. E os códigos SWH, SBE, SBL, e SER foram usados para indicar os tecidos sintéticos com coloração branca, bege, preta e vermelha, respectivamente.

Todos os voluntários assinaram o Termo de Livre Consentimento Esclarecido – TCLE (Apêndice B), que foi aprovado pelo Comitê de Ética em Pesquisa envolvendo Seres Humanos da Universidade Federal de Pernambuco (CEP-UFPE), vinculado ao

Projeto cujo Certificado de Apresentação de Apreciação Ética é o n° 50883721.2.0000.5208.

### 5.2.1 Aquisição das Imagens Hiperespectrais

As imagens hiperespectrais foram adquiridas diariamente em um intervalo aproximado de 24 horas durante 15 dias, com a exceção da primeira imagem, a qual foi adquirida após o tempo de deposição de três horas. Para realizar a aquisição das imagens foi utilizada uma câmera Specim FX17 cuja faixa espectral abrange de 900 nm a 1700 nm. O sistema é controlado pelo software Lumo Scanner e, antes de cada medição, imagens escuras (obturador fechado) e brancas (haste Spectralon® de 99% de refletância) foram automaticamente registradas e armazenadas para serem usadas para calcular o valor de refletância espectral (R) para cada pixel e comprimento de onda.

A lente usada produzia imagens com dimensões de pixels com 15 × 15 μm, e resolução espacial de 3,5 nm e resolução espectral de 8 nm. Os parâmetros utilizados durante a aquisição das imagens foram: taxa de quadros igual a 50,00 Hz, velocidade de varredura de 20 mm/s e tempo de exposição de 4 ms. Além disso, o foco manual foi ajustado antes da digitalização. Ao escanear toda a superfície da amostra, foi criada uma imagem hiperespectral tridimensional completa, na qual as duas primeiras dimensões representam a informação espacial enquanto a terceira é a coordenada espectral.

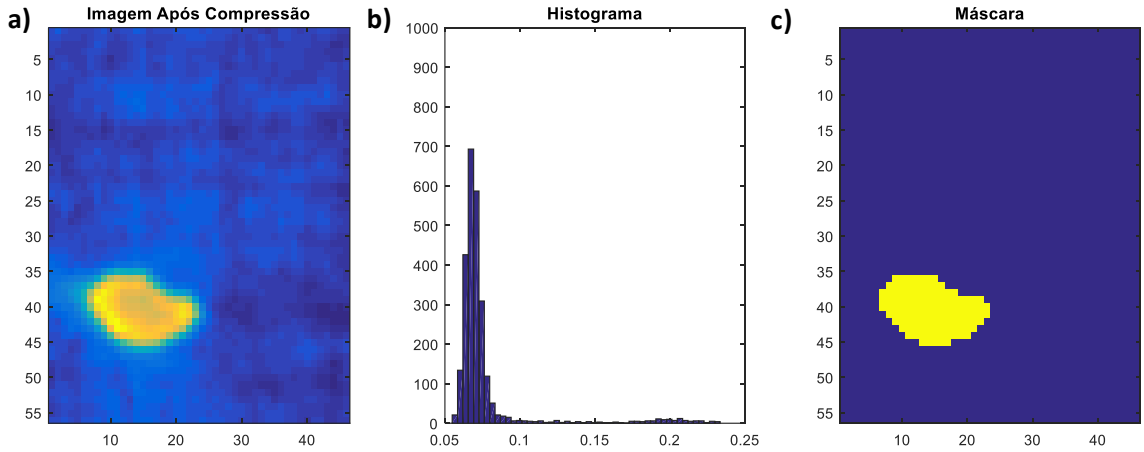
Os espectros foram registrados no modo de reflexão (escala de refletância) no formato *.raw* era e foram convertidos em *.mat* utilizando o software *Evince* que já calculava a informação espectral após a subtração dos valores obtidos para as imagens escuras e brancas. Uma placa de teflon foi colocada na base da bandeja do equipamento para servir de suporte para as amostras e duas barras de teflon foram utilizadas para manter os tecidos na posição e diminuir possíveis efeitos físicos causados por dobras ou enrugamentos dos tecidos.

### 5.2.2 Análise de Dados

Para realizar a análise quimiométrica dos dados espectrais, foi utilizado o software Matlab (MATLAB®, MathWorks) com o PLS Toolbox versão 8.6.2 (Eigenvector Research, Inc). Para realizar a seleção dos pixels da região de interesse (RoI, do inglês *Region of Interest*) foi utilizada uma rotina para construir uma máscara capaz de extrair

os pixels a partir de um histograma de frequência construído através da compressão da intensidade dos pixels em um comprimento de onda, de modo que o *background* foi excluído e apenas os pixels referentes a mancha de sangue foram selecionados, conforme pode ser visto na Figura 30.

**Figura 30** – a) Imagem reconstruída da mancha de sangue no tecido após a compressão; b) Histograma de frequência; c) Imagem falsa evidenciando o recorte dos pixels selecionados pela máscara.




**Fonte:** A autora (2024).

Após a seleção da RoI, o hipercubo foi desdobrado em uma matriz de dados em que cada pixel passou a representar uma amostra. Como a quantidade de pixels por mancha de sangue varia de acordo com o tamanho da mancha de sangue, foi calculada a média dos espectros de todos os pixels de cada mancha. Para realizar a análise exploratória e para a construção dos modelos foi utilizada a média espectral dos pixels de cada mancha de sangue em cada intervalo de tempo de deposição.

As imagens hiperespectrais no NIR são, geralmente, afetadas pelas irregularidades da superfície das amostras o que gera efeitos indesejados como espalhamento da radiação e ruído espectral. Por esse motivo, diferentes ferramentas de pré-processamento foram avaliados para atenuar esses efeitos espectrais, como o filtro de suavização de Savitzky-Golay (usando janelas com 3, 5 e 7 pontos), SNV, GLSW, 1ª Derivada de Savitzky-Golay (usando janelas com 3, 5 e 7 pontos), e centragem na média. A análise exploratória foi realizada através da construção de modelos de PCA.

Para realizar a escolha do parâmetro α referente ao filtro da ferramenta de pré-processamento GLSW, foram construídas superfícies de respostas, seguindo a mesma metodologia referenciada no Capítulo 4 da presente tese. Os modelos de regressão foram

construídos utilizando PLSR, e foram testadas diferentes abordagens: a primeira delas, envolveu a construção de um modelo de regressão único que pudesse ser utilizado tanto para os tecidos de algodão quanto para os tecidos sintéticos, também foram construídos modelos de regressão para cada um dos tipos de fibra de tecido (algodão e sintético) utilizando a média dos espectros dos pixels de cada mancha de sangue. Essa abordagem visou a construção de modelos que pudessem ser utilizados de acordo com o tipo de tecido, sem importar o seu tingimento, apenas a fibra da sua estrutura.

Considerando que em um contexto criminal, as manchas de sangue podem ser encontradas em diversos tipos de tecidos, e podem ser submetidas a diferentes condições ambientais, a segunda abordagem foi mais específica, e tratou-se de uma abordagem caso-a-caso, em que foi construído um modelo específico para um único tipo de tecido (simulando um tecido supostamente encontrado em um local de crime). Para essa construção, os dados coletados de um único doador depositados em um tecido ao longo do tempo foram utilizados para a construção do modelo e os dados de um doador externo depositados no mesmo tecido foram utilizados como conjunto de predição.

## 5.3 RESULTADOS E DISCUSSÃO

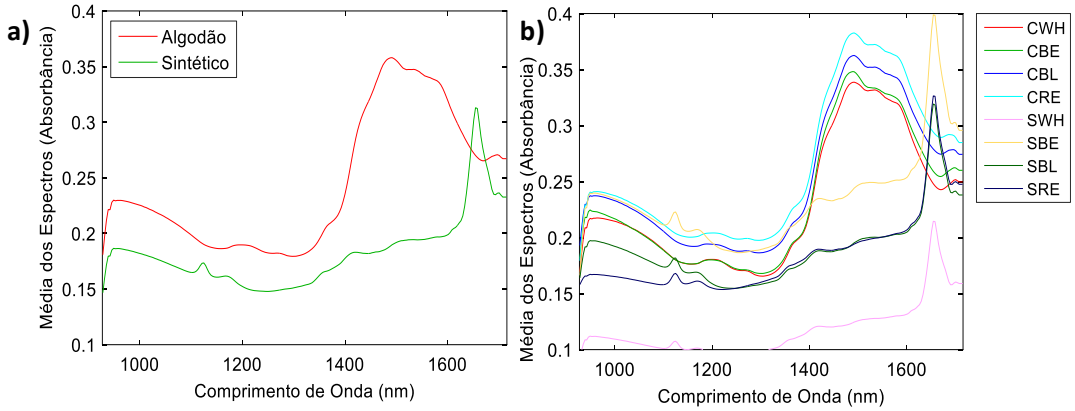
### 5.3.1 Abordagem Global

#### 5.3.1.1 Análise Espectral e Exploratória

A seleção da região de interesse (RoI) foi aplicada a cada imagem hiperespectral das manchas de sangue depositadas nos diferentes tecidos e intervalos de tempo. Após a seleção dos pixels de interesse, a média dos espectros dos pixels de cada uma das manchas foi calculada e os dados foram organizados em um único conjunto para a avaliação dos espectros originais e escolha do melhor pré-processamento. A Figura 31 mostra a média dos espectros originais de sangue depositados nos diferentes tipos de tecido. Na Figura 31a, é possível observar a interferência do tipo de tecido no qual o sangue foi depositado nos espectros brutos, sem pré-processamento. Os espectros de sangue depositados em tecido sintético apresentam bandas mais estreitas em duas regiões: a primeira, em torno de 1128 nm e a segunda em torno de 1660 nm, sendo essas regiões comumente encontradas em espectros de poliéster confeccionados a partir de tereftalato de etileno (PET, do inglês *Polyethylene terephthalate*) (BURNS; CIURCZAK, 2009), sendo a região de 1660 nm referente à banda de absorbância principal para PET devido a

sobretons de estiramento da ligação $-C-H$ presente nesses compostos (RANI et al., 2019).

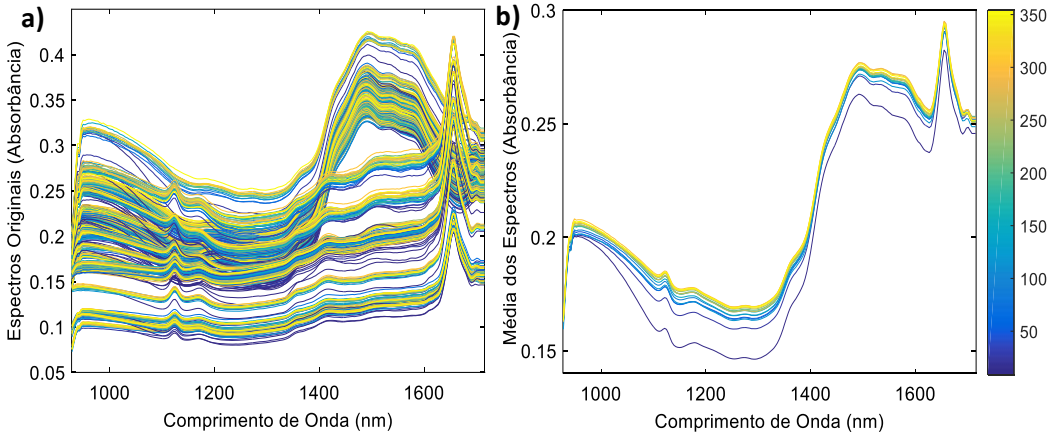
**Figura 31** – Média dos espectros das manchas de sangue depositadas nos tecidos a) colorido de acordo com as classes de tipo de tecido algodão e sintético; b) coloridos de acordo com as diferentes classes de tecido.




**Fonte:** A autora (2024).

A Figura 32 mostra os espectros obtidos a partir da média de cada pixel das manchas de sangue nos diferentes tempos de deposição - coloridos de acordo um gradiente de tempo variando de 3 h (azul) a 360 h (amarelo). Na Figura 32a pode-se observar que a maior diferença espectral entre as amostras ocorre, principalmente, no intervalo entre as primeiras horas (3 h a 24 h) e entre os primeiros três dias. E essa mudança ocorre principalmente na região entre 1000 nm e 1200 nm (região da banda de combinação de vibrações de estiramento da ligação $-O-H$ presente em moléculas de água que se encontra em torno de 985 nm). Além da região, entre 1300 nm e 1500 nm, que também está relacionada a grupos metil ($-CH_3$) - entre 1347-1367 nm - e ao primeiro sobretom de estiramento da ligação $-O-H$ (1440 e 1485 nm). É importante destacar que a região em torno de 1690 nm pode estar relacionada ao primeiro sobretom de vibrações de estiramento de ligações $-C-H$ presentes em proteínas como hemoglobina, globulina e albumina (EDELMAN et al., 2012).

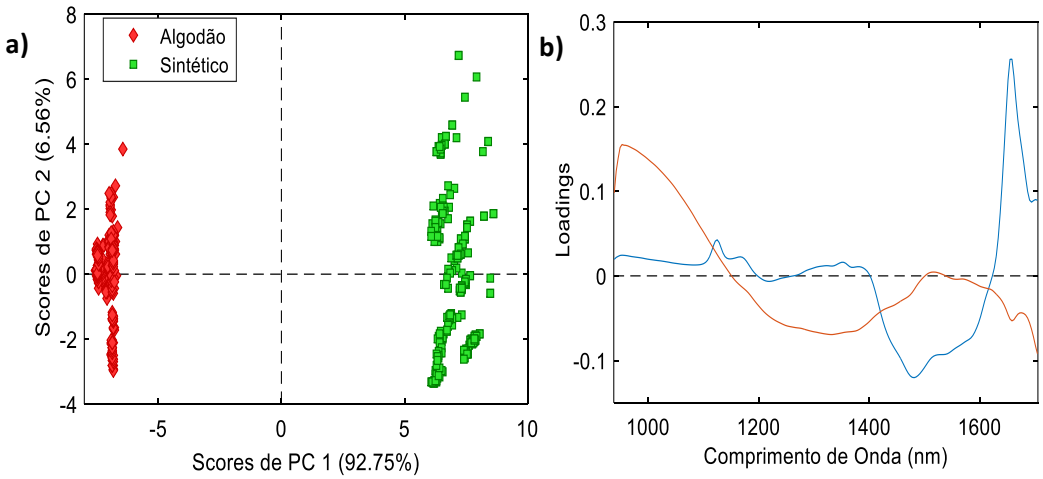
**Figura 32** – a) Espectros originais obtidos a partir da média dos pixels das manchas de sangue coloridos de acordo com a escala de tempo; b) Média dos espectros obtidos a partir da média dos pixels das manchas de sangue coloridos de acordo com a escala de tempo;




**Fonte:** A autora (2024).

Após a aplicação das ferramentas de pré-processamento (Suavização de Savitzky-Golay com 7 pontos e polinômio do 2°grau, SNV e centragem na média), foi realizada a Análise Exploratória construindo um modelo de PCA (Figura 33), no qual foi possível constatar que a primeira componente principal (PC1) separa os *scores* das manchas de sangue depositadas nos tecidos de algodão dos *scores* das manchas de sangue depositadas nos tecidos sintético (Figura 33a). Ao observar os *loadings* dessa mesma componente (PC1, em azul), nota-se que a região da banda em torno de 1660 nm é a principal responsável por essa separação, assim como a região em torno de 1128 nm, sendo essas regiões características de tecidos de poliéster (Figura 33b).

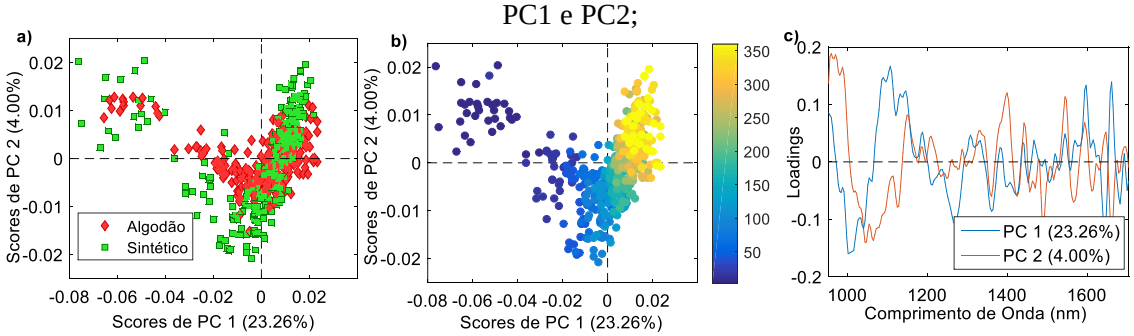
**Figura 33** – Modelo de PCA das amostras depositadas nos tecidos de algodão e sintéticos pré-processadas usando Suavização de Savitzky-Golay com 7 pontos e polinômio do 2°grau, SNV e centragem na média; a) *Scores* de PC1xPC2; b) *Loadings* de PC1 e PC2.




**Fonte:** A autora (2024).

Quando a ferramenta de GLSW ($\alpha = 0.02$) é adicionada ao pré-processamento e o modelo de PCA é novamente construído, observou-se uma mudança na tendência de separação dos *scores*. Na Figura 34a, é possível constatar que as diferenças entre as amostras depositadas nos tecidos de algodão e nos tecidos sintéticos são minimizadas. A Figura 34b, retrata o mesmo modelo PCA com coloração de acordo com o intervalo de tempo de deposição, evidenciando que após o uso do GLSW a tendência de separação das amostras passa a estar relacionada ao tempo de deposição da mancha de sangue.

**Figura 34** – Modelo de PCA das amostras depositadas nos tecidos de algodão e sintéticos pré-processadas usando Suavização de Savitzky-Golay com 7 pontos e polinômio do 2°grau, SNV, GLSW ($\alpha = 0.02$) e centragem na média; a) *Scores* de PC1xPC2; b) *Loadings* de PC1 e PC2;




**Fonte:** A autora (2024).

### *5.3.1.2 Modelos de Regressão*

Para a construção dos modelos de regressão, também foram utilizadas as ferramentas de pré-processamento de Suavização, SNV e GLSW ($\alpha = 0.02$). Para escolher o valor do parâmetro $\alpha$ a ser utilizado, foram construídas superfícies de resposta seguindo a mesma metodologia mencionada no Capítulo 4 desta tese. Inicialmente, foi construído um modelo de regressão único, utilizando a média dos espectros das manchas de sangue depositadas nos tecidos de algodão e sintético. Para formar o conjunto de treinamento, utilizou-se os dados obtidos a partir das manchas dos doadores H1, H2, H3, H4, H5 e M2, totalizando 384 espectros médios. Enquanto o conjunto de predição foi formado pelos dados obtidos a partir das manchas dos doadores M1 e H6, totalizando 128 espectros médios. Em seguida, foram construídos mais dois modelos de regressão: um modelo para manchas de sangue depositadas em tecidos de algodão, e outro para manchas de sangue depositadas em tecido sintético.

A Tabela 7 mostra os resultados das figuras de mérito dos diferentes modelos de regressão construídos. Todos os modelos usaram os dados de todos os intervalos de tempo de deposição, ou seja, de 3 h a 15 dias (360 h), utilizaram as mesmas ferramentas de pré-processamento e duas variáveis latentes. Todos os valores do coeficiente de Pearson da validação cruzada ($R^2_{CV}$) foram acima de 0,89 indicando um bom ajuste do modelo.

É evidente que os modelos construídos para as manchas de sangue depositadas especificamente em um tipo de tecido obtiveram menores valores de erros quadráticos médios de validação cruzada (RMSECV) e de predição (RMSEP). Os modelos específicos construídos para as manchas de sangue depositadas em um único tipo de tecido obtiveram RMSEP de 43,91 h (menos de dois dias) para as manchas de sangue depositadas em algodão, e RMSEP de 38,89 h para as manchas de sangue depositadas em tecido sintético, enquanto o modelo que englobava as manchas depositadas em ambos os tipos de tecido obteve um valor de RMSEP de 50,98 h.

Esse resultado era esperado, visto que o tipo de fibra interfere na absorção da mancha de sangue por parte do tecido e, consequentemente, na cinética de degradação e secagem da mancha, e uma das principais diferenças entre os tecidos de algodão e os tecidos sintéticos (em sua maioria de poliéster) é em relação a sua hidrofobicidade. Os tecidos de algodão absorvem a água com mais facilidade, enquanto a maioria dos tecidos sintéticos são hidrofóbicos. Nesse sentido, é esperado que a molhabilidade do tecido interfira no processo de secagem da mancha de sangue e, consequentemente, no processo de degradação.

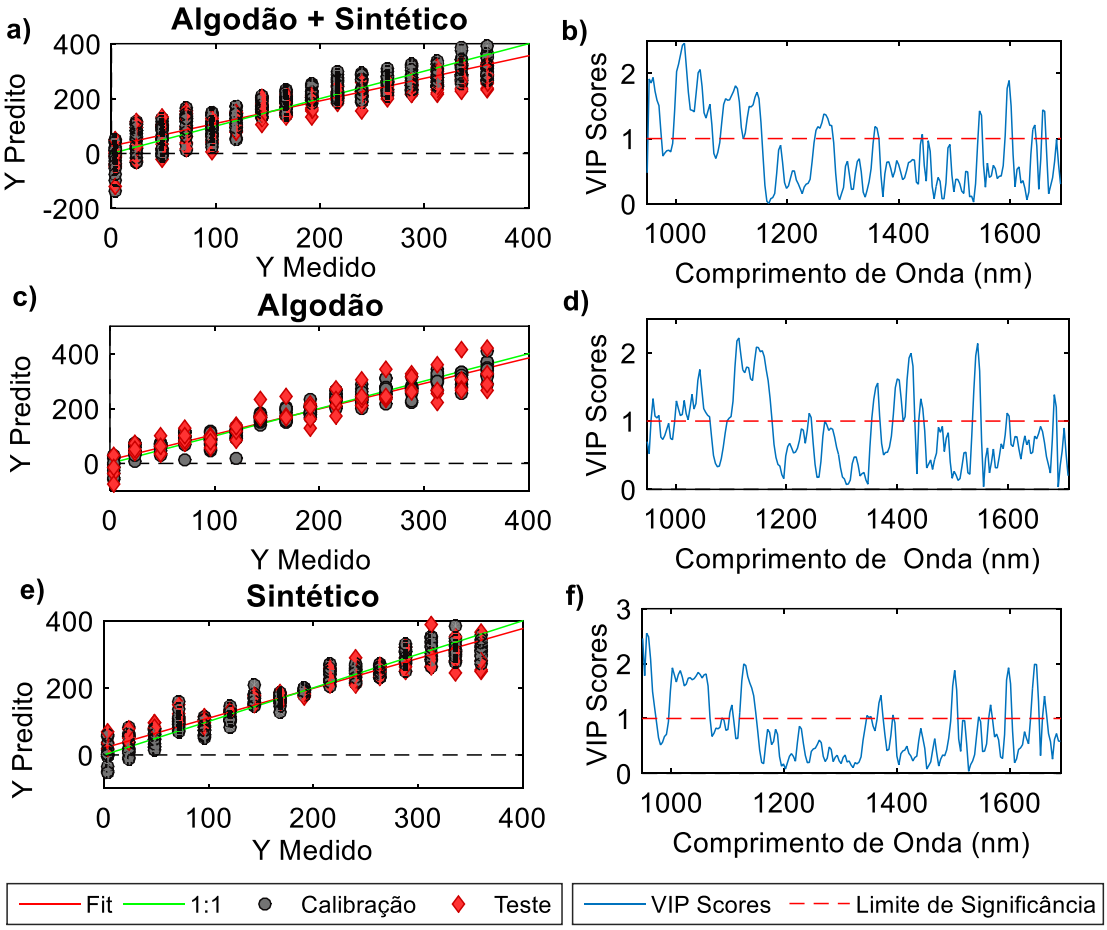
**Tabela 7.** Valores obtidos das figuras de mérito para os modelos.

| Conjunto de Treinamento | Conjunto de Predição | Pré-Processamento | LV | $R^2_{CV}$ | $R^2_{Pred}$ | RMSECV (h) | RMSEP (h) | $Bias_{CV}$ (h) | $Bias_{Pred}$ (h) |
|---|---|---|---|---|---|---|---|---|---|
| Algodão + Sintético | M1 e H6 | SNV; GLSW ($\alpha$ = 0.02) | 2 | 0.89 | 0.84 | 49,34 | 50,98 | -0,37 | -16,79 |
| Algodão | M1 | SNV; GLSW ($\alpha$ = 0.02) | 2 | 0.89 | 0.84 | 36,51 | 43,91 | -0,38 | 5,02 |
| Sintético | H6 | SNV; GLSW ($\alpha$ = 0.02) | 2 | 0,89 | 0,89 | 36,91 | 38,89 | -0,03 | 3,12 |

O *Bias* de Predição ($Bias_{Pred}$) também apresentou valores mais elevados no primeiro modelo, sendo -16,79 h, enquanto os valores para o modelo construído exclusivamente para manchas de sangue depositadas em tecido de algodão foi 5,02 h e o $Bias_{Pred}$ do modelo específico para manchas de sangue depositadas em tecido sintético foi

3,12 h. A Figura 35 mostra os gráficos de resposta do Y Predito versus Y Medido e os VIP *Scores* obtidos para cada modelo.

**Figura 35** – Modelos de regressão PLS construídos utilizando a 1ª abordagem. a) Y Predito x Y Medido e b) Gráfico de VIP *Scores* do modelo construído com as manchas de sangue depositadas em tecido sintético e algodão; c) Y Predito x Y Medido e d) Gráfico de VIP *Scores* do modelo construído com as manchas de sangue depositadas em tecido de algodão; e) Y Predito x Y Medido e f) Gráfico de VIP *Scores* do modelo construído com as manchas de sangue depositadas em tecido sintético.




**Fonte:** A autora (2024).

Um estudo recente elaborado por Manis e colaboradores, (2022) utilizou imagens hiperespectrais obtidas de manchas de sangue depositadas em tecidos de algodão e sintéticos adquiridas a cada 1 hora durante as primeiras 48 horas e a cada 12 horas durante os 19 dias subsequentes, e a faixa espectral operacional do HSI foi de 1000 nm a 2500 nm. A média espectral dos pixels da mancha foi utilizada para construir modelos de regressão de mínimos quadrados parciais (PLS) e apresentou valores de RMSECV de 11,4 h e RMSEP de 10,8 h para amostras de sangue de algodão; RMSECV de 15,5 h e

RMSEP 16,5 h para amostras de sangue em poliéster. Apesar desse estudo ter obtido menores valores de erros em comparação ao do presente trabalho, é importante destacar que ele foi realizado utilizando um equipamento que abrange uma faixa espectral mais larga, que contém uma região de bandas de combinações de grupos hidroxila e de grupos amida. Esses grupos podem estar relacionados às substâncias presentes no sangue, como proteínas, albumina, globulina e hemoglobina, o que pode fornecer informações espectrais mais eficientes para relacioná-las ao processo de envelhecimento dos elementos presentes no sangue. Além disso, no estudo desenvolvido por eles, foram coletadas mais imagens hiperespectrais e com menor intervalo de tempo entre elas, o que permitiu a construção de um modelo de regressão com um maior ajuste a cinética de degradação das manchas de sangue.

### 5.3.2 Abordagem Caso-Específico

De modo análogo ao discutido no Capítulo 4, foi realizada uma segunda abordagem com o intuito de simular uma situação caso-específica, em que um tecido com uma mancha de sangue teria sido encontrado em um local de crime. Nesse sentido, foi construído um modelo de regressão com os dados de duas manchas de sangue fornecidas pelo mesmo doador (H1) e depositadas no tecido de algodão bege (CBE). O conjunto de predição foi composto pelos dados de duas manchas de sangue depositadas sob o mesmo tecido, mas fornecidas por outro doador (M1). Dessa forma, o tamanho do conjunto de predição foi o mesmo do conjunto de treinamento, sendo constituído por 32 espectros médios (sendo 16 espectros por mancha de sangue referente a cada intervalo de tempo) cada.

De maneira análoga, foi realizada a mesma abordagem para o tecido sintético bege (SBE), o conjunto de treinamento foi composto pelas amostras das duas manchas obtidas através do doador H2, enquanto o conjunto de predição foi obtido através dos dados do doador M4.

O pré-processamento utilizado para a construção dos modelos de regressão usando PLS foi o mesmo descrito na seção anterior: Suavização de Savistzky-Golay (polinômio do 2° grau e janela de 5 pontos), SNV e GLSW ($\alpha = 0,02$).

A Tabela 8 mostra as figuras de mérito obtidas para cada modelo, e a Figura 36 mostra os ajustes obtidos dos modelos PLS construídos utilizando a abordagem caso-específico para o tecido de algodão CBE. Quando se observa apenas os dados da Tabela

8 obtidos na construção dos modelos utilizando as manchas depositadas no tecido CBE, os resultados são muito semelhantes aos obtidos na abordagem global, apresentando uma pequena melhora. O RMSEP passou de 43,91 h para 38,11 h, diminuindo o erro em aproximadamente 6 h. No entanto, quando o teste F é realizado para comparar os valores de RMSEP foi calculado que eles são estatisticamente equivalentes a um nível de 95% de confiança.

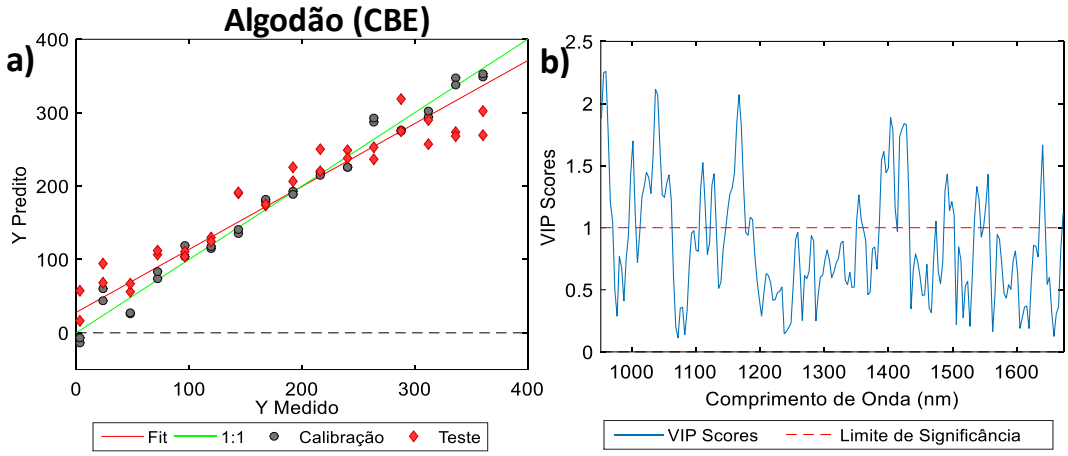
Em relação ao *Bias* de Predição ($Bias_{Pred}$) observou-se pouca variação, de 5,02 h para 4,45 h. Contudo, esses valores ainda são bastante elevados e, ao observar o gráfico da Figura 36, percebe-se que as amostras mais recentes, com tempo inicial de 3 h parecem ter uma maior influência no valor do $Bias_{Pred}$. Como já foi mencionado anteriormente, a cinética de degradação das manchas de sangue tem um comportamento diferente nas primeiras horas de deposição, algo que pode influenciar diretamente no comportamento na inclinação da reta de regressão e, consequentemente, no valor do bias.

**Tabela 8.** Valores obtidos das figuras de mérito para os modelos utilizando a 2ª abordagem nas amostras do tecido sintético CBE.

| Conjunto de Treinamento | Conjunto de Predição | Pré-Processamento | LV | $R^2_{CV}$ | $R^2_{Pred}$ | RMSECV (h) | RMSEP (h) | $Bias_{CV}$ (h) | $Bias_{Pred}$ (h) |
|---|---|---|---|---|---|---|---|---|---|
| H1 (CBE) | M1 (CBE) | SNV; GLSW ($\alpha$ = 0,02) | 3 | 0,93 | 0,92 | 29,31 | 38,11 | -1,90 | 4,45 |

A Figura 36 mostra o ajuste de Y Predito versus Y Medido para o modelo de regressão construído com as manchas de sangue depositadas no tecido CBE. Pode-se observar que o comportamento das manchas de sangue utilizadas no conjunto de predição/teste (M1) – em vermelho – têm um comportamento semelhante ao das amostras utilizadas no conjunto de treinamento (H1) que pode ser percebido através do menor valor de RMSEP e no padrão aleatório de distribuição das amostras de predição em torno da reta de calibração.

**Figura 36** – Modelos de regressão PLS construídos utilizando a abordagem caso-específico nas amostras do tecido de algodão (CBE). a) Y Predito x Y Medido e b) Gráfico de VIP *Scores* do modelo.




**Fonte:** A autora (2024).

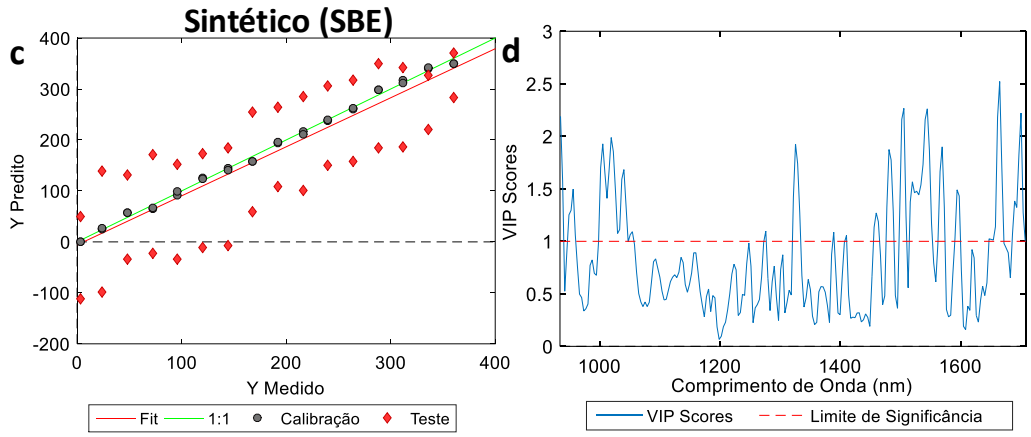
Contudo, ao observar os resultados obtidos para o modelo construído com as manchas de sangue depositadas no tecido sintético bege (SBE) na Tabela 9, percebe-se que houve uma piora significativa nos resultados, não só em relação aos valores de RMSEP e $Bias_{Pred}$ como também do $R^2_{Pred}$.

**Tabela 9.** Valores obtidos das figuras de mérito para os modelos utilizando a abordagem caso-específico nas amostras do tecido sintético SBE.

| Conjunto de Treinamento | Conjunto de Predição | Pré-Processamento | LV | $R^2_{CV}$ | $R^2_{Pred}$ | RMSECV (h) | RMSEP (h) | $Bias_{CV}$ (h) | $Bias_{Pred}$ (h) |
|---|---|---|---|---|---|---|---|---|---|
| H4 (SBE) | M2 (SBE) | SNV; GLSW (α = 0,02) | 2 | 0,96 | 0,67 | 21,19 | 68,74 | 1,58 | -13,51 |

Ao visualizar o gráfico de Y Predito x Y Medido obtido a partir do modelo PLS, pode-se constatar que ainda que as duas amostras de mancha de sangue do conjunto de predição/teste tenham sido fornecidas pelo mesmo doador, elas aparentam ter um comportamento diferente entre si. E que o comportamento das manchas de sangue do conjunto de treinamento (H4) parece ter pouca variabilidade. É importante destacar que cada amostra está relacionada a média dos espectros obtidos em cada pixel referente à mancha de sangue, nesse sentido, apesar da média ser representativa da totalidade da mancha de sangue, utilizar a média diminui a variabilidade do conjunto.

**Figura 37** – Modelos de regressão PLS construídos utilizando a abordagem caso-específico nas amostras do tecido sintético (SBE). a) Y Predito x Y Medido e b) Gráfico de VIP *Scores* do modelo.




**Fonte:** A autora (2024).

## 5.4 CONCLUSÃO

Este capítulo apresenta um estudo que buscou estimar o tempo de deposição de manchas de sangue em diferentes tipos de tecidos, utilizando imagens obtidas por meio de uma câmera hiperespectral no infravermelho próximo, na faixa de 1000 a 1700 nm, aliadas a ferramentas quimiométricas e considerando fins forenses.

Uma das principais vantagens de utilizar imagens hiperespectrais NIR para estimar a idade de manchas de sangue é o fato de ser uma técnica completamente não destrutiva e não invasiva, o que a torna especialmente adequada para a detecção direta de vestígios biológicos no local original em que foi encontrado sem risco de danos e permitindo análises futuras, como a de DNA. Além disso, trata-se de um método rápido que, quando combinado com ferramentas quimiométricas, pode fornecer respostas de forma ágil. Ambas as abordagens oferecem possibilidades para adequação ao contexto forense.

Pela primeira vez, investigou-se a possibilidade de utilizar um único modelo quimiométrico para estimar o envelhecimento de manchas de sangue depositadas em tecidos de algodão e materiais sintéticos de diferentes colorações. Além disso, os resultados obtidos através dos modelos construídos com maior variabilidade (abordagem global) foram comparados com os resultados obtidos através de um modelo de regressão construído especificamente para se adequar ao contexto em que a mancha de sangue teria sido encontrada (abordagem caso-específica).

Atualmente, ainda não existe uma técnica amplamente validada para o estudo do envelhecimento de vestígios de manchas de sangue. A maioria das pesquisas nessa área está em fase de desenvolvimento, necessitando ampliação dos estudos para a adequação não apenas aos diferentes substratos onde as manchas podem ser depositadas, como também às diferentes condições ambientais nas quais elas podem ser encontradas, como umidade, temperatura, contaminantes e outras interferências que podem desempenhar um papel fundamental na cinética de degradação do vestígio.

Nesse contexto, a abordagem caso-específica é interessante por possibilitar a adequação da metodologia às condições do local no qual a mancha foi encontrada. No entanto, é importante destacar que essa abordagem demanda tempo do especialista forense para realizar o exame, o que nem sempre é possível devido à urgência e rapidez exigidas nas investigações criminais. Por essa razão, uma abordagem mais abrangente, com modelos que considerem uma maior variabilidade de tecidos e que possam ser aplicados a diferentes contextos, torna-se fundamental. O presente estudo mostrou que, de fato, os modelos desenvolvidos utilizando HSI e ferramentas quimiométricas podem contribuir para uma estimativa mais precisa da idade das manchas de sangue depositadas em tecidos.

# 6 PERSPECTIVAS FUTURAS

O presente trabalho apresenta vários avanços no uso de espectroscopia NIR combinada com ferramentas quimiométricas para aplicações forenses na área de identificação e estimativa do tempo de deposição de manchas de sangue. Contudo, ainda existem muitos aspectos que ainda precisam ser estudados antes da implementação de uma metodologia validada para aplicação em cenários forenses reais. Nesse sentido, como perspectivas futuras para ampliar os estudos na área, é necessário acrescentar um maior número de doadores para uma maior representativa da população, assim como considerar um maior número de substratos (pisos e tecidos) para aumentar a variabilidade dos conjuntos de treinamento.

Além disso, um dos principais desafios é avaliar as consequências no processo de degradação de manchas de sangue relacionadas a condições ambientais as quais elas podem ser expostas, como temperatura, umidade, incidência de luz solar, contaminação microbiana, entre outros fatores que podem alterar a cinética de degradação e, consequentemente, a estimativa do tempo de deposição das manchas de sangue.

## REFERÊNCIAS

ABDULLAH SANI, Muhamad Shirwan; ISMAIL, Azilawati Mohd; AZID, Azman; SAMSUDIN, Mohd Saiful. Establishing forensic food models for authentication and quantification of porcine adulterant in gelatine and marshmallow. **Food Control**, v. 130, 2021. DOI: 10.1016/j.foodcont.2021.108350.

AGUDELO, Juliana; HALÁMKOVÁ, Lenka; BRUNELLE, Erica; RODRIGUES, Roselyn; HUYNH, Crystal; HALÁMEK, Jan. Ages at a Crime Scene: Simultaneous Estimation of the Time since Deposition and Age of Its Originator. **Analytical Chemistry**, v. 88, p. 6479–6484, 2016. DOI: 10.1021/acs.analchem.6b01169.

AIRLIE, Melissa; CHASELING, Janet; KROSCH, Matt N.; WRIGHT, Kirsty. An evaluation of infrared photography for detecting bloodstains on dark-coloured floor coverings commonly encountered at crime scenes. **Australian Journal of Forensic Sciences**, p. 1–13, 2020. DOI: 10.1080/00450618.2020.1805012.

AMIGO, José Manuel; BABAMORADI, Hamid; ELCOROARISTIZABAL, Saioa. **Hyperspectral image analysis. A tutorial**. **Analytica Chimica Acta**. 2015. DOI: 10.1016/j.aca.2015.09.030.

ANDERSON, Stacey E.; HOBBS, Gerald R.; BISHOP, Clifton P. Multivariate Analysis for Estimating the Age of a Bloodstain. **Journal of Forensic Sciences**, v. 56, p. 186–193, 2011. DOI: 10.1111/j.1556-4029.2010.01551.x.

ATTINGER, Daniel; MOORE, Craig; DONALDSON, Adam; JAFARI, Arian; STONE, Howard A. **Fluid dynamics topics in bloodstain pattern analysis: Comparative review and research opportunities**. **Forensic Science International**, 2013. DOI: 10.1016/j.forsciint.2013.04.018.

BALLABIO, Davide; CONSONNI, Viviana. **Classification tools in chemistry. Part 1: Linear models. PLS-DA**. **Analytical Methods**, 2013. DOI: 10.1039/c3ay40582f.

BASANTIA, NC; NOLLET, Leo ML; KAMRUZZAMAN, Mohammed. **Hyperspectral Imaging Analysis and Applications for Food Quality**. Boca Raton - FL, 2019.

BEĆ, Krzysztof B.; GRABSKA, Justyna; HUCK, Christian W. **Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers**. **Chemistry - A European Journal**, 2021. DOI: 10.1002/chem.202002838.

BEEBE, Kenneth R.; PELL, Randy J.; SEASHOLTZ, Mary Beth. **Chemometrics: A Practical Guide**. **Technometrics**, 1998. DOI: 10.2307/1271366.

BENABDELHALIM, H.; BRUTIN, David. **Drying of Complex Fluid Drops Fundamentals and Applications**.: The Royal Society of Chemistry, 2022.

BEVEL, Tom; GARDNER, Ross M. **Bloodstain pattern analysis with an introduction to crime scene reconstruction,**. 3rd. ed., 2008.

BREMMER, Rolf H.; DE BRUIN, Daniel M.; DE JOODE, Maarten; BUMA, Wybren Jan; VAN LEEUWEN, Ton G.; AALDERS, Maurice C. G. Biphasic oxidation of Oxy-Hemoglobin in bloodstains. **PLoS ONE**, v. 6, n. 7, 2011. a. DOI: 10.1371/journal.pone.0021845.

BREMMER, Rolf H.; DE BRUIN, Karla G.; VAN GEMERT, Martin J. C.; VAN LEEUWEN, Ton G.; AALDERS, Maurice C. G. Forensic quest for age determination of bloodstains. **Forensic Science International**, v. 216, p. 1–11, 2012. DOI: 10.1016/j.forsciint.2011.07.027.

BREMMER, Rolf H.; NADORT, Annemarie; VAN LEEUWEN, Ton G.; VAN GEMERT, Martin J. C.; AALDERS, Maurice C. G. Age estimation of blood stains by hemoglobin derivative determination using reflectance spectroscopy. **Forensic Science International**, v. 206, p. 166–171, 2011. b. DOI: 10.1016/j.forsciint.2010.07.034.

BRERETON, Richard G. **Chemometrics: Data Analysis for the Laboratory and Chemical Plant**. Bristol, UK, 2003. v. 8

BRERETON, Richard G.; JANSEN, Jeroen; LOPES, João; MARINI, Federico; POMERANTSEV, Alexey; RODIONOVA, Oxana; ROGER, Jean Michel; WALCZAK, Beata; TAULER, Romà. Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools. **Analytical and Bioanalytical Chemistry**, v. 409, p. 5891–5899, 2017. DOI: 10.1007/s00216-017-0517-1.

BRO, Rasmus; SMILDE, Age K. **Principal component analysis**. **Analytical Methods**, 2014. DOI: 10.1039/c3ay41907j.

BROWN, Steven D. The chemometrics revolution re-examined. **Journal of Chemometrics**, v. 31, p. 1–23, 2017. DOI: 10.1002/cem.2856.

BRUTIN, D.; SOBAC, B.; LOQUET, B.; SAMPOL, J. Pattern formation in drying drops of blood. **Journal of Fluid Mechanics**, v. 667, p. 85–95, 2011. DOI: 10.1017/S0022112010005070.

BUENDIA GARCIA, Jhon; GORNAY, Julien; LACOUE-NEGRE, Marion; MAS GARCIA, Silvia; ER-RMYLY, Jihane; BENDOULA, Ryad; ROGER, Jean Michel. A novel methodology for determining effectiveness of preprocessing methods in reducing undesired spectral variability in near infrared spectra. **Journal of Near Infrared Spectroscopy**, v. 30, p. 74–88, 2022. DOI: 10.1177/09670335211047959.

BURNS, Donald A.; CIURCZAK, Emil W. **Handbook of Near-Infrared Analysis**. 3rd. ed. CRC Press: Taylor & Francis Group, 2009.

CANO-TRUJILLO, Cristina; GARCÍA-RUIZ, Carmen; ORTEGA-OJEDA, Fernando E.; ROMOLO, Francesco; MONTALVO, Gemma. **Forensic analysis of biological fluid stains on substrates by spectroscopic approaches and chemometrics: A review**. **Analytica Chimica Acta**Elsevier B.V., 2023. DOI: 10.1016/j.aca.2023.341841.

CARNEIRO, Caroline R.; SILVA, Carolina S.; WEBER, Ingrid T. Preliminary study of fingerprint aging using Near Infrared Hyperspectral Imaging (NIR-HIS). **Analytical Methods**, 2023.

CHANG, Chein-I. **Hyperspectral Imaging: Techniques for Spectral Detection and Classification**. New York: Kluwer Academic/Plenum Publishers, 2003.

CHENG, Feng; LI, Wanting; JI, Zhimin; LI, Junli; HU, Wenjing; ZHAO, Mengyang; YU, Daijing; SIMAYIJIANG, Halimureti; YAN, Jiangwei. Estimation of bloodstain deposition time within a 24-h day-night cycle with rhythmic mRNA based on a machine learning algorithm. **Forensic Science International: Genetics**, p. 102910, 2023. DOI: 10.1016/j.fsigen.2023.102910.

CHEVALLIER, Sylvie; BERTRAND, Dominique; KOHLER, Achim; COURCOUX, Philippe. Application of PLS-DA in multivariate image analysis. **Journal Of Chemometrics,** 2006. p. 221–229. DOI: 10.1002/cem.994.

DA SILVA, Neirivaldo Cavalcante; HONORATO, Ricardo Saldanha; PIMENTEL, Maria Fernanda; GARRIGUES, Salvador; CERVERA, Maria Luisa; DE LA GUARDIA, Miguel. Near Infrared Spectroscopy Detection and Quantification of Herbal Medicines Adulterated with Sibutramine. **Journal of Forensic Sciences**, v. 60, p. 1199–1205, 2015. DOI: 10.1111/1556-4029.12884.

DOS SANTOS, Breno Henrique Dantas. **Avaliação de equipamentos NIR e Raman portáteis e técnicas de aprendizagem de máquina para estimar o tempo de deposição de sangue em pisos para fins forenses**. 2020. Dissertação de Mestrado - Universidade Federal de Pernambuco, Recife, 2020.

DAS, Tanurup; HARSHEY, Abhimanyu; NIGAM, Kriti; YADAV, Vijay K.; SRIVASTAVA, Ankit. Analytical approaches for bloodstain aging by vibrational spectroscopy: Current trends and future perspectives. **Microchemical Journal**, v. 158, 2020. DOI: 10.1016/j.microc.2020.105278.

DE CASTRO, Therese C.; TAYLOR, Michael C.; KIESER, Jules A.; CARR, Debra J.; DUNCAN, W. Systematic investigation of drip stains on apparel fabrics: The effects of prior-laundering, fibre content and fabric structure on final stain appearance. **Forensic Science International**, v. 250, p. 98–109, 2015. DOI: 10.1016/j.forsciint.2015.03.004.

DE CASTRO, Therese; NICKSON, Tania; CARR, Debra; KNOCK, Clare. Interpreting the formation of bloodstains on selected apparel fabrics. **International Journal of Legal Medicine**, v. 127, p. 251–258, 2013. DOI: 10.1007/s00414-012-0717-3.

DE JUAN, Anna; MAEDER, Marcel; HANCEWICZ, Thomas; DUPONCHEL, Ludovic,; TAULER, Roma. **Infrared and Raman spectroscopic imaging**., Wiley-VCH, 2009.

DICKEN, L.; KNOCK, C.; CARR, D. J.; BECKETT, S. The effect of fabric mass per unit area and blood impact velocity on bloodstain morphology. **Forensic Science International**, v. 301, p. 12–27, 2019. a. DOI: 10.1016/j.forsciint.2019.05.001.

DICKEN, L.; KNOCK, C.; CARR, D. J.; BECKETT, S. Investigating bloodstain dynamics at impact on the technical rear of fabric. **Forensic Science International**, v. 301, p. 142–148, 2019. b. DOI: 10.1016/j.forsciint.2019.05.025.

DOTY, Kyle C.; MCLAUGHLIN, Gregory; LEDNEV, Igor K. A Raman “spectroscopic clock” for bloodstain age determination: the first week after deposition. **Analytical and Bioanalytical Chemistry**, v. 408, n. 15, p. 3993–4001, 2016. DOI: 10.1007/s00216-016-9486-z.

DOTY, Kyle C.; MURO, Claire K.; LEDNEV, Igor K. Predicting the time of the crime: Bloodstain aging estimation for up to two years. **Forensic Chemistry**, v. 5, p. 1–7, 2017. DOI: 10.1016/j.forc.2017.05.002.

DURANTE, Caterina; BRO, Rasmus; COCCHI, Marina. A classification tool for N-way array based on SIMCA methodology. **Chemometrics and Intelligent Laboratory Systems**, v. 106, n. 1, p. 73–85, 2011. DOI: 10.1016/j.chemolab.2010.09.004.

EDELMAN, Gerda; MANTI, Vicky; VAN RUTH, Saskia M.; VAN LEEUWEN, Ton; AALDERS, Maurice. Identification and age estimation of blood stains on colored backgrounds by near infrared spectroscopy. **Forensic Science International**, v. 220, n. 1–3, p. 239–244, 2012. a. DOI: 10.1016/j.forsciint.2012.03.009.

EDELMAN, Gerda; MANTI, Vicky; VAN RUTH, Saskia M.; VAN LEEUWEN, Ton; AALDERS, Maurice. Identification and age estimation of blood stains on colored backgrounds by near infrared spectroscopy. **Forensic Science International**, v. 220, n. 1–3, p. 239–244, 2012. b. DOI: 10.1016/j.forsciint.2012.03.009.

EDELMAN, Gerda; VAN LEEUWEN, Ton G.; AALDERS, Maurice C. G. Hyperspectral imaging for the age estimation of blood stains at the crime scene. **Forensic Science International**, v. 223, n. 1–3, p. 72–77, 2012. DOI: 10.1016/j.forsciint.2012.08.003.

EIGENVECTOR. **GLSW**. 2015. Disponível em: http://wiki.eigenvector.com/index.php?title=Glsw. Acesso em: 05 de Março de 2024.

ELKINS, Kelly M. **Introduction to Forensic Chemistry**., Taylor & Francis Group, 2019.

EWING, Andrew V.; KAZARIAN, Sergei G. Infrared spectroscopy and spectroscopic imaging in forensic science. **Analyst**, v. 142, n. 2, p. 257–272, 2017. DOI: 10.1039/c6an02244h.

FABER, Nicolaas (Klaas). Estimating the uncertainty in estimates of root mean square error of prediction: application to determining the size of an adequate test set in multivariate calibration. **Chemometrics and Intelligent Laboratory Systems**, v. 49, p. 79–89, 1999.

FEARN, Tom; RICCIOLI, Cecilia; GARRIDO-VARO, Ana; GUERRERO-GINEL, José Emilio. On the geometry of SNV and MSC. **Chemometrics and Intelligent Laboratory Systems**, v. 96, n. 1, p. 22–26, 2009. DOI: 10.1016/j.chemolab.2008.11.006.

FERREIRA, Márcia M. C. **Quimiometria: Conceitos, Métodos e Aplicações**. Editora da Unicamp, 2015.

FEUDALE, Robert N.; WOODY, Nathaniel A.; TAN, Huwei; MYLES, Anthony J.; BROWN, Steven D.; FERRÉ, Joan. Transfer of multivariate calibration models: a review. **Chemometrics and intelligent laboratory systems**, v. 64, p. 181–192, 2002.

FONSECA, Aline C. S.; PEREIRA, José F. Q.; HONORATO, Ricardo S.; BRO, Rasmus; PIMENTEL, Maria Fernanda. Hierarchical classification models and Handheld NIR spectrometer to human blood stains identification on different floor tiles. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 267, 2022. DOI: 10.1016/j.saa.2021.120533.

FONSECA, Aline Coelho Silva. **Identificação de vestígios de sangue em diferentes substratos empregando equipamento NIR portátil e modelo hierárquico**. 2020. Dissertação de Mestrado - Universidade Federal de Pernambuco (UFPE), Recife, PE, 2020.

FORINA, M.; LANTERI, S.; CASALE, M. **Multivariate calibration**. **Journal of Chromatography A**, 2007. DOI: 10.1016/j.chroma.2007.03.082.

FULGÊNCIO, A. C. C.; RESENDE, Glaucimar Alex Passos; TEIXEIRA, Marden Claret Fontoura; BOTELHO, Bruno Gonçalves; SENA, Marcelo Martins. Screening method for the rapid detection of diethylene glycol in beer based on chemometrics and portable near-infrared spectroscopy. **Food Chemistry**, v. 391, 2022. DOI: 10.1016/j.foodchem.2022.133258.

GARIGLIO, Sara; MALEGORI, Cristina; MENŻYK, Alicja; ZADORA, Grzegorz; VINCENTI, Marco; CASALE, Monica; OLIVERI, Paolo. Determination of time since deposition of bloodstains through NIR and UV–Vis spectroscopy – A critical comparison. **Talanta**, v. 278, 2024. DOI: 10.1016/j.talanta.2024.126444.

GEMPERLINE, Paul. **Practical Guide to Chemometrics**. Boca Raton, FL: CRC Press: Taylor & Francis Group, 2006. DOI: 10.2307/1269627.

GROBÉRIO, Tatiane S.; ZACCA, Jorge J.; BOTELHO, Élvio D.; TALHAVINI, Marcio; BRAGA, Jez W. B. Discrimination and quantification of cocaine and adulterants in seized drug samples by infrared spectroscopy and PLSR. **Forensic Science International**, v. 257, p. 297–306, 2015. DOI: 10.1016/j.forsciint.2015.09.012.

HART, Andrew J.; BARNES, Geoffrey C.; FULLER, Fiona; CORNWELL, Amy M.; GYULA, Judit; MARSH, Nick P. Finding blood in the dark: A comparison of infrared imaging devices for the detection of bloodstains on dark fabrics based on their resolution. **Forensic Science International**, v. 330, 2022. DOI: 10.1016/j.forsciint.2021.111124.

JIAO, Yiping; LI, Zhichao; CHEN, Xisong; FEI, Shumin. Preprocessing methods for near-infrared spectrum calibration. **Journal of Chemometrics**, v. 34, n. 11, 2020. DOI: 10.1002/cem.3306.

KARGER, B.; RAND, S. P.; BRINKMANN, B. Experimental bloodstains on fabric from contact and from droplets. **International Journal of Legal Medicine**, v. 111, p. 17–21, 1998. DOI: 10.1007/s004140050104.

KUENSTNER, J. Todd; NORRIS, Karl H. Spectrophotometry of human hemoglobin in the near infrared region from 1000 to 2500 nm. **Journal of Near Infrared Spectroscopy**, v. 65, p. 59–65, 1994.

KUMAR, Raj; SHARMA, Kajal; SHARMA, Vishal. Bloodstain age estimation through infrared spectroscopy and Chemometric models. **Science and Justice**, v. 60, n. 6, p. 538–546, 2020. a. DOI: 10.1016/j.scijus.2020.07.004.

KUMAR, Raj; SHARMA, Kajal; SHARMA, Vishal. Bloodstain age estimation through infrared spectroscopy and Chemometric models. **Science and Justice**, v. 60, n. 6, p. 538–546, 2020. b. DOI: 10.1016/j.scijus.2020.07.004.

KUMAR, Raj; SHARMA, Vishal. **Chemometrics in forensic science**. **TrAC - Trends in Analytical Chemistry**Elsevier B.V., , 2018. DOI: 10.1016/j.trac.2018.05.010. Acesso em: 23 set. 2019.

LI, Bo; BEVERIDGE, Peter; O'HARE, William T.; ISLAM, Meez. The age estimation of blood stains up to 30 days old using visible wavelength hyperspectral image analysis and linear discriminant analysis. **Science and Justice**, v. 53, n. 3, p. 270–277, 2013. DOI: 10.1016/j.scijus.2013.04.004.

LI, Richard. **Forensic Biology**., CRC Press , 2011.

LIN, Hancheng; ZHANG, Yinming; WANG, Qi; LI, Bing; FAN, Shuanliang; WANG, Zhenyuan. Species identification of bloodstains by ATR-FTIR spectroscopy: the effects of bloodstain age and the deposition environment. **Int J Legal Med**, v. 132, p. 667–674, 2018. DOI: 10.1007/s00414-017-1634-2.

LIN, Hancheng; ZHANG, Yinming; WANG, Qi; LI, Bing; HUANG, Ping; WANG, Zhenyuan. Estimation of the age of human bloodstains under the simulated indoor and outdoor crime

scene conditions by ATR-FTIR spectroscopy. **Scientific Reports**, v. 7, n. 1, p. 1–9, 2017. DOI: 10.1038/s41598-017-13725-1.

MAJDA, Alicja; WIETECHA-POSŁUSZNY, Renata; MENDYS, Agata; WÓJTOWICZ, Anna; ŁYDŻBA-KOPCZYŃSKA, Barbara. Hyperspectral imaging and multivariate analysis in the dried blood spots investigations. **Applied Physics A: Materials Science and Processing**, v. 124, 2018. DOI: 10.1007/s00339-018-1739-6.

MALEGORI, Cristina; ALLADIO, Eugenio; OLIVERI, Paolo; MANIS, Cristina; VINCENTI, Marco; GAROFANO, Paolo; BARNI, Filippo; BERTI, Andrea; DIFAR, Farmacia. Identification of invisible biological traces in forensic evidences by hyperspectral NIR imaging combined with chemometrics. **Talanta**, v. 215, n. March, p. 120911, 2020. DOI: 10.1016/j.talanta.2020.120911.

MANIS, Cristina; MALEGORI, Cristina; ALLADIO, Eugenio; VINCENTI, Marco; GAROFANO, Paolo; BARNI, Filippo; BERTI, Andrea; OLIVERI, Paolo. Non-destructive age estimation of biological fluid stains: An integrated analytical strategy based on near-infrared hyperspectral imaging and multivariate regression. **Talanta**, v. 245, 2022. DOI: 10.1016/j.talanta.2022.123472.

MARCHI, Lorenzo; KRYLOV, Ivan; ROGINSKI, Robert T.; WISE, Barry; DI DONATO, Francesca; NIETO-ORTEGA, Sonia; PEREIRA, José Francielson Q.; BRO, Rasmus. Automatic hierarchical model builder. **Journal of Chemometrics**, v. 36, n. 12, 2022. DOI: 10.1002/cem.3455.

MARIOTTI, Kristiane de Cássia; ORTIZ, Rafael Scorsatto; FERRÃO, Marco Flôres. Hyperspectral imaging in forensic science: An overview of major application areas. **Science and Justice**, v. 63, n. 3, p. 387–395, 2023. DOI: 10.1016/j.scijus.2023.04.003.

MARRONE, April; BALLANTYNE, Jack. Changes in dry state hemoglobin over time do not increase the potential for oxidative DNA damage in dried blood. **PLoS ONE**, v. 4, n. 4, 2009. DOI: 10.1371/journal.pone.0005110.

MC SHINE, S.; SUHLING, K.; BEAVIL, A.; DANIEL, B.; FRASCIONE, N. The applicability of fluorescence lifetime to determine the time since the deposition of biological stains. **Analytical Methods**, v. 9, n. 13, p. 2007–2013, 2017. DOI: 10.1039/c6ay03099h.

MENŻYK, Alicja; MARTYNA, Agnieszka; DAMIN, Alessandro; VINCENTI, Marco; ZADORA, Grzegorz. Breaking with trends in forensic dating: A likelihood ratio-based approach for estimating time elapsed since bloodstains deposition. **Forensic Science International**, v. 349, n. 111763, 2023.

MISTEK, Ewelina; LEDNEV, Igor K. Identification of species' blood by attenuated total reflection (ATR) Fourier transform infrared (FT-IR) spectroscopy. **Analytical and Bioanalytical Chemistry**, v. 407, n. 24, 2015. DOI: 10.1007/s00216-015-8909-6.

MORILLAS, Alvaro Varela; GOOCH, James; FRASCIONE, Nunzianda. Feasibility of a handheld near infrared device for the qualitative analysis of bloodstains. **Talanta**, v. 184, n. February, p. 1–6, 2018. DOI: 10.1016/j.talanta.2018.02.110.

NETTO, Jose Marcelino; HONORATO, Fernanda A.; CELSO, Paulo Gustavo; PIMENTEL, Maria Fernanda. Authenticity of almond flour using handheld near infrared instruments and one class classifiers. **Journal of Food Composition and Analysis**, v. 115, 2023. DOI: 10.1016/j.jfca.2022.104981.

O'BRIEN, Nada A.; HULSE, Charles A.; FRIEDRICH, Donald M.; VAN MILLIGEN, Fred J.; VON GUNTEN, Marc K.; PFEIFER, Frank; SIESLER, Heinz W. **Miniature near-infrared (NIR) spectrometer engine for handheld applications.** *Em*: **Next-Generation Spectroscopic Technologies**, 2012, p. 837404. DOI: 10.1117/12.917983.

OLIVERI, Paolo. **Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues – A tutorial**. **Analytica Chimica Acta**, 2017. DOI: 10.1016/j.aca.2017.05.013.

OLIVERI, Paolo; DOWNEY, Gerard. **Multivariate class modeling for the verification of food-authenticity claims**. **TrAC - Trends in Analytical Chemistry**, 2012. DOI: 10.1016/j.trac.2012.02.005.

OLIVIERI, Alejandro C. **Introduction to Multivariate Calibration**., Springer, 2018. DOI: 10.1007/978-3-319-97097-4.

OTTO, Mathias. **Chemometrics: Statistics and Computer Application in Analytical Chemistry**. 3rd. ed. Leipziger, Germany: Wiley-VCH Verlag GmbH &Co, 2017. DOI: 10.1198/tech.2001.s596.

PASQUINI, Celio. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. **J. Braz. Chem. Soc**, v. 14, n. 2, p. 198–219, 2003.

PEREIRA, José F. Q.; PIMENTEL, Maria Fernanda; HONORATO, Ricardo S.; BRO, Rasmus. Hierarchical method and hyperspectral images for classification of blood stains on colored and printed fabrics. **Chemometrics and Intelligent Laboratory Systems**, v. 210, n. January, p. 104253, 2021. DOI: 10.1016/j.chemolab.2021.104253.

PEREIRA, José F. Q.; SILVA, Carolina S.; VIEIRA, Maria Júlia L.; PIMENTEL, Maria Fernanda; BRAZ, André; HONORATO, Ricardo S. Evaluation and identification of blood stains with handheld NIR spectrometer. **Microchemical Journal**, v. 133, p. 561–566, 2017. DOI: 10.1016/j.microc.2017.04.038.

PEREIRA, José Francielson Q.; PIMENTEL, Maria Fernanda; AMIGO, José Manuel; HONORATO, Ricardo S. Detection and identification of Cannabis sativa L. using near infrared hyperspectral imaging and machine learning methods. A feasibility study. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 237, p. 118385, 2020. DOI: 10.1016/j.saa.2020.118385.

POMERANTSEV, Alexey L.; RODIONOVA, Oxana Ye. Popular decision rules in SIMCA: Critical review. **Journal of Chemometrics**, v. 34, n. 8, p. 1–14, 2020. DOI: 10.1002/cem.3250.

RANI, M. et al. Miniaturized near-infrared (MicroNIR) spectrometer in plastic waste sorting. Materials, v. 12, n. 7, 2019.

RIBAUX, Olivier; BAYLON, Amélie; ROUX, Claude; DELÉMONT, Olivier; LOCK, Eric; ZINGG, Christian; MARGOT, Pierre. Intelligence-led crime scene processing. Part I: Forensic intelligence. **Forensic Science International**, v. 195, n. 1–3, p. 10–16, 2010. DOI: 10.1016/j.forsciint.2009.10.027.

RODRIGUES E BRITO, Lívia; BRAZ, André; SALDANHA HONORATO, Ricardo; PIMENTEL, Maria Fernanda; PASQUINI, Celio. Evaluating the potential of near infrared hyperspectral imaging associated with multivariate data analysis for examining crossing ink lines. **Forensic Science International**, v. 298, p. 169–176, 2019. DOI: 10.1016/j.forsciint.2019.02.043.

RUIZ-PEREZ, Daniel; GUAN, Haibin; MADHIVANAN, Purnima; MATHEE, Kalai; NARASIMHAN, Giri. So you think you can PLS-DA? **BMC Bioinformatics**, v. 21, 2020. DOI: 10.1186/s12859-019-3310-7.

SHARMA, Vishal; KUMAR, Raj. **Trends of chemometrics in bloodstain investigations**. **TrAC - Trends in Analytical Chemistry**, 2018. DOI: 10.1016/j.trac.2018.08.006.

SILVA, Carolina S.; BRAZ, André; PIMENTEL, Maria Fernanda. Vibrational spectroscopy and chemometrics in forensic chemistry: Critical review, current trends and challenges. **Journal of the Brazilian Chemical Society**, v. 30, n. 11, p. 2259–2290, 2019. DOI: 10.21577/0103-5053.20190140.

SILVA, Carolina S.; FERNANDA, Maria; AMIGO, Manuel; GARCÍA-RUIZ, Carmen; ORTEGA-OJEDA, Fernando. Chemometric approaches for document dating : Handling paper variability. **Analytica Chimica Acta**, p. 1–10, 2018. DOI: 10.1016/j.aca.2018.06.031.

SILVA, Carolina S.; PIMENTEL, Maria Fernanda; AMIGO, José Manuel; HONORATO, Ricardo S.; PASQUINI, Celio. **Detecting semen stains on fabrics using near infrared hyperspectral images and multivariate models**. **TrAC - Trends in Analytical Chemistry**Elsevier B.V., , 2017. DOI: 10.1016/j.trac.2017.07.026.

SILVA, Erklaylle G. C.; SILVA, Carolina S.; PIMENTEL, M. Fernanda. Infrared spectroscopy for document dating - Advantages, challenges and limitations. **Talanta Open**, v. 5, 2022. DOI: 10.1016/j.talo.2022.100102.

SILVA, Maurício A. M.; FERREIRA, Marcus H.; BRAGA, Jez W. B.; SENA, Marcelo M. Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy. **Talanta**, v. 89, p. 342–351, 2012. DOI: 10.1016/j.talanta.2011.12.039.

SKOOG, Douglas; HOLLER, James; CROUCH, Stanley. **Principles of Instrumental Analysis**. 7ª ed., 2018. DOI: 10.1515/pac-2015-0305.

SMITH, F. R.; NICLOUX, C.; BRUTIN, D. A new forensic tool to date human blood pools. **Scientific Reports**, v. 10, 2020. DOI: 10.1038/s41598-020-65465-4.

SOBAC, B.; BRUTIN, D. Structural and evaporative evolutions in desiccating sessile drops of blood. **Physical Review E - Statistical, Nonlinear, and Soft Matter Physics**, v. 84, n. 1, 2011. DOI: 10.1103/PhysRevE.84.011603.

SOBAC, B.; BRUTIN, D. Desiccation of a sessile drop of blood: Cracks, folds formation and delamination. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, v. 448, n. 1, p. 34–44, 2014. DOI: 10.1016/j.colsurfa.2014.01.076.

SUN, Da-Wen. **Hyperspectral Imaging for Food Quality Analysis and Control**. Elsevier, 2010.

TAKAMURA, Ayari; WATANABE, Daisuke; SHIMADA, Rintaro; OZAWA, Takeaki. Comprehensive modeling of bloodstain aging by multivariate Raman spectral resolution with kinetics. **Communications Chemistry**, v. 2, n. 1, 2019. DOI: 10.1038/s42004-019-0217-1.

TASTEKIN, Burak; AKCAN, Ramazan; EVRAN, Eylul; TAMER, Ugur; ZENGIN, H. Yagmur; YILDIRIM, Mahmut Serif; BOYACI, Ismail Hakki. Estimation of time since deposition of semen stain on different fabric types using ATR-FTIR spectroscopy and chemometrics. **Forensic Science International**, v. 354, 2024. DOI: 10.1016/j.forsciint.2023.111885.

TAYLOR, Michael C.; LABER, Terry L.; KISH, Paul E.; OWENS, Glynn; OSBORNE, Nikola K. P. The Reliability of Pattern Classification in Bloodstain Pattern Analysis—PART 2: Bloodstain Patterns on Fabric Surfaces. **Journal of Forensic Sciences**, v. 61, n. 6, p. 1461–1466, 2016. DOI: 10.1111/1556-4029.13191.

TSURUGA, Mie; MATSUOKA, Ariki; HACHIMORI, Akira; SUGAWARA, Yoshiaki; SHIKAMA, Keiji. The molecular mechanism of autoxidation for human oxyhemoglobin. Tilting of the distal histidine causes nonequivalent oxidation in the β chain. **Journal of Biological Chemistry**, v. 273, n. 15, p. 8607–8615, 1998. DOI: 10.1074/jbc.273.15.8607.

VIDAL, Maider; AMIGO, José Manuel. Pre-processing of hyperspectral images. Essential steps before image analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 117, p. 138–148, 2012. DOI: 10.1016/j.chemolab.2012.05.009.

VIRKLER, Kelly; LEDNEV, Igor K. Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene. **Forensic Science International**, v. 188, p. 1–17, 2009. DOI: 10.1016/j.forsciint.2009.02.013.

VITALE, Raffaele; COCCHI, Marina; BIANCOLILLO, Alessandra; RUCKEBUSCH, Cyril; MARINI, Federico. Class modelling by Soft Independent Modelling of Class Analogy: why, when, how? A tutorial. **Analytica Chimica Acta**, 2023. DOI: 10.1016/j.aca.2023.341304.

WANG, Fujun; GALLARDO, Vanessa; MICHIELSEN, Stephen; FANG, Tiegang. Fundamental study of porcine drip bloodstains on fabrics: Blood droplet impact and wicking dynamics. **Forensic Science International**, v. 318, 2021. DOI: 10.1016/j.forsciint.2020.110614.

WEYERMANN, Céline; RIBAUX, Olivier. Situating forensic traces in time. **Science and Justice**, v. 52, n. 2, p. 68–75, 2012. DOI: 10.1016/j.scijus.2011.09.003.

WILLIAMS, Elisabeth M. P.; DODDS, Margaret; TAYLOR, Michael C.; LI, Jingyao; MICHIELSEN, Stephen. Impact dynamics of porcine drip bloodstains on fabrics. **Forensic Science International**, v. 262, p. 66–72, 2016. DOI: 10.1016/j.forsciint.2016.02.037.

WILLIAN, Jez; BRAGA E RONEI, B.; POPPI, Jesus. Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo. **Quim. Nova**, v. 27, n. 6, 2004.

WINNEPENNINCKX, Astrid; VERHOEVEN, Elke; VERMEULEN, Steve; BEKAERT, Bram. Evaluation of infrared photography for latent bloodstain visualization and the influence of time. **Forensic Science International**, v. 331, 2022. DOI: 10.1016/j.forsciint.2021.111167.

WISE, Barry M.; MARTENS, Harald; HØY, Martin; BRO, Rasmus; BROCKHOFF, Per B. Calibration Transfer by Generalized Least Squares. **Seventh Scandinavian Symposium on chemometrics (SSC7)**, p. 19–23, 2001.

WOLD, Svante; SJOSTROM, Michael; ERIKSSON, Lennart; SWEDEN˚˚, Sweden˚. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 109–130, 2001.

ZADORA, Grzegorz; MENZYK, Alicja. In the pursuit of the holy grail of forensic science e Spectroscopic studies on the estimation of time since deposition of bloodstains. **Trends in Analytical Chemistry**, v. 105, p. 137–165, 2018. a. DOI: 10.1016/j.trac.2018.04.009.

ZHANG, Linna; ZHANG, Shengzhao; SUN, Meixiu; WANG, Zhennan; LI, Hongxiao; LI, Yingxin; LI, Gang; LIN, Ling. Blood species identification using Near-Infrared diffuse transmitted spectra and PLS-DA method. **Infrared Physics and Technology**, v. 76, p. 587–591, 2016. DOI: 10.1016/j.infrared.2016.04.025.

ZHAO, Yuefeng; HU, Nannan; WANG, Yunuan; LIU, Yonglei; LI, Xiaofei; WANG, Jingjing. The application of near-infrared reflectance hyperspectral imaging for the detection and extraction of bloodstains. **Cluster Computing**, v. 22, p. 8453–8461, 2019. DOI: 10.1007/s10586-018-1869-9.

ZHI, Hui; LUO, Jianrang; LIU, Ruiwen. The Application Progress and Prospect of Miniature Spectrometer in Precision Agriculture. **Converter Magazine**, v. 2021, n. 3, 2020.

# APÊNDICE A – ARTIGO PUBLICADO

Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Anal. Methods*, 2023, 15, 5459

Published on 14 September 2023. Downloaded on 10/30/2023 2:20:36 PM.

## Classification of bloodstains deposited at different times on floor tiles using hierarchical modelling and a handheld NIR spectrometer†

Aline C. S. Fonseca, [a] José F. Q. Pereira, [*b] Ricardo S. Honorato, [c] Rasmus Bro [d] and Maria Fernanda Pimentel [*e]

Bloodstains are commonly encountered at crime scenes, especially on floor tiles, and can be deposited over different periods and intervals. Therefore, it is crucial to develop techniques that can accurately identify bloodstains deposited at different times. This study builds upon a previous investigation and aims to enhance the performance of three distinct hierarchical models (HMs) designed to differentiate and identify stains of human blood (HB), animal blood (AB), and common false positives (CFPs) on nine different types of floor tiles. Soft Independent Modeling Class Analogies (SIMCA), and Partial Least Squares-Discriminant Analysis (PLS-DA) were employed as decision rules in this process. The originally published model was constructed using a training set that included samples with a known time of deposit of six days. This model was then tested to predict samples with various deposition times, including human blood samples aged for 0, 1, 9, 20, 30, and 162 days, as well as animal blood samples aged for 0, 1, 10, 13, 20, 29, 105, and 176 days. To improve the identification of human blood, the models were modified by adding zero-day and one-day-old bloodstains to the original training set. All models showed improvement when fresher samples were included in the training set. The best results were achieved with the hierarchical model that used partial least squares-discriminant analysis as the second decision rule and incorporated one-day-old samples in the training set. This model yielded sensitivity values above 0.92 and specificity values above 0.7 for samples aged between zero and 30 days.

Received 14th July 2023
Accepted 10th September 2023

DOI: 10.1039/d3ay01204b

rsc.li/methods

## Introduction

Blood traces are important pieces of information for elucidating how the events happened in violent crimes. Ideally, the evidence needs to be identified and collected still while at the crime scene, to speed up the action of investigators. Therefore, fast, non-destructive (for future analyses such as DNA testing) and confirmatory methods must be used. Methodologies that use vibrational techniques such as near-infrared are capable of preserving the integrity of the evidence and allow counterproof or complementary analyses to be carried out.[1–4]

Recent studies have shown the usefulness of portable equipment to identify bloodstains at suspected crime scenes.[5–7] Many challenges must be overcome for a practical application, such as the diversity of surfaces where these stains are found,[8–10] the environmental conditions to which they have been exposed,[8,11] and the length of time since the stains were deposited.[12] Different approaches have tested the application of both vibrational and chemometric techniques to identify bloodstains. Most of these studies, however, have used fresh bloodstains to build classification models.[13–17]

In the work developed by Lin *et al.*,[12] samples of blood from humans, mammals (rabbits and dogs), and fowl (chickens and ducks) were collected and analysed using Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy (ATR-FTIR). From these, Principal Component Analysis-*K*-Means Clustering (PCA-K-Means) and partial least squares-discriminant analysis models were built. They were able to distinguish human blood (HB) samples from other species. In addition, they noticed that fresh and old HB samples also showed differences; and as the age of deposit increased, it became more difficult to differentiate human blood samples from samples from other mammals. Their work confirmed that it is possible to identify both fresh and aged human blood

[a]Department of Fundamental Chemistry, Federal University of Pernambuco, Av. Jornalista Aníbal Fernandes, Cidade Universitária, 50.740-560, Recife, Brazil
[b]Federal Rural University of Pernambuco, Serra Talhada Academic Unit, Av. Gregório Ferraz Nogueira, s/n, Serra Talhada, PE, 56909-535, Brazil
[c]Federal Police, Cais do Apolo, 321, 50030-907, Recife, PE, Brazil
[d]Department of Food Science, University of Copenhagen, Rolighedsvej 26, DK-1958 Frederiksberg, Denmark
[e]Department of Chemical Engineering, Federal University of Pernambuco, Av. dos Economistas, Cidade Universitária, s/n, 50.740-590, Recife, PE, Brazil. E-mail: mfernanda.pimentel@ufpe.br

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3ay01204b

# APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

**UNIVERSIDADE FEDERAL DE PERNAMBUCO**
**CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA**
**DEPARTAMENTO DE QUÍMICA FUNDAMENTAL**

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
## (PARA MAIORES DE 18 ANOS OU EMANCIPADOS - Resolução 466/12)

Convidamos o (a) Sr. (a) para participar como voluntário (a) da pesquisa **"Identificação de vestígios sangue humano depositados em diferentes substratos e envelhecidos em tempos diferentes utilizando espectroscopia e imagem no infravermelho e técnicas quimiométricas"**, que está sob a responsabilidade da pesquisadora **Aline Coelho Silva Fonseca** ([redacted] alinecsfonseca@ufpe.br) e está sob a orientação de: Mª Fernanda Pimentel, Telefones para contato: (81)21267233; (81)21267235, e-mail (mfernanda.pimentel@ufpe.br) e coorientação de: José Francielson Queiroz Pereira (josefrancielson.queiroz@gmail.com), Telefones para contato: [redacted].

Este Termo de Consentimento pode conter informações que o/a senhor/a não entenda. Caso haja alguma dúvida, pergunte à pessoa que está lhe entrevistando para que o/a senhor/a esteja bem esclarecido (a) sobre sua participação na pesquisa. Após ser esclarecido (a) sobre as informações a seguir, caso aceite em fazer parte do estudo, rubrique as folhas e assine ao final deste documento, que está em duas vias. Uma delas é sua e a outra é do pesquisador responsável. Em caso de recusa o (a) Sr. (a) não será penalizado (a) de forma alguma. Também garantimos que o (a) Senhor (a) tem o direito de retirar o consentimento da sua participação em qualquer fase da pesquisa, sem qualquer penalidade.

**INFORMAÇÕES SOBRE A PESQUISA**:

Essa pesquisa tem como objetivo desenvolver métodos confiáveis, rápidos e não destrutivos para identificação e datação de vestígios de fluidos corporais em cenas de crimes utilizando imagens químicas das manchas dos fluidos doados em diferentes tecidos. Amostras de material biológico (sangue) serão depositadas em diferentes superfícies para identificação. A coleta dos materiais será realizada:

a. A coleta será realizada utilizando material descartável em ambiente limpo e apropriado. A coleta será realizada pela enfermeira participante da equipe.
b. O período de participação do voluntário na pesquisa consiste em apenas um dia de visita para a doação do material.
c. Serão colhidos aproximadamente 0,1 ml de sangue (equivalente a duas gotas) por punção na falange distal do 4º dedo da mão direita.

**RISCOS:** O risco associado ao procedimento de coleta é um leve desconforto no local puncionado. Será pressionado um algodão estéril no local da punção, e para aqueles voluntários sensíveis a dor, será utilizada compressa com gelo para minimizá-la.

**BENEFÍCIOS diretos/indiretos para os voluntários:** Não há benefícios diretos à saúde dos voluntários. Os benefícios dessa pesquisa estão associados a questões de segurança pública.

As informações desta pesquisa serão confidenciais e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos voluntários, a não ser entre os

responsáveis pelo estudo, sendo assegurado o sigilo sobre a sua participação. Os dados coletados nesta pesquisa (material biológico), ficarão armazenados em computador pessoal, sob a responsabilidade do pesquisador, no endereço acima informado, pelo período de mínimo 5 anos.

O (a) senhor (a) não pagará nada para participar desta pesquisa. Se houver necessidade, as despesas para a sua participação serão assumidos pelos pesquisadores (ressarcimento de transporte e alimentação). Fica também garantida indenização em casos de danos, comprovadamente decorrentes da participação na pesquisa, conforme decisão judicial ou extra-judicial.

Em caso de dúvidas relacionadas aos aspectos éticos deste estudo, você poderá consultar o Comitê de Ética em Pesquisa Envolvendo Seres Humanos da UFPE no endereço: (Avenida da Engenharia s/n – 1º Andar, sala 4 - Cidade Universitária, Recife-PE, CEP: 50740-600, Tel.: (81) 2126.8588 – e-mail: cepccs@ufpe.br).

_____
(assinatura do pesquisador)

**CONSENTIMENTO DA PARTICIPAÇÃO DA PESSOA COMO VOLUNTÁRIO (A)**

Eu, _____, CPF _____, abaixo assinado, após a leitura (ou a escuta da leitura) deste documento e de ter tido a oportunidade de conversar e ter esclarecido as minhas dúvidas com o pesquisador responsável, concordo em participar do estudo **"Identificação de vestígios sangue humano depositados em diferentes substratos e envelhecidos em tempos diferentes utilizando espectroscopia e imagem no infravermelho e técnicas quimiométricas"**, como voluntário(a).

Fui devidamente informado (a) e esclarecido (a) pelo(a) pesquisador (a) sobre a pesquisa, os procedimentos nela envolvidos, assim como os possíveis riscos e benefícios decorrentes de minha participação. Foi-me garantido que posso retirar o meu consentimento a qualquer momento, sem que isto leve a qualquer penalidade (ou interrupção de meu acompanhamento/assistência/tratamento).

Local e data: _____

Assinatura do participante: _____

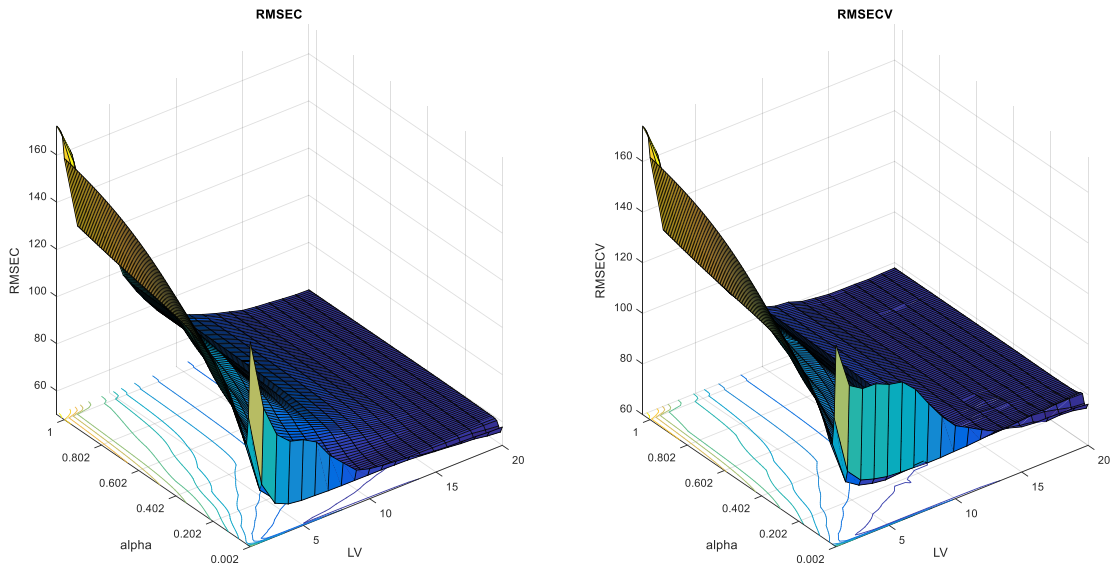
Impressão Digital (Opcional)

**Presenciamos a solicitação de consentimento, esclarecimentos sobre a pesquisa e o aceite do voluntário em participar.** (02 testemunhas não ligadas à equipe de pesquisadores):

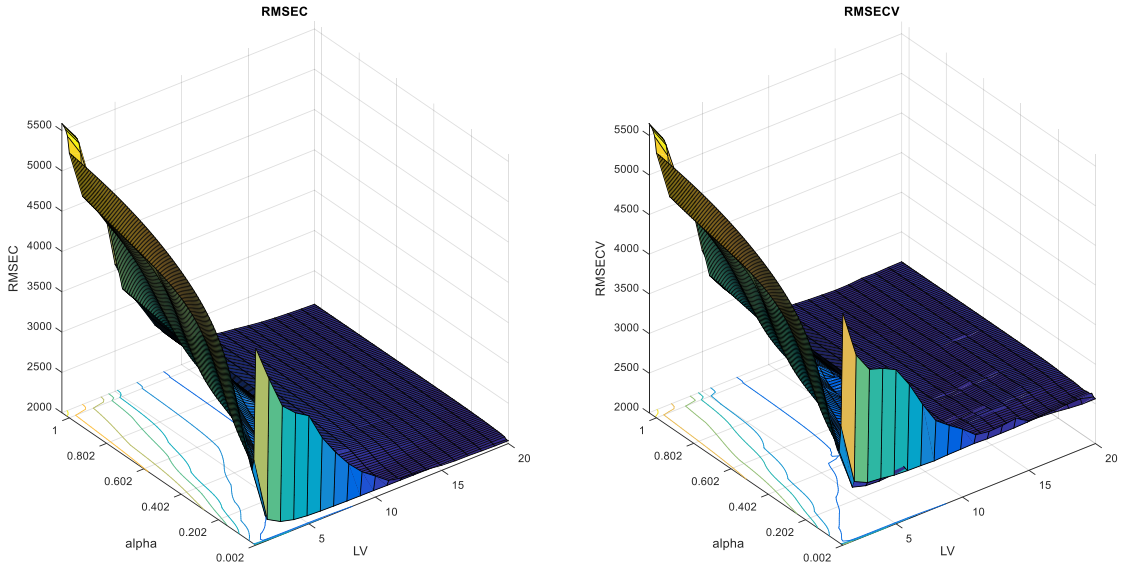
| | |
|---|---|
| Nome: | Nome: |
| Assinatura: | Assinatura: |

# APÊNDICE C – GRÁFICOS DE CONTORNO DAS SUPERFÍCIES DE RESPOSTAS PARA OS DADOS COLETADOS UTILIZANDO O NEOSPECTRA

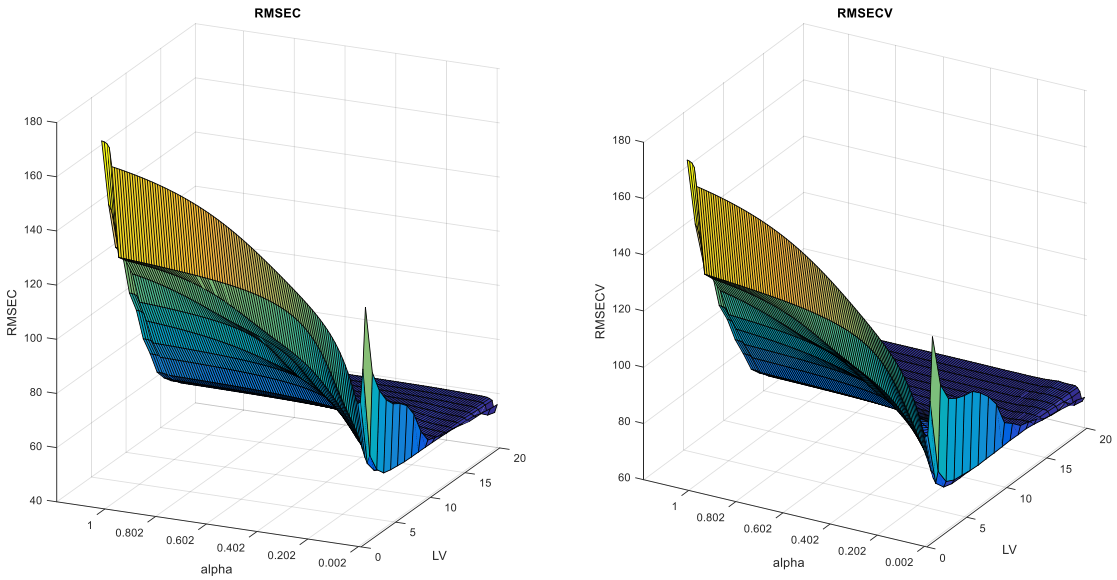
1. Gráfico da Superfície de Resposta para o Conjunto de Treinamento 1 (30 min a 10h):



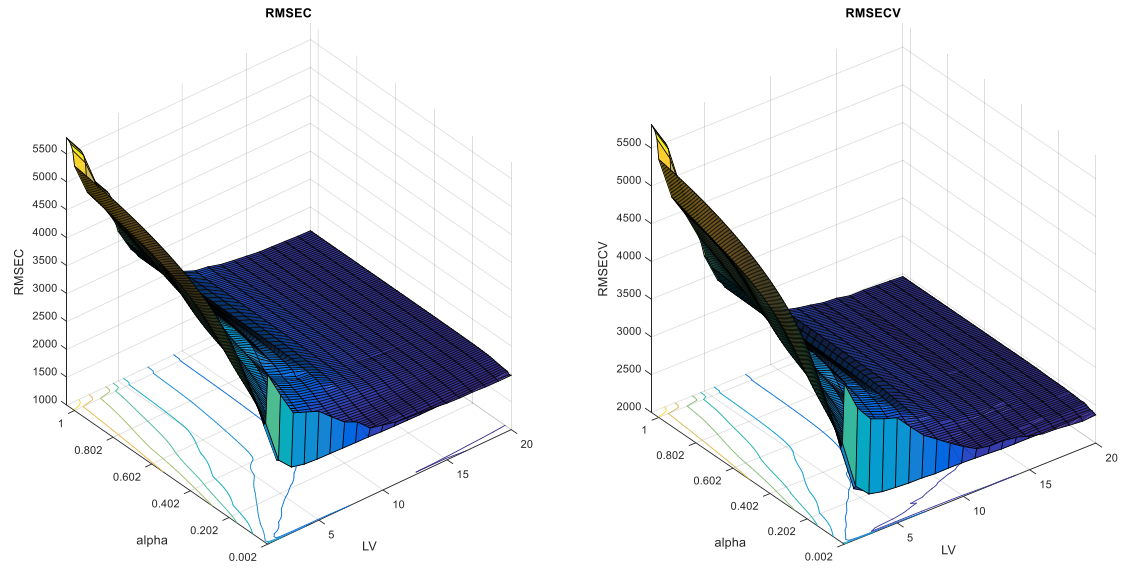

2. Gráfico da Superfície de Resposta para o Conjunto de Treinamento 1 (24h a 15 dias).



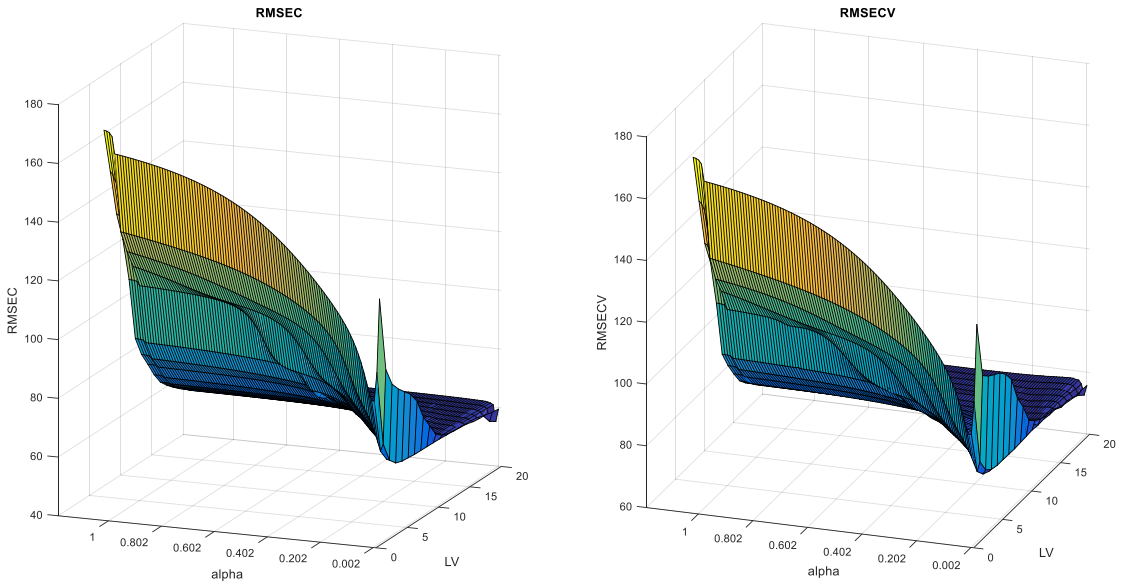
3. Gráfico da Superfície de Resposta para o Conjunto de Treinamento 2 (30 min a 10h):



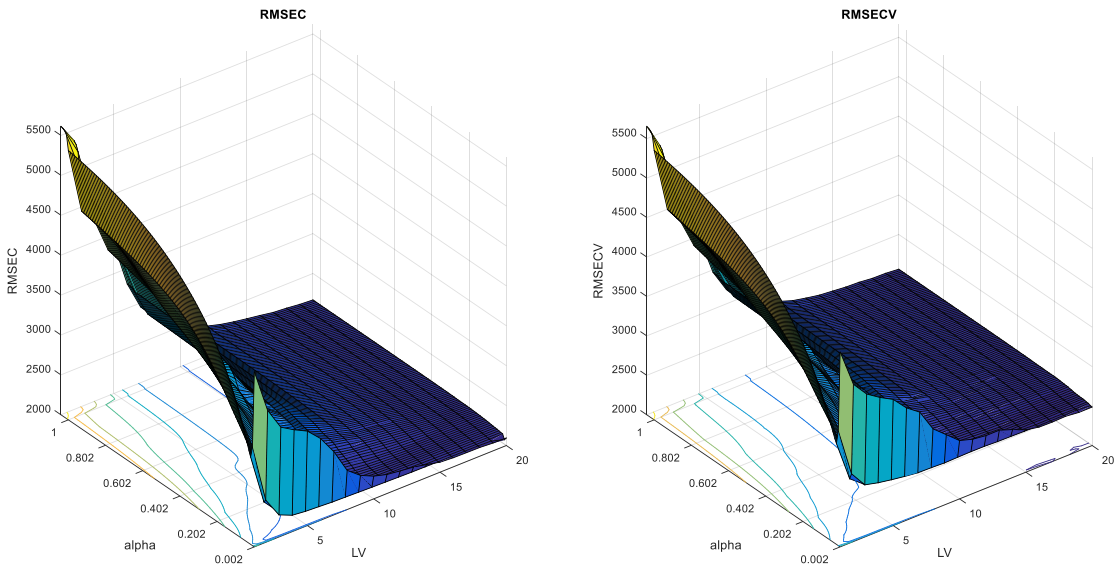

4. Gráfico da Superfície de Resposta para o Conjunto de Treinamento 2 (24h a 15 dias):



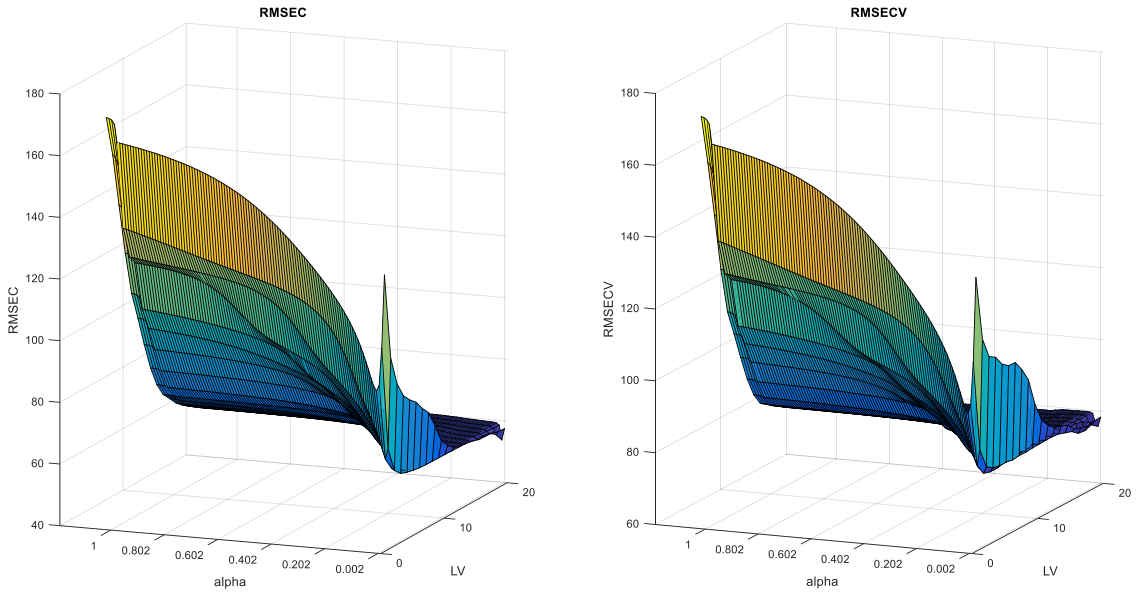
5. Gráfico da Superfície de Resposta para o Conjunto de Treinamento 3 (30 min a 10h):



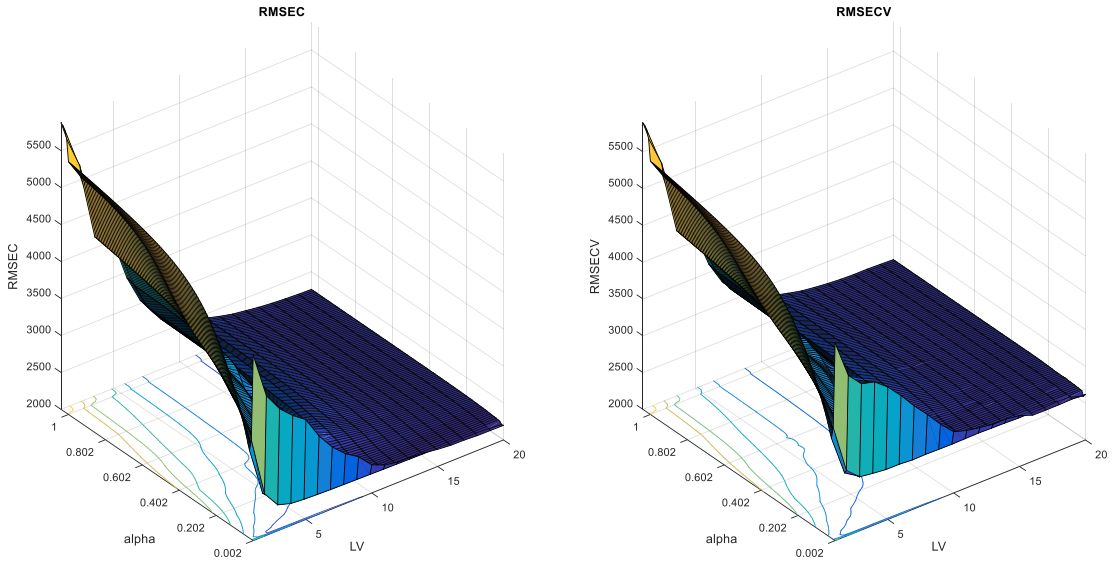

6. Gráfico da Superfície de Resposta para o Conjunto de Treinamento 3 (24h a 15 dias):



7. Gráfico da Superfície de Resposta para o Conjunto de Treinamento 4 (30 min a 10h):




8. Gráfico da Superfície de Resposta para o Conjunto de Treinamento 4 (24h a 15 dias):

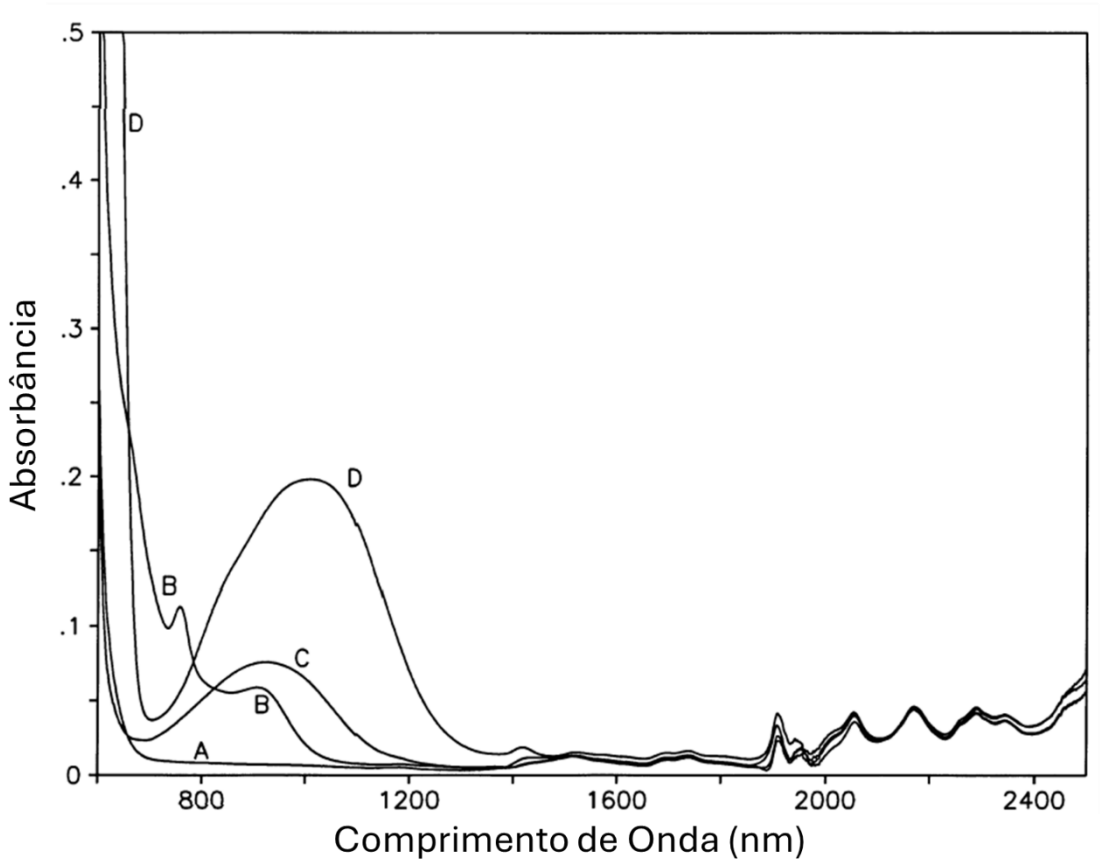


## APÊNDICE D – NOTA DE IMPRENSA

Aline Coelho Silva Fonseca, sob orientação da Dra. Maria Fernanda Pimentel e coorientação do Dr. José Francielson Queiroz Pereira e desenvolvido sob o fomento de bolsa de doutorado fornecida pela agência CAPES, apresenta um estudo que explora a utilização de espectroscopia no infravermelho próximo (NIR) para monitorar o envelhecimento de vestígios de sangue humano depositados em diferentes substratos, como ladrilhos de piso e tecidos. Este trabalho é pioneiro no uso de equipamentos portáteis de NIR e imagens hiperespectrais para essa finalidade, proporcionando novas possibilidades no campo da investigação forense.

O estudo utiliza ferramentas avançadas de quimiometria, como a Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA) e a Regressão de Mínimos Quadrados Parciais (PLSR), para desenvolver modelos hierárquicos que possibilitam a estimativa da idade das manchas de sangue. Os resultados são promissores: foi possível determinar a idade de manchas de sangue depositadas em pisos com um erro médio de predição (RMSEP) de cerca de 1 hora para intervalos de 2 a 10 horas, e de aproximadamente 2 dias para intervalos de 1 a 15 dias. Para as manchas depositadas em tecidos, foi possível estimar o tempo de deposição com RMSEP de aproximadamente 2 dias. Este avanço abre portas para a utilização de tecnologias portáteis na análise de cenas de crime, melhorando a precisão na determinação do tempo de deposição de vestígios biológicos, um fator crucial em investigações criminais.

**ANEXO A – ESPECTRO DE ABSORÇÃO DAS DIFERENTES ESPÉCIES DE HEMOGLOBINA APÓS A SUBTRAÇÃO DOS SUBSTRATOS.**




**Fonte:** (KUENSTNER et al., 1994)

A – Carboxiemoglobina;

B – Desoxihemoglobina;

C – Oxiemoglobina;

D – Metemoglobina.