



UNIVERSIDADE DE BRASÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS E TECNOLOGIAS EM SAÚDE
FACULDADE DE CEILÂNDIA

**QUIMIOMETRIA APLICADA À DISCRIMINAÇÃO DE VEÍCULOS PELA
ANÁLISE DE FRAGMENTOS DE TINTAS AUTOMOTIVAS POR ATR-FTIR**

JULIANA MELO DUARTE

Brasília, 14 de julho de 2021

JULIANA MELO DUARTE

**QUIMIOMETRIA APLICADA À DISCRIMINAÇÃO DE VEÍCULOS PELA
ANÁLISE DE FRAGMENTOS DE TINTAS AUTOMOTIVAS POR ATR-FTIR**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências e Tecnologias em Saúde da Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Ciências e Tecnologias em Saúde. Área de concentração: Mecanismos Básicos e Processos Biológicos em Saúde, linha de pesquisa: Nanobiotecnologia Aplicada à Saúde, temática: Nanomateriais Aplicados à Saúde e Ambiente.

Orientador: Prof. Dr. Marcelo Henrique Sousa

Coorientador: Dr. Juliano de Andrade Gomes

Brasília, 2021

Melo Duarte, Juliana

MJ94q Quimiometria aplicada à discriminação de veículos pela análise de fragmentos de tintas automotivas por ATR-FTIR / Juliana Melo Duarte; orientador Marcelo Henrique Sousa; coorientador Juliano de Andrade Gomes. -- Brasília, 2021.
76 p.

Dissertação (Mestrado - Mestrado em Ciências e Tecnologias em Saúde) -- Universidade de Brasília, 2021.

1. tinta automotiva. 2. espectroscopia ATR-FTIR. 3. quimiometria. 4. PLS-DA. 5. Ciência Forense. I. Henrique Sousa, Marcelo, orient. II. de Andrade Gomes, Juliano, coorient. III. Título.

JULIANA MELO DUARTE

**QUIMIOMETRIA APLICADA À DISCRIMINAÇÃO DE VEÍCULOS PELA
ANÁLISE DE FRAGMENTOS DE TINTAS AUTOMOTIVAS POR ATR-FTIR**

Data da defesa: 14 de julho de 2021

BANCA EXAMINADORA:

Prof. Dr. MARCELO HENRIQUE SOUSA (Orientador e Presidente)

Universidade de Brasília/Faculdade de Ceilândia - UnB/FCE

Programa de Pós-Graduação em Ciências e Tecnologias em Saúde

Profa. Dra. MARIANA RAMOS DE ALMEIDA. (1º Membro)

Universidade Federal de Minas Gerais - UFMG

Profa. Dra. SANDRA MARIA DA LUZ (2º Membro)

Universidade de Brasília/Faculdade do Gama - UnB/FGA

Prof. Dr. JULIANO ALEXANDRE CHAKER (Suplente)

Universidade de Brasília/Faculdade de Ceilândia - UnB/FCE

Aos meus espelhos, Élio e Lyz.
Ao Antônio Luiz, companheiro de vida, de ideias e de ideais.
Ao Ravi e à Sabrina, os grandes motivos da minha caminhada.

AGRADECIMENTOS

Ao Orientador Marcelo, pela orientação no formato exato. Perfeita.

Ao Juliano, pela linha de pesquisa, pelo apoio enorme e constante e por ter mentido que seria fácil.

Ao Professor Jez, por partilhar a sua genialidade de forma irrestrita e não parcimoniosa. É coorientador das análises deste trabalho, de forma não oficial por mera intempestividade de prazo hábil.

À Nádia. Depois eu mando por áudio, pois é muita coisa para digitar. Brincadeira à parte, em especial, pela inteligência que emana, intelectual e emocional. Cuide bem do seu brilhante futuro.

À Eliane, pela verdadeira irmandade acadêmica. Aos queridos Thayná, Wesley, Luciano, Eliude e Mavignier, pelo coleguismo e amizade nas disciplinas.

A Nagao, Jorge, Vinícius e Almeida, pelo apoio à pesquisa, recebido enquanto chefes/diretor, quando estive na SDT. A Bárbara e Juliano, pelo apoio atual.

Aos colegas da SDT e colegas agentes que ajudaram nas coletas dos fragmentos.

Aos colegas da SPAL pela troca constante de conhecimentos, em especial ao brilhante Diego. Ao Luciano, por toda a diferença científica já promovida no Instituto, abrindo os caminhos, e também pelo nitrogênio. À Luísa, pela leitura desta dissertação.

À Fundação de Peritos em Criminalística Ilaraine Acácio Arce- FPCIAA, em nome do Fábio, pelo apoio em forma de projeto de PIBIC, pelos equipamentos, pelos livros e pela disponibilidade. Destaque para o espectrômetro Bruker Lumos, utilizado nesta pesquisa.

A Souza, Marcelo e demais diretores, pela mentalidade aberta, o que faz toda a diferença.

À Deliane e à Dora, por terem muitas vezes cuidado dos meus mais preciosos bens enquanto eu estudava.

A Élio, Lyz, Eleniza e Sílvio, por terem me construído. A Agnaldo e Lênia pelo fundamental inglês. A Luciano, Vinícius e Marina pelos exemplos de persistência apaixonada, de aprendizado através da reinvenção e de coerência científica cartesiana.

Ao Antônio por ter me oferecido a pílula vermelha. Aos Professores Ravi e Sabrina, pelos ensinamentos diários.

A Tatiana e Flávia, pela minha complementariedade.

A todos os familiares e amigos, pelos preciosos momentos de descontração.

“Mestre é aquele que tomou consciência de que nada sabe. Talvez, um dia, decida fazer o doutorado, na esperança de que ainda aprenda alguma coisa.”

Juliana Duarte, 2021.

SUMÁRIO

LISTA DE FIGURAS	x
LISTA DE ABREVIATURAS E SIGLAS	xiv
LISTA DE SÍMBOLOS	xvi
RESUMO	xvii
ABSTRACT	xviii
1 INTRODUÇÃO	19
1.1 COMPOSIÇÃO DAS TINTAS AUTOMOTIVAS.....	21
1.2 ANÁLISE DE TINTAS AUTOMOTIVAS NOS ÚLTIMOS DOZE ANOS	23
1.3 REFERENCIAL TEÓRICO.....	27
1.3.1 Espectroscopia de absorção na região do infravermelho	27
1.3.2 Pré-processamentos dos dados	29
1.3.3 Análise de Componentes Principais (PCA)	30
1.3.4 Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)	31
2 OBJETIVOS	32
3 METODOLOGIA	33
3.1 COLETA DAS AMOSTRAS E PREPARO	33
3.2 TRATAMENTO DOS DADOS	36
3.2.1 Pré-processamentos e PCA	36
3.2.2 PLS-DA	37
3.2.2.1 Construção do modelo	39
3.2.2.2 Validação do modelo	43
3.2.2.3 Figuras de mérito	46
4 RESULTADOS E DISCUSSÃO	47
4.1 RESULTADO DA ANÁLISE DE PCA	50
4.2 RESULTADO DA ANÁLISE DE PLS-DA	52
4.2.1 Validação do modelo	55
4.2.2 Figuras de mérito	59
4.3 AVALIAÇÃO GERAL	62
4.4 PERSPECTIVAS FUTURAS	64

5 CONCLUSÃO	66
REFERÊNCIAS	67

LISTA DE FIGURAS

Figura 1- Principais camadas de tinta automotiva, a partir da sua representação, vista em corte.	22
Figura 2- Estudo bibliométrico realizado no artigo: “Automotive paint analysis: How far has science advanced in the last ten years?”, o que possibilita a visualização da relação entre as publicações e citações entre os principais autores do tema.	24
Figura 3- Comprimentos de ondas aproximados de cada tipo de radiação eletromagnética. ...	27
Figura 4- Exemplo de lâmina de vidro preparada com 19 amostras de tintas automotivas, aderidas por fita dupla face e devidamente identificadas, preparadas para as análises com espectroscopia de infravermelho.....	35
Figura 5- a, b, c- Ilustram o equipamento LUMOS® Bruker Optics, com uma lâmina de amostras de tintas automotivas em sua mesa de trabalho (elipses), durante a aquisição dos espectros infravermelho. A imagem ampliada da amostra em análise, pelo uso da microscopia, é vista no monitor de trabalho (seta).....	36
Figura 6- Gráfico de Hotelling T2 versus o resíduo Q da plotagem inicial dos dados, quando da construção do modelo. A área circundada pela linha vermelha destaca os espectros que tiveram baixos valores tanto de T2 quanto Q, e, portanto, bem adaptados ao modelo. Os demais espectros são considerados outliers.....	39
Figura 7- Análise PLS-DA da classe Volkswagen Cristal. A seta destaca a linha de discriminação entre o grupo de espectros classificados como Volkswagen Cristal (espectros posicionados acima da linha) e o grupo de espectros classificados como não pertencentes à classe Volkswagen Cristal (posicionados abaixo da linha). Os espectros Volkswagen Cristal posicionados acima da linha são verdadeiros positivos. Os posicionados abaixo são falsos negativos. Os espectros de outras classes posicionados acima da linha são falsos positivos e as abaixo da linha são verdadeiros negativos.	40
Figura 8- Sequência para a criação do modelo de calibração definitivo PLS-DA. Após a aquisição dos espectros, a primeira identificação de outliers foi feita de forma visual, pela inspeção simultânea dos cinco espectros da mesma amostra. As classes principais foram selecionadas conforme o número de amostras (classes mais numerosas). Para a análise PCA, foram utilizadas todas as classes. Para a construção do modelo, foi utilizado o conjunto de calibração, pertencente às classes principais. Após a seleção do número de variáveis latentes (aqui denominadas VLs), a identificação e exclusão de outliers foi feita baseado em Hotelling T2 e Q residual. Após a exclusão dos outliers, foi feito o recálculo do modelo, gerando-se, dessa forma, o modelo PLS-DA otimizado.	41

Figura 9- Divisão das classes de amostras em grupos de calibração, de validação e de teste externo a partir das 24 classes de tintas automotivas, definidas previamente conforme as marcas e tonalidades de cor dos veículos.	42
Figura 10- Critérios para a definição da classe da amostra (definida pela marca e tonalidade de cor do veículo), conforme a classificação de suas cinco replicatas. Caso três ou mais das cinco replicatas tenham sido classificadas para uma determinada classe, então a amostra é positiva para aquela classe. Nas demais situações, a sua classificação é dada como inconclusiva).	43
Figura 11- Sequência para a validação do modelo PLS-DA de análise de tintas automotivas, usando-se os grupos de validação e de teste externo. Após a predição das amostras (em cada uma das etapas ilustradas), foram calculadas as figuras de mérito, para a avaliação da qualidade do modelo construído.	44
Figura 12- Avaliação visual das cinco replicatas de espectros ATR/FTIR das amostras de tintas automotivas (a) WhiFI16aM: replicata WhiFI16aM ii (de cor laranja) e replicata WhiFI16aM v (de cor azul escura) apresentam comportamento anômalo, quando comparadas às demais replicatas da mesma amostra e (b) WhiMB1aM: replicata WhiMB1aM iv (de cor amarela) apresenta comportamento anômalo, quando comparada às demais replicatas da mesma amostra.	48
Figura 13– a) Espectros brutos ATR-FTIR de uma amostra de cada uma das 24 classes das tintas automotivas, b) espectros após os pré-processamentos SNV (standard normal variate) e transformação de transmitância para absorbância e remoção das faixas não úteis dos espectros.	49
Figura 14 – Porcentagens de variância explicada por cada uma das componentes principais na análise de PCA.	50
Figura 15- Plotagem dos escores em sistema bidimensional da análise de componentes principais de tintas automotivas, com duas componentes principais.	51
Figura 16- Plotagem dos escores em sistema tridimensional de análise das componentes principais de tintas automotivas, com três componentes principais.	51
Figura 17 – Raiz quadrada do erro médio quadrático de validação para a determinação do número de variáveis latentes a serem utilizadas no modelo PLS-DA, das classes principais. Os erros tendem a se estabilizar após 14 variáveis latentes.	52
Figura 18 – Calibração do modelo PLS-DA: Hotelling T2 versus o resíduo Q após eliminação de outliers. Análise de discriminação de espectros, para cada uma das classes principais na construção modelo PLS-DA: (a): Fiat Banchisa, (b): Ford Artico, (c): General Motors, (d): Honda Taffeta, (e): Hyundai Creamy, (f): Renault Glacier, (g): Citroen Banquise, (h):	

Volkswagen Cristal. Os círculos pontilhados destacam a amostra VW25a, que resultou como falso negativo para a própria classe e falso positivo para a classe Renault Glacier. 54

Figura 19 – Etapa de validação do modelo PLS-DA: Hotelling T2 versus o resíduo Q após eliminação de outliers. Análise de discriminação de espectros, para cada uma das classes principais na validação do modelo PLS-DA: (a): Fiat Banchisa, (b): Ford Artico, (c): General Motors, (d): Honda Taffeta, (e): Hyundai Creamy, (f): Renault Glacier, (g): Citroen Banquise, (h): Volkswagen Cristal. 56

Figura 20 - Gráfico de Hotelling T2 versus o resíduo Q da predição do grupo de teste externo pelo modelo PLS-DA. O espectros com valores elevados de T2 e Q são considerados outliers (o comportamento esperado para esta etapa), e os com baixos valores, podem estar erroneamente classificados como bem adaptados ao modelo. 57

Figura 21 – Etapa de predição dos espectros do grupo teste externo, no modelo PLS-DA: Hotelling T2 versus o resíduo Q após eliminação de outliers. Análise de discriminação de espectros do grupo teste externo para cada uma das classes principais, no modelo PLS-DA: (a): Fiat Banchisa, (b): Ford Ártico, (c): General Motors, (d): Honda Taffeta, (e): Hyundai Creamy, (f): Renault Glacier, (g): Citroen Banquise, (h): Volkswagen Cristal. As setas evidenciam as linhas de discriminação entre os espectros pertencentes ou não à respectiva classe e as elipses destacam os espectros posicionados acima das linhas de discriminação, e, portanto, passíveis de análise. Os demais tiveram o comportamento esperado. 58

LISTA DE TABELAS E QUADRO

Tabela 1- Relação e número de amostras de tintas automotivas utilizadas, relacionadas conforme as respectivas classes, definidas pela marca e tonalidade de cor dos fabricantes. ...	34
Tabela 2- Relação e número das amostras nomeadas conforme as respectivas classes de tintas automotivas, utilizadas nas análises, após separadas dentre os grupos de treinamento, de validação e de teste.	38
Tabela 3 – Relação e número de amostras das classes principais (aquelas mais numerosas, contendo cinco ou mais amostras) – conforme marcas e tonalidades de cores dos fabricantes (classes), divididas dentre os conjuntos de calibração e de validação.....	42
Tabela 4 – Equações matemáticas para se calcular as figuras de mérito do modelo e seus grupos de aplicabilidade. FP = número de falsos positivos; FN = número de falsos negativos; VP = número de verdadeiros positivos; VN = número de verdadeiros negativos, N = número de amostras do conjunto analisado e I = número de amostras testadas.....	46
Tabela 5 – Figuras de mérito da calibração. NA = número de amostras; FP = número de falsos positivos; FN = número de falsos negativos; VP = número de verdadeiros positivos; VN = número de verdadeiros negativos; TFP = taxa de falsos positivos; TFN = taxa de falsos negativos; TST = taxa de sensibilidade; TSP = taxa de especificidade; TEF = taxa de eficiência. As taxas estão expressas em porcentagens.	60
Tabela 6 - Figuras de mérito da validação. NA = número de amostras; FP = número de falsos positivos; FN = número de falsos negativos; VP = número de verdadeiros positivos; VN = número de verdadeiros negativos; TFP = taxa de falsos positivos; TFN = taxa de falsos negativos; TST = taxa de sensibilidade; TSP = taxa de especificidade; TEF = taxa de eficiência; TCC = taxa de classificação correta. As taxas estão expressas em porcentagens.	61
Tabela 7 – Figuras de mérito do teste externo. NA = número de amostras; FP = número de falsos positivos; TCC = taxa de classificação correta, expressa em porcentagem.	61

LISTA DE ABREVIATURAS E SIGLAS

A – Absorbância

ATR – Refletância total atenuada (*attenuated total reflection*)

B – Campo magnético

CO₂ – Dióxido de carbono

DF – Distrito Federal

E – Campo elétrico

EPG – *European Paint Group*

FBI – *Federal Bureau of Investigation*

FDA - Análise discriminante de Fisher (*Fisher discriminant analysis*)

FN – Falso negativo

FP – Falso positivo

FTIR – Infravermelho com transformada de Fourier (*Fourier transform infrared spectroscopy*)

I – Intensidade de radiação

IR – Infravermelho (*infrared*)

LD – Limiar de discriminação

MLPNN – Rede neural perceptron multicamadas (*multi-layer perceptron neural network*)

MLRM – Espectroscopia micro-Raman (*micro-laser Raman spectroscopy*)

NA – Número de amostras

NIV – Número de identificação veicular

OH – Hidroxila

PC1 – Primeira componente principal

PC2 – Segunda componente principal

PC3 – Terceira componente principal

PCA – Análise por componentes principais (*principal component analysis*)

PCDF – Polícia Civil Do Distrito Federal

PDQ – Paint Data Query

PLS-DA – Análise discriminante por mínimos quadrados parciais (*partial least squared – discriminant analysis*)

RMSECV – Raiz quadrada do erro médio quadrático de validação cruzada (*root mean square of cross validation*)

SNV – Padronização normal de sinal (*standard normal variate*)

SVM – Máquina de vetores de suporte (*support vector machine*)

T – Transmitância

TCC – Taxa de classificação correta

TEF – Taxa de eficiência

TFN – Taxa de falso negativo

TFP – Taxa de falso positivo

TINDEF – Taxa de resultados indefinidos

TSP – Taxa de especificidade

TST – Taxa de sensibilidade

UV – Ultravioleta

VIS – Luz visível

VL – Variável latente

VN – Verdadeiro negativo

VP – Verdadeiro positivo

LISTA DE SÍMBOLOS

e_i - Decomposição residual para i

y_i - Valor de classe real atribuído a uma amostra

\hat{y}_i - Valor estimado obtido pelo modelo

ΔE – Variação entre dois níveis energéticos

c - Velocidade da luz

cm^{-1} – Número de onda

E_{rad} – Energia da radiação

ν – Frequência

h – Constante de Planck

I – Intensidade de radiação que atravessa a amostra

I_0 – Intensidade de radiação incidente sobre a amostra

λ – Comprimento de onda

Λ - Matriz diagonal de autovalores

RESUMO

Introdução: A conduta de evasão do local após um acidente de trânsito (popularmente, bater e fugir) é punível no Brasil com detenção de seis meses a um ano ou multa. Ainda assim, em média, duas a três ocorrências são registradas por dia no Distrito Federal. As possíveis consequências são: o agravamento do quadro clínico dos feridos, maior carga para o sistema hospitalar, e, muitas vezes, a morte de vítimas. A identificação do veículo evasor é um desafio para as equipes policiais. **Objetivo:** A partir de amostras de tintas automotivas (que muitas vezes são encontradas por peritos no local do acidente), pretende-se construir um modelo matemático que seja capaz de identificar a marca e a tonalidade de cor do veículo evasor. **Metodologia:** Foram utilizadas 143 amostras de cor branca, previamente coletadas em locais de acidentes de trânsito, por peritos criminais e agentes da Polícia Civil do Distrito Federal (PCDF). As superfícies das amostras foram analisadas por espectroscopia de infravermelho com transformada de Fourier e refletância total atenuada (ATR-FTIR), com microscopia acoplada. Com os dados dos espectros, foi feita uma análise exploratória empregando a análise de componentes principais (PCA) e análise discriminante por mínimos quadrados parciais (PLS-DA). Para a análise discriminante, as amostras foram divididas em três grupos: calibração, validação e teste externo. Com os resultados obtidos, foram calculadas as figuras de mérito, para avaliação da qualidade do modelo. **Resultados:** Para o grupo de calibração da análise discriminante, as taxas de sensibilidade, especificidade e eficiência foram de, respectivamente, 98,9%, 98,4% e 98,6%. Para o grupo de validação, foram todas de 100%. Foram obtidas 100% de classificações corretas das amostras do grupo de validação interna e 79,1% do grupo de teste externo. **Conclusões:** O método desenvolvido apresenta potencialidade para aplicações forenses, uma vez que emprega uma técnica rápida, limpa e não destrutiva e o modelo matemático apresentou desempenho adequado. O microscópio acoplado é importante para se evitar regiões avariadas das amostras. Ter-se limitado a diferença de idade das amostras em 10 anos (devido ao desgaste natural) se mostrou efetivo. A taxa de classificação correta do grupo de teste externo (de 79,1%), como se refere a classes sub representadas numericamente, pode ser potencialmente aprimorada a qualquer momento, pois o modelo permite a inserção de novas amostras. Foi construído, portanto, um modelo promissor para a aplicação forense.

Palavras-Chave: tinta automotiva, espectroscopia ATR-FTIR, quimiometria, PCA, PLS-DA, Ciência Forense.

ABSTRACT

Introduction: Getting away from a traffic accident scene (hit and run) is punishable with six to twelve months of detention or a fine, in Brazil. Two to three incidents are recorded daily in Brazilian Federal District. The consequences are: more serious injuries of the victims, health system overload and victim's death. The vehicle identification is a challenge for law enforcement agents. **Objective:** We intent to build a mathematical model that is able to identify the make and color shade of the hit-and-run vehicle, by using automotive paint samples (which are often found by the forensic experts at the accident site). **Methods:** 143 white samples were collected at traffic accident scenes, by forensic staff of the Brazilian Federal District Civil Police (PCDF). The samples were *in situ* analyzed by using Fourier transform infrared spectroscopy with attenuated total reflectance (ATR-FTIR), and coupled microscopy. Data analysis was performed with principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). The samples were split into three groups: calibration set, validation set and external test set. The quality of the model was assessed by the figures of merit. **Results:** For the calibration set, sensitivity, specificity and efficiency rates were, respectively, 98,9%, 98.4% and 98.6%. For the validation group, they were all 100%. Internal validation set was 100% correctly classified, as well as 79.1% of the external test set. **Conclusions:** the technique is fast, clean and non-destructive. The attached microscope is important to avoid damaged regions of the samples. The age limitation of the samples in 10 years (due to natural wear) proved to be effective. The correct classification rate of the external test group (79.1%) refers to underrepresented classes. It can be potentially improved at any time, since the model allows the insertion of new samples. Therefore, the model built is promising, for forensic purposes.

Keywords: automotive paint, ATR-FTIR spectroscopy, chemometrics, PCA, PLS-DA, Forensic Science.

1 INTRODUÇÃO

No Brasil, o ato de um motorista se evadir do local após um acidente de trânsito é punível, conforme o Código de Trânsito Brasileiro (1):

*Art. 305. Afastar-se o condutor do veículo do local do acidente, para fugir à responsabilidade penal ou civil que lhe possa ser atribuída:
Penas - detenção, de seis meses a um ano, ou multa.*

Em novembro de 2018, o Supremo Tribunal Federal firma a sua jurisprudência sobre o assunto, tipificando a conduta como crime, e ratifica o entendimento de constitucionalidade desse artigo, em outubro de 2020 (2).

Pesquisa realizada através de sistema interno da Polícia Civil do Distrito Federal (PCDF), o Sistema Millenium, mostrou que durante o ano de 2020 foram registradas no Distrito Federal (DF) 839 ocorrências policiais de evasão de local de acidente de trânsito, o que dá uma média de 2,3 ocorrências por dia. Considerando-se que o ano de 2020 foi um ano atípico devido às restrições de funcionamento de escolas e comércio, motivadas pela pandemia de Covid-19, pegou-se também para referência o ano de 2019, quando foram registradas 1089 ocorrências dessa natureza, o que gera uma média de 3,0 ocorrências por dia registradas no DF, ou, em média, uma ocorrência a cada oito horas. É, portanto, uma conduta, além de criminosa, frequente.

A correlação entre acidentes de trânsito e o emprego de recursos do Sistema Único de Saúde é direta. Quanto maior o número de ocorrências, maior o número de internações e cirurgias, maiores os custos e os investimentos materiais e humanos. Uma vez que um condutor se afasta do local de um acidente de trânsito, ele deixa de prestar socorro a possíveis vítimas. Isso pode significar o agravamento do quadro clínico dos feridos, e muitas vezes, a morte.

Dentre as teorias do crime, existe uma que diz: “Não é o rigor do suplício que previne os crimes com mais segurança, mas a certeza do castigo (...). A perspectiva de um castigo moderado, mas inevitável causará sempre uma forte impressão mais forte do que o vago temor de um suplício terrível, em relação ao qual se apresenta alguma esperança de impunidade.” (3). Ou seja, a certeza de que será punido é o que mais inibe a conduta criminosa, e nem tanto a previsão de uma pena grave, que pode não vir a ser aplicada. E é nesse sentido que devem ser elaborados mecanismos mais certos quanto à identificação de veículos evasores de acidentes, de forma a possibilitar a punição e dessa forma, inibir a conduta criminosa de evasão do local.

Em ocorrências dessa natureza, as autoridades policiais precisam contar com fontes alternativas de informação para identificar o veículo e o motorista, como por exemplo, através

de testemunhas oculares ou de imagens de câmeras de vigilância. Muitas vezes, tais informações não estão disponíveis ou não são confiáveis. E em muitos casos, fragmentos da pintura automotiva (ou seja, lascas de tinta) são recuperados na cena do crime. Tais fragmentos de tinta podem fornecer informações cruciais sobre a marca, modelo e ano do veículo. Podem atuar, portanto, como um elo para orientar e se iniciar a investigação policial (4). A análise de tintas é, portanto, um importante ramo da Ciência Forense, e muito útil nesses casos em que apenas fragmentos de tinta permanecem no local da colisão.

Tal situação, de evasão de local de acidente, não é exclusiva do DF, ou tampouco do Brasil. Em outros países, também motivados pela necessidade de criação de mecanismos para punir e inibir o crime de evasão após acidentes, foram criados bancos de tinta automotiva com finalidade forense. Como exemplos, há uma base de dados europeia, o *European Paint Group* (EPG) da instituição *European Network of Forensic Science Institutes*, assim como uma base de dados canadense, o *Paint Data Query* (PDQ) da Polícia Real Montada do Canadá (*Royal Canadian Mounted Police*), que é utilizado há mais de 35 anos, sendo hoje em dia acessado por 102 laboratórios forenses, localizados em 24 países diferentes, incluindo laboratórios policiares, como do *Federal Bureau of Investigation* – FBI (4-7).

Pinturas automobilísticas realizadas em diferentes fábricas (ainda que de uma mesma marca veicular), apresentam diferenças entre si (8,9). Tais particularidades locais dos processos de pintura industrial impedem a comparação dos carros brasileiros com amostras de bancos de tinta já existentes.

Além dos exames demandados por ocorrências de acidentes veículos evasores, existe também outra demanda semelhante, gerada por exames de justaposição de veículos, realizados rotineiramente por institutos forenses. Trata-se de um exame para se avaliar se houve colisão recíproca entre dois veículos envolvidos em acidentes de trânsito, cuja preservação do local por algum motivo não tenha sido mantida e, portanto, não tenha havido a perícia no local do acidente. Nesses casos, os veículos são encaminhados para exame em pátios ou depósitos, onde os peritos avaliam apenas as características das avarias nos veículos, de forma a se determinar todos os elementos pertinentes à colisão. Nesses casos, sempre que o conjunto probatório permite, o perito estabelece se houve ou não colisão recíproca entre os veículos avaliados, ainda que nenhum deles tenha se evadido do local.

Observa-se, portanto, a utilidade e a importância de se criar e se manter um banco de tinta automotiva forense brasileiro, o que permitiria a realização de um novo tipo de perícia no Brasil, uma vez que a demanda é alta e o exame é sabidamente tradicional e confiável.

A análise da tinta automotiva visa a comparar, por exemplo, uma amostra de tinta coletada no local do crime, com um fragmento de tinta de algum veículo suspeito, ou mesmo comparar com uma amostra de tinta proveniente de um banco de dados de tinta (4). Atualmente, o protocolo internacional mais utilizado para avaliação e identificação de tintas automotivas nesses casos consiste nas seguintes etapas: avaliação do número de camadas e originalidade da pintura; a comparação visual da cor com tabelas de fabricantes de tintas; a obtenção do espectro infravermelho (IR) de cada camada; a comparação dos espectros com os espectros obtidos a partir de bases de dados IR de referência. A combinação desses resultados gera uma lista de possíveis veículos que podem ter se envolvido na cena do crime, de forma a facilitar a investigação policial (5,10).

A literatura revela o uso de diferentes técnicas para a análise de tintas automotivas, como é o caso da espectroscopia Raman (11,12), imagem hiperespectral (13,14). Entretanto, é o uso da espectroscopia de infravermelho com transformada de Fourier (FTIR) a técnica que se encontra mais bem estabelecida para a análise forense das tintas automotivas (4,8-10,15-22). Trata-se de uma técnica rotineiramente utilizada pela maioria dos institutos forenses, e também por instituições de ensino e pesquisa, e, portanto, tais equipamentos encontram-se frequentemente disponíveis para o uso, por peritos e por pesquisadores em geral. Por esses motivos, foi a técnica selecionada para esta pesquisa. Será detalhada oportunamente.

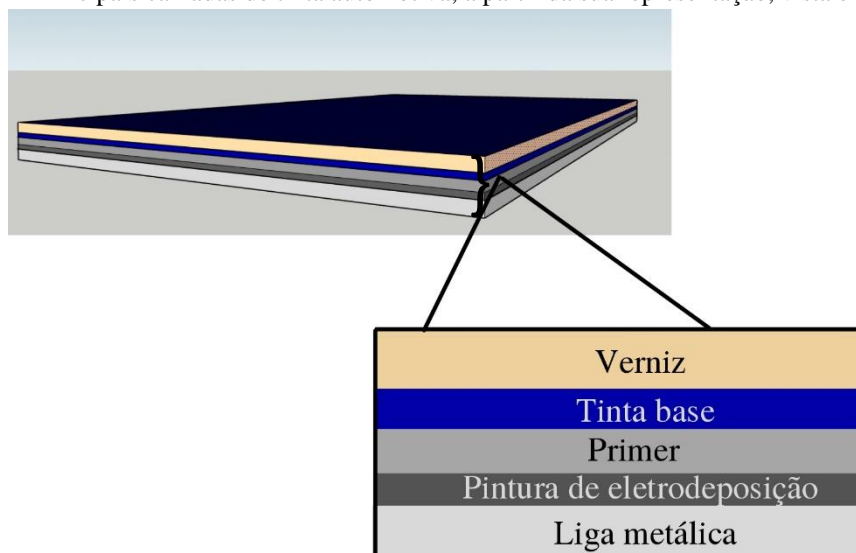
Uma vez selecionada a técnica, é fundamental a escolha das metodologias para o manejo dos dados coletados, a partir dos espectros obtidos. São necessárias técnicas estatísticas de análise multivariada para a interpretação dos dados e obtenção dos resultados. A quimiometria apresenta uma vasta gama de ferramentas (15–39) que precisam ser cuidadosamente escolhidas para cada momento da análise. A metodologia selecionada deve incluir a validação do modelo proposto, o que garante a confiabilidade de sua aplicação futura. Portanto, é uma etapa crítica e imprescindível para a criação do banco de dados.

1.1 COMPOSIÇÃO DAS TINTAS AUTOMOTIVAS

Tipicamente, as tintas automotivas são compostas por quatro camadas, seguindo-se a sequência da camada mais profunda para a mais superficial: pintura eletrodeposição, *primer*, tinta base e verniz. A pintura eletrodeposição é sobreposta à lataria do veículo. Ela confere resistência mecânica e proteção contra corrosão. A camada seguinte, o *primer*, confere aderência, fazendo a junção entre a camada anterior e a tinta base. A tinta base é a camada que confere a cor, propriamente dita, já que contém os pigmentos responsáveis pela cor e pelos efeitos visuais associados. O verniz é uma camada geralmente não pigmentada, que confere

proteção às demais camadas contra degradações causadas pela exposição a intempéries, como chuvas e exposição ultravioleta, além de produtos químicos utilizados na limpeza dos veículos (40). O número de camadas do tipo *primer* pode variar, a depender do tipo de metal constituinte do substrato e também da configuração de presença de certos elementos, muitas vezes adicionados a uma camada extra de verniz, dependendo do tipo de efeito visual final a que se propõe a pintura, como é o caso de tintas com brilho metálico (41). A espessura das camadas pode ser também muito variável, o que representa, por si só, um item individualizador na comparação entre duas formulações químicas. Mas a título de exemplo, como valores aproximados, podemos citar espessura de aproximadamente 20 μm para a pintura eletrodeposição, 35 μm para o *primer*, 15 μm para a tinta base e 40 μm para o verniz (42). A Figura 1 traz a representação dessas camadas.

Figura 1- Principais camadas de tinta automotiva, a partir da sua representação, vista em corte.



Fonte: Autoria Própria.

Na composição das camadas, uma infinidade de diferentes compostos químicos de proporções nanométricas é empregada. Fazem parte da composição: resinas alquídicas, acrílicas, epóxis, aglutinantes, aditivos, pigmentos, polímeros, estabilizantes e cargas. A título de exemplos, podemos citar óxidos férricos hidratados, mica, titanato de níquel, cromato de chumbo (amarelo cromo), vanadato de bismuto, azul de ftalocianina, violeta dioxazina, ftalocianina de cobre clorada, indantrona, melamine, dióxido de titânio anatase e rutilo e molibdato de cromato de chumbo. As diferentes composições e quantidades de cada substância nas formulações de tintas conferem a cada formulação características únicas, o que possibilita

a caracterização das tintas pelos seus dados espectrais, gerados a partir da utilização das técnicas espectroscópicas (17,43,44).

Não será objeto deste trabalho a caracterização química de grupos funcionais ou a comparação das bandas constituintes dos espectros gerados. A própria geração do espectro de cada amostra de tinta automotiva é por si só uma análise química dos seus componentes. Entretanto, por estarmos lidando com um grande número de amostras e devido à semelhança que as composições naturalmente apresentam, as análises serão feitas apenas com base nas técnicas quimiométricas, ou seja, cálculos estatísticos aplicados sobre os valores numéricos obtidos a partir dos dados espectrais gerados. Entretanto, não descartamos, para análises futuras, a utilização da caracterização química individual das tintas estudadas, já que se constitui de uma importante ferramenta.

1.2 ANÁLISE DE TINTAS AUTOMOTIVAS NOS ÚLTIMOS DOZE ANOS

Como parte componente dessa pesquisa de mestrado, foi realizada uma publicação científica de revisão bibliográfica (45). Sobre a contribuição do artigo publicado à literatura científica, justificativa e os objetivos propostos, algumas considerações serão feitas a seguir.

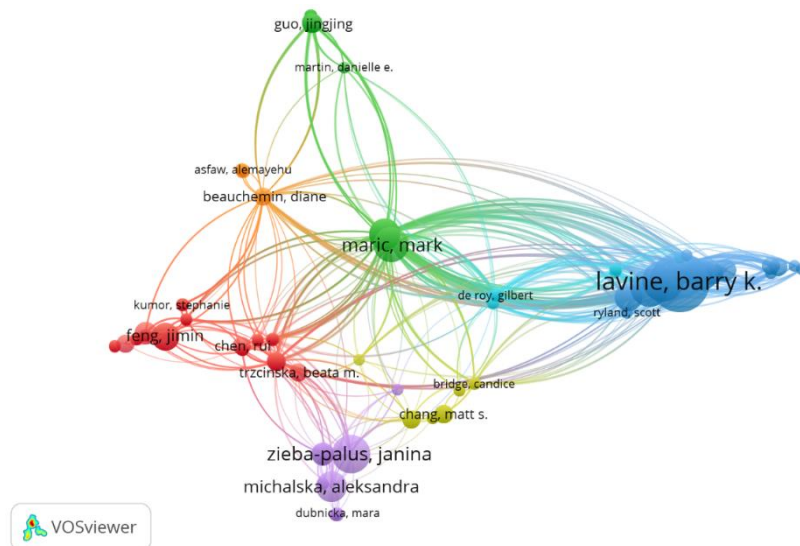
Artigos de revisão importantes já haviam sido escritos sobre tópicos semelhantes, tais como espectroscopia Raman (10,11,33), espectroscopia de infravermelho e imagem hiperespectral (13,14) aplicadas à Criminalística, e análise de dados multivariados e quimiometria (32). No entanto, nenhum dos artigos especificamente enfatizava o estudo de tintas automotivas, tampouco os diferentes métodos de análise que poderiam ser usados. Portanto, a publicação de um artigo científico de revisão bibliográfica com esse tema foi uma novidade na literatura.

O artigo publicado teve como título: “*Automotive paint analysis: How far has science advanced in the last ten years?*”. O trabalho analisou as metodologias utilizadas para a identificação de veículos evasores de colisões, utilizadas nas publicações sobre o tema, nos últimos dez anos. Resgatou e reuniu, portanto, todo o conhecimento de relevância científica a respeito do tema, além de ter sugerido a otimização e inovação dos procedimentos listados nos artigos trabalhados. Tornou-se, portanto, leitura fundamental para os pesquisadores que desejem iniciar novos bancos de tintas automotivas forenses, e, ainda, para os que já trabalham com o tema se atualizarem e terem uma visão geral das publicações a respeito.

O artigo utilizou a ferramenta estatística bibliometria para o estudo dos grupos e dos pesquisadores que trabalharam sobre o tema dentro do intervalo proposto. Essa ferramenta permite uma visualização interessante do relacionamento entre os artigos estudados,

ênfatizando quais são os grupos de pesquisadores que mais publicaram sobre o tema, e quais as relações de citações entre eles, como ilustra a Figura 2.

Figura 2- Estudo bibliométrico realizado no artigo: “*Automotive paint analysis: How far has science advanced in the last ten years?*”, o que possibilita a visualização da relação entre as publicações e citações entre os principais autores do tema.



Fonte: J.M. Duarte et. al, 2020 (45).

Foi realizado o levantamento bibliográfico entre os anos de 2010 a 2019, em três bases de dados: Web of Science, Scopus e Medline, com os termos para a busca: “(automotive OR car OR vehicle) AND (paint* OR “paint trace” OR “clear coat*” OR basecoat OR “primer surface” OR electrocoat) AND (identification OR examination OR classification OR analysis) AND (forensic OR justice OR crime)”. A partir da seleção realizada, que resultou em 65 artigos relevantes ao tema, pode-se traçar um perfil da literatura ao longo desses 10 anos. Para esta dissertação, acrescentamos, no presente item, os arquivos publicados, utilizando-se os mesmos termos, porém aplicados aos anos de 2020 e 2021, para fins de atualização da pesquisa.

A técnica mais comumente utilizada para a análise dos fragmentos de tintas automotivas é a espectroscopia de infravermelho (4,8,9,15–22,34–39,43,44,46–61), seguida por espectroscopia Raman, que muitas vezes é utilizada como técnica complementar à primeira (11,17,18,24–26,31,44,47–49,54–56,60,62–69). Outras técnicas comumente reportadas são a pirólise acoplada à cromatografia gasosa e espectrometria de massa (Py-GC/MS) (30,50,70), análise elementar, incluindo microscopia eletrônica de varredura (SEM), espectroscopia de raios X por dispersão em energia (EDS) e espectroscopia de fluorescência de raio-X (XRF) (43,53–55,71), espectrometria de massa com análise direta em tempo real (DART-MS) (27,51)

e sistema de imagem hiperespectral (13,23,72). A maioria delas são técnicas muito utilizadas para a análise de nanomateriais em geral.

Para o tratamento dos dados, as técnicas por quimiometria se tornaram fundamentais para a avaliação das tintas automotivas, já que se tratam de análises geralmente de grupos amostrais numerosos e muitas vezes com composições muito semelhantes. A análise de componentes principais (PCA) é o método mais frequentemente empregado para o tratamento não supervisionado dos dados (8,9,34–38,73). A análise hierárquica de agrupamentos (HCA), que, assim como a PCA, é uma análise exploratória, não supervisionada, também é frequentemente utilizada (9,15,29,30,35,36,38). A análise discriminante linear talvez seja a mais frequente dentre as análises supervisionadas (15,19–24,29,37,59,74–79). Coeficientes *wavelet* (8,9,17–20,22,31,34–38,65,73) e algoritmos genéticos para reconhecimento de padrões (9,19–22,36–38) também têm sido utilizados com muito sucesso nas análises de tintas automotivas.

A validação para os modelos supervisionados é fundamental para a confiabilidade e reprodutibilidade do modelo e é utilizada com frequência por autores do tema (8,9,15,19–22,24,34–38,50,65,73). Deve-se levar em consideração, ainda, a limitação da diferença de idade das amostras analisadas. O tema é objeto de estudo de alguns artigos, na literatura, que demonstram as diferenças químicas entre amostras envelhecidas (16,44,80,81). Quanto ao substrato das amostras, ou seja, se provenientes de partes plásticas ou de partes metálicas dos veículos, deve-se ter o cuidado de separar diferentes origens, para análise, já que se encontra estabelecida na literatura a impossibilidade de comparação de amostras de diferentes substratos, ou seja, tintas provenientes de partes plásticas do veículo não são comparáveis a tintas provenientes de partes metálicas (17,18,23,26,27,34,43,55,60). Quanto à origem de coleta das amostras, alguns estudos usam, como fontes de coletas, oficinas de instalação de teto solar (pois tais oficinas desprezam parte da lataria removida, do teto) (15,16,24,39,52,81). Muitos estudos se utilizam de bancos de fragmentos já estabelecidos, como é o caso do PDQ (8,9,19–22,34–38,46,59,73). Há estudos que tenham se utilizado de fragmentos de tinta coletados de locais de acidentes de trânsito (17,18,49,51,55,58,65,82).

O artigo deu como contribuições várias conclusões. 1) Para pesquisadores que desejem iniciar novos bancos de tinta automotiva, é interessante que se considere, primeiramente, técnicas consagradas, especialmente quando se tratam de técnicas limpas, de fácil manuseio, que exigem pouco ou nenhum preparo da amostra e que seja não destrutiva. A espectroscopia de infravermelho atende a todas essas características. 2) A validação do modelo é fundamental para se ter a real quantificação da qualidade do modelo e da sua aplicabilidade ao que se propõe.

3) É importante limitar a diferença de idade das amostras, de forma que a deterioração pelo tempo e condições climáticas não interfiram nas análises. 4) Devem-se separar os fragmentos coletados de partes metálicas do veículo, daqueles coletados de partes plásticas, uma vez que se tratam de tintas de composições químicas diferentes e, portanto, não comparáveis. 5) A forma de coleta com potencialmente menor viés de seleção dentre as observadas na literatura foi a coleta de fragmentos em locais de acidentes de trânsito, já que não privilegia cores, marcas ou padrão econômico dos veículos.

Podemos citar alguns artigos de relevância que foram publicados após a edição desse artigo de revisão bibliográfica. Destaca-se um artigo de revisão bibliográfica de Lavine et al. (83), que divide as análises das tintas automotivas em duas seções: 1) busca em bibliotecas (comparação de um espectro com espectros de biblioteca), e nesse caso, os algoritmos reportam um "índice de qualidade" ou uma "escala de similaridade" entre as amostra comparadas e 2) "reconhecimento de padrões", e nesse caso, características especiais fazem a identificação com espectros mais semelhantes possíveis, comparando com informações de um grupo treinado. Para aplicações forenses, não é possível individualizar uma amostra, mas apenas estimar a classe (no caso, marca e restringir modelos). Kwofie et al. (61) sugerem o preparo de amostras para seção transversal sem o uso de resina para fixação, o que pode contaminar a amostra e impedir a sua devida análise. Grant et al. (79), se utilizam de espectroscopia FTIR e análise LDA para a discriminação de classes de tintas de películas de vidros automotivos. Zieba-Palus (60) analisa as composições químicas de amostras retiradas de 30 veículos, sendo uma de parte metálica e uma de parte plástica de cada veículo, utilizando espectroscopias FTIR e Raman, o que reforça a necessidade de diferenciação dos substratos. A autora também encontra diferenças individuais das composições, o que pode ser resultado da tecnologia de produção da peça ou consequência de reparos ou repinturas durante a manutenção do veículo. Sauzier et al. (84) em artigo de revisão bibliográfica, elencam as principais análises quimiométricas para as aplicações forenses. Incluem as análises de tintas automotivas na revisão, o que reafirma a importância dessa ferramenta nas análises de tintas. Wei et al. (85), analisando fragmentos de tinta por espectroscopia ATR-FTIR, construíram dois modelos de classificação para comparação de resultados, com análise discriminante de Fisher - *Fisher discriminant analysis* (FDA) e máquina de vetores de suporte - *support vector machine* (SVM). Conseguiram, com SMV, maior acurácia de classificação. Qiu et al. (86), utilizando espectroscopia micro-Raman (MLRM) e ATR-FTIR, com análises de PCA, de rede neural perceptron multicamadas - *multi-layer perceptron neural network* (MLPNN) e de FDA, criaram modelos otimizados, com fusão de espectros, utilizando-se as regiões úteis para discriminação. Em comparação dentre as

modelagens testadas, FDA com o uso de espectros otimizados foi o que proveu melhores resultados.

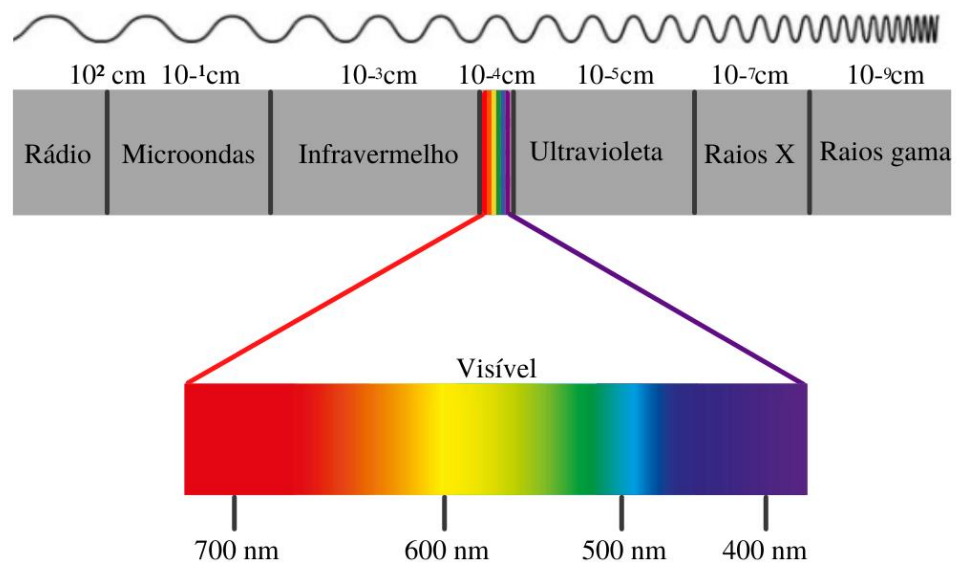
1.3 REFERENCIAL TEÓRICO

1.3.1 Espectroscopia de absorção na região do infravermelho

A teoria apresentada nesse tópico foi estudada nos livros Química Forense Experimental de Martins e Oliveira e Fundamentos de Química Forense de Bruni, Velho e Oliveira (87,88).

As técnicas espectroscópicas, muito utilizadas para materiais nanométricos, de maneira geral, se baseiam na interação da radiação eletromagnética com a matéria (por exemplo raios gama, ultravioleta (UV), luz visível (Vis), infravermelho (IR), micro-ondas). Enquanto a radiação UV-Vis muda a energia dos elétrons em íons e moléculas, a radiação por infravermelho altera a amplitude de vibração dos átomos. A Figura 3 mostra os comprimentos de ondas característicos aproximados de cada tipo de radiação eletromagnética.

Figura 3- Comprimentos de ondas aproximados de cada tipo de radiação eletromagnética.



Fonte: Autoria Própria. Adaptado de (41).

Da interação da energia com a molécula, resulta-se um gráfico (o espectro) da energia absorvida ou emitida (após a irradiação) em função da frequência, comprimento de onda ou número de onda.

A espectroscopia de infravermelho é um exemplo de espectroscopia vibracional. A energia com que os átomos vibram estão de acordo com o tipo e a força de ligação com outros átomos, com sua massa, com o meio e com a natureza de interação com íons ou moléculas

vizinhas. Portanto, caracterizam a substância de forma muito precisa e no caso específico das análises de tintas automotivas, de forma não destrutiva.

A energia dos átomos e moléculas (eletrônica, vibracional e rotacional) não é alterada de forma contínua, mas em “saltos”, a depender da quantidade de energia fornecida ou retirada do átomo ou molécula. Esse pacote de energia necessária para a alteração é chamado *quantum* (ou *quanta*, no plural). A radiação eletromagnética é caracterizada por uma frequência (ν) ou um comprimento de onda (λ), sendo que $\nu=c/\lambda$ (c =velocidade da luz) e a frequência é diretamente proporcional à energia da radiação (E_{rad}), sendo que $E_{\text{rad}}=h\nu$ (h =constante de Planck); ou $E_{\text{rad}}=h c/\lambda$. A radiação eletromagnética é composta por um campo elétrico (\vec{E}) e um campo magnético (\vec{B}), perpendiculares entre si e a direção de propagação da radiação.

Só haverá absorção ou emissão de radiação se a energia for igual à variação (ΔE) entre dois níveis energéticos daquele átomo ou molécula. A absorção é chamada de fenômeno de ressonância. A molécula ou íons está no chamado estado vibracional fundamental quando tem a mínima quantidade de energia vibracional possível. Demais estados vibracionais irão ocorrer caso haja a interação do íon ou molécula com o campo elétrico oscilante da radiação. Caso o momento de dipolo da molécula oscile com a mesma frequência da radiação, ocorrerá o acoplamento entre ambas e ocorrerá a absorção. Ou seja, a vibração terá sido ativa no infravermelho. Quanto maior for a magnitude do dipolo oscilante, maior será a transferência de energia da radiação e, portanto, maior será a intensidade de absorção no infravermelho. Após receber a energia, a matéria retorna para o estado original, emitindo radiação, e emitindo, por exemplo, fluorescência ou fosforescência.

A intensidade é dada por:

-Porcentagem de transmitância (% T) $= (I/I_0) \times 100$, sendo I =intensidade de radiação que atravessa a amostra, I_0 =intensidade de radiação que incide sobre a amostra; ou

-Absorbância $A = \log 1/T$ (logaritmo do inverso da transmitância).

Unidade usual: número de onda (cm^{-1}).

Cada substância apresentará diferentes picos de intensidade de absorção de energia em diferentes comprimentos de onda já que as vibrações dos átomos ocorrerão em valores específicos de frequência, a depender dos grupos funcionais que a constituem. É o que permite a caracterização da substância, na forma de espectro.

Emprega-se a expressão FTIR (infravermelho com transformada de Fourier, do inglês, *Fourier transform infrared spectroscopy*) para a designar, pois o espectro é extraído do

interferograma através de um conjunto de operações matemáticas, a transformada de Fourier. Os equipamentos FTIR são constituídos basicamente por uma fonte policromática contínua de radiação (um material cerâmico aquecido pela passagem de corrente elétrica), o interferômetro (com janelas e lentes) e um detector.

ATR (refletância total atenuada, do inglês, *attenuated total reflection*) - é um acessório baseado em reflexão, ao invés de transmissão. Os componentes da radiação que não são absorvidos são refletidos pela amostra. Após o devido processamento, chega-se a um espectro equivalente. ATR usa a reflexão total atenuada. O nome é explicado pois existe um ângulo de incidência crítico do feixe de radiação a partir do qual ocorre reflexão total. As ondas evanescentes são formadas na fronteira entre os meios e quando em contato com um cristal com alto índice de refração, a absorção das ondas evanescentes atenuará a reflexão total. O elemento a ser analisado é colocado em contato direto com o cristal. A radiação penetra a amostra com profundidade entre 0,5 e 10 μm , a depender da região do espectro (menor nas regiões de alta frequência), e do tipo de cristal. É necessário, portanto, fazer-se a correção da distorção do espectro, de forma que se assemelhe ao espectro no modo de transmissão. Tais correções já são realizadas pelo software do equipamento. O acessório ATR permite a acomodação de pequenos objetos, desde que tenha a superfície plana. Portanto, se torna interessante para a análise de lascas de tintas, já que as amostras apresentam essas características.

Microscopia FTIR – quando operando no modo de reflexão, não há necessidade de manipulação da amostra. Os espectros são obtidos de pontos específicos após a inspeção visual, em técnica realmente não destrutiva.

Como limitações da técnica de espectroscopia por infravermelho, pode-se citar que ela não fornece informações sobre a composição elementar das substâncias, tampouco faz a detecção de traços. Sofre interferência de água ou de grupos –OH (hidroxila).

1.3.2 Pré-processamentos dos dados

A partir da matriz X com os dados espectrais obtidos, são feitos pré-processamentos de forma a otimizar os resultados e facilitar o manejo dos dados.

Transformação de transmitância para absorbância: logaritmo do inverso da transmitância (87).

$$\text{Absorbância } A = \log 1/T \quad (\text{Equação 1})$$

Padronização normal de sinal (*standard normal variate* – SNV): usada em dados espectrais para corrigir efeitos de deslocamento de linha base. Cada linha de X é subtraída de sua média e dividida por seu desvio padrão (89).

$$X_{i,snv} = \frac{x_i^T - x_i^{-T}}{s_i^T} \quad (\text{Equação 2})$$

Centralização na média (*mean centering*): a média de cada coluna é subtraída dos dados. Elimina-se um fator constante dos dados e translada os dados da origem natural para a origem na média (90).

$$x_{ij}^* = x_{ij} - \bar{x}_j \quad (\text{Equação 3})$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (\text{Equação 4})$$

1.3.3 Análise de Componentes Principais (PCA)

As componentes principais (PCA, do inglês, *principal component analysis*) correspondem a combinações lineares das variáveis, de forma a darem o melhor ajuste para a análise de dados (91). A variância dos dados é maximizada para que haja o seu maior desvio ao longo da primeira componente principal. A definição da segunda componente principal e todas as demais é feita em função da máxima variância dos dados que ainda não foram quantificados pelas componentes principais anteriores. Uma componente principal será sempre ortogonal à variável principal anterior, de forma a não haver informação redundante (92).

É um método usado para decompor uma matriz X (em cujas linhas estão dispostas as amostras e nas colunas, valores de números de onda pré-estabelecidos) em uma soma de r matrizes, sendo r o número de componentes principais (93):

$$X = M_1 + M_2 + M_3 + \dots + M_r \quad (\text{Equação 5})$$

As matrizes M podem ser escritas como produtos de dois vetores:

$$X = t_1 p'_1 + t_2 p'_2 + t_3 p'_3 + \dots + t_r p'_r \text{ ou } \sum_{i=1}^r t_i p'_i + E \quad (\text{Equação 6}),$$

sendo t os escores e p os pesos (*loadings*). Pode também ser escrita na forma de matriz: $X = TP' + E$ (Equação 7), sendo que P' é composta por p' em suas linhas e T é composta por t em suas colunas. No novo eixo de coordenadas (das componentes principais), os escores são as novas coordenadas das amostras. Os pesos são os cossenos dos ângulos formados entre as variáveis originais e os novos eixos formados e contém a informação do peso de cada variável original no novo sistema de coordenadas. Cada linha de P' e cada coluna de T são correspondentes a uma componente principal e E é a matriz de resíduos, que o modelo não é capaz de explicar.

1.3.4 Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)

O método PLS-DA (do inglês, *partial least squared – discriminant analysis*) é desenvolvido pela correlação da matriz de treinamento \mathbf{X} com um vetor \mathbf{y} que indica se a amostra pertence à classe de interesse (valor de classe 1) ou às outras classes (valor de classe 0). Após a otimização do modelo, um limiar de discriminação (LD) é estimado com base na dispersão dos valores de classe estimados obtidos para o conjunto de treinamento de forma a minimizar a ocorrência de erros falsos positivos e falsos negativos. Como resultado, os modelos PLS-DA realizam uma discriminação binária, de modo que se trata de uma análise discriminante, o que significa que as amostras analisadas são classificadas como uma das classes predefinidas. No entanto, caso a amostra não corresponda a nenhuma das classes (valor estimado da classe (\hat{y}) menor que LD para todas as classes) ou tenha sido identificada como um *outlier* (ou seja, uma amostra com comportamento anômalo), é considerada uma análise inconclusiva para o método (94-97).

$$X = TP' + E \text{ ou } \sum_{i=1}^A t_i p'_i + E \quad (\text{Equação 8}),$$

$$y = Tq' + f \text{ ou } \sum_{i=1}^A t_i q'_i + f \quad (\text{Equação 9}),$$

cujas interpretações são análogas à PCA e A é o número de variáveis latentes a ser utilizado no modelo.

As estimativas de um conjunto de amostras, feitas para as suas propriedades de interesse (\hat{y}) são dadas pela multiplicação entre os dados de X e o vetor regressão b :

$$\hat{y} = Tq' = XW(P'W)^{-1}q' = Xb \quad (\text{Equação 10})$$

sendo que W é outra matriz pesos (*weights*), diferente de *loadings*, determinada pela regressão por PLS.

Primeiramente deve ser feita a escolha do número de variáveis latentes (VL) a ser utilizado. As variáveis latentes são uma entidade análoga aos componentes principais, da PCA, já que na PLS-DA também ocorre uma redefinição das coordenadas originais e é criado um novo sistema de coordenadas, com tantas dimensões quanto for o número de variáveis latentes. Esse cálculo do número de dimensões (ou coordenadas, ou variáveis), a serem trabalhadas, é feito baseado no erro médio de classificação do modelo. Para se estimar esse erro, calcula-se a raiz quadrada do erro médio quadrático de validação cruzada (*root mean square of cross validation – RMSECV*), dada pela Equação 11 (97):

$$RMSECV = \sqrt{\sum_{l=1}^n \frac{(y_l - \hat{y}_l)^2}{n}} \quad (\text{Equação 11})$$

Sendo que y_i é o valor de classe real atribuído a uma amostra i , \hat{y}_i é o valor estimado obtido pelo modelo, e n é o número de amostras.

A escolha correta do número de variáveis latentes é fundamental para não se utilizar informação redundante, o que gera a possibilidade de sobreajuste do modelo e, portanto, perda da sua capacidade preditiva. Por outro lado, a utilização de menos variáveis latentes do que o necessário gera subajuste do modelo.

Durante cada etapa, deve ser feita a identificação e exclusão de *outliers*. Essa identificação é baseada nos dados espectrais. Devem ser excluídas as amostras que excedem o nível de confiança definidos, tendo como referência dois parâmetros: T^2 de *Hottelling* e resíduos Q , sendo que T^2 quantifica o quão distante se encontra uma amostra do centro do modelo e Q é a somatória dos quadrados dos resíduos deixados pelo modelo, de acordo com as equações 12 e 13 (97).

$$Q = e_i e_i^T \quad (\text{Equação 12})$$

e_i é a decomposição residual obtida para a amostra i .

$$T_i^2 = t_i \Lambda^{-1} t_i^T \quad (\text{Equação 13})$$

Λ é a matriz diagonal contendo os autovalores associados ao número de variáveis latentes do modelo.

Os conceitos teóricos aplicáveis às demais etapas serão detalhados em seção futura, já que a sua compreensão depende primeiramente da exposição da metodologia utilizada nesse particular trabalho.

2 OBJETIVOS

O objetivo desse trabalho é, a partir de amostras de fragmentos de tintas automotivas, obter os espectros de absorção na região do infravermelho e construir um modelo matemático, que 1) seja capaz de classificar essas amostras pela marca do veículo e pela tonalidade da sua cor, e 2) que permita a predição de novas amostras, quando aplicadas ao modelo, de forma a) a identificar amostras que sejam pertencentes às classes (de marca e tonalidades) testadas e b) identificar como não pertencentes a tais classes as demais amostras. Isso permite o início da investigação criminal, mesmo quando não se tem qualquer outra informação de um veículo evasor de colisão, ou do seu condutor.

3 METODOLOGIA

3.1 COLETA DAS AMOSTRAS E PREPARO

Foram coletadas amostras de tinta automotiva, de diferentes marcas e modelos, de veículos envolvidos em acidentes de trânsito examinados pela Polícia Civil do Distrito Federal (PCDF), entre os anos de 2018 e 2020. As amostras foram coletadas por peritos criminais e agentes de polícia do Instituto de Criminalística, treinados para identificar amostras viáveis (pinturas originais do fabricante, em regiões mais íntegras, não avariadas pela colisão). As coletas foram realizadas com a inclusão de todas as cores disponíveis. Após a seleção das amostras de veículos com ano de fabricação entre 2011 e 2020, considerando-se apenas as amostras de pintura original do fabricante, conforme avaliação visual, somaram 453 no total. As amostras brancas totalizavam 189 itens. Para esse estudo, foram selecionadas apenas as amostras de cor branca, originárias de partes metálicas dos veículos, o que resultou em 143 amostras. Não foram feitas quaisquer outras restrições.

A escolha da cor branca, em detrimento das demais cores, dentre as amostras originalmente coletadas, se deveu apenas pelo fato de ser uma das cores com maior número de amostras, já que o número de amostras coletadas de acidentes acaba por refletir proporcionalmente o número de veículos circulantes (e a cor branca é muito frequente dentre a frota de veículos circulantes em geral).

Foi criado um banco de dados contendo as informações de cada amostra. As informações coletadas para cada amostra continham, sobre o veículo: a marca, o modelo, a cor, o tom da cor (nome comercial, da respectiva montadora), a placa de licenciamento, o ano de fabricação, o ano modelo, o número de identificação veicular -NIV- (o popular chassi do veículo), a fábrica onde o veículo foi produzido (informação constante no NIV), a parte do veículo de onde a amostra foi coletada, se havia repintura, e se a base da tinta era metálica (lataria do veículo) ou plástica (de partes plásticas). Cada amostra foi identificada com um código em que as três primeiras letras se referem à cor da tinta, as duas seguintes, em maiúsculo, se referem à marca do veículo, seguida pelo número da amostra e de uma letra minúscula para indicar outra parte de um mesmo veículo e em seguida, a letra M ou P, para indicar a base da tinta. Por exemplo, uma amostra WhiFI1aM representa uma amostra branca (Whi, do inglês *white* = branco), “FI” por ser da Fiat, “1” por ser o primeiro veículo branco da Fiat cadastrado, “a” por ser a primeira amostra cadastrada daquele veículo, e “M” por ter base metálica. Já uma amostra nomeada WhiFI1bM, significa que uma segunda amostra foi cadastrada desse mesmo veículo, tendo sido retirada de outra parte do veículo (veículo “1”, parte “b” do veículo). Exemplos de “partes” do veículo são: para-lamas, portas, capô, teto, capa do para-choque.

A restrição de idade (tempo de circulação) do veículo para o máximo de 10 anos se deveu pelo fato de esse grupo representar mais de 50% da frota circulante no Distrito Federal (quando do início das análises), conforme anexo contendo estatísticas do Departamento de Trânsito – Detran (veja Anexo). Restringindo-se a idade dos veículos, evita-se trabalhar com amostras muito degradadas por ação climática e desgastes do uso, e, portanto, com possíveis alterações em sua composição química, tanto devido ao desgaste quanto à migração de componentes através das diferentes camadas (16,44), o que impediria de se construir um modelo robusto.

A Tabela 1 ilustra as amostras coletadas e a sua separação por classes predefinidas, conforme a sua marca e tonalidade de cor (nomes dados pelos respectivos fabricantes).

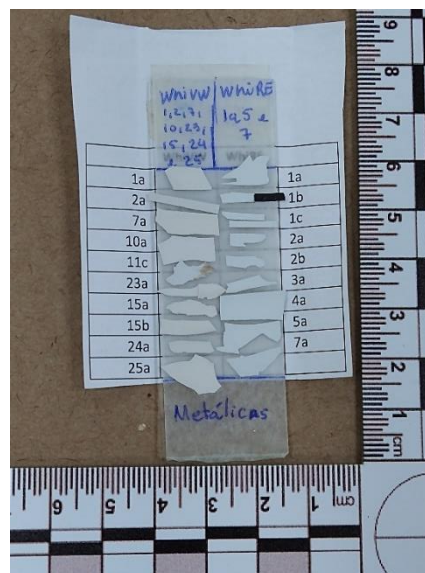
Tabela 1- Relação e número de amostras de tintas automotivas utilizadas, relacionadas conforme as respectivas classes, definidas pela marca e tonalidade de cor dos fabricantes.

Classes	Número de amostras	Classes	Número de amostras
Fiat Branco Banchisa	37	Toyota Branco Polar	3
Fiat Branco Perolizado	1	Audi Branco Geleira	2
Ford Branco Ártico	18	Citroen Branco Banquise	5
Ford Branco Pérola	2	Jeep Branco Perolizado	2
GM Branco Mahler	12	Kia Branco Clear	1
Honda Branco Taffeta	6	Lifan Branco	1
Honda Branco Perolizado	3	Mercedes Benz Branco	1
Hyundai Branco Creamy	9	Mercedes Benz CalcitWeiss	2
Hyundai Branco Perolizado	1	Mitsubishi Fairy White	1
Nissan Branco Polar	3	Peugeot Blanc Nacré	3
Nissan Branco Diamond	1	VW Branco Cristal	17
Renault Branco Glacier	9	VW Branco Puro	3
Total de amostras		143	

Para facilitar o seu manuseio, as amostras foram fixadas, em grupos com no máximo 20 delas, em lâminas de vidro, com a utilização de fita dupla face. As amostras foram fixadas de forma a expor a camada mais superficial da tinta, ou seja, o verniz. A região corresponde à

pintura de eletrodeposição (aquela que originalmente fica em contato com o veículo) ficava fixada à fita dupla face. Cada amostra era devidamente identificada quando dessa fixação, como ilustra a Figura 4. O uso desse aparato permitiu mais rapidez durante a coleta dos espectros (a ser descrita), já que evitava a necessidade de movimentação da mesa (do equipamento) a cada análise, pois essa movimentação passou a ser feita a cada 20 análises, ou seja, após a análise de 20 amostras. O uso da fita dupla face se mostrou muito vantajoso para deixar as amostras planas, uma vez que o fragmento de tinta, após removido do veículo, tem uma certa tendência a emborcar a sua superfície. Uma vez aderido à lâmina de vidro, o fragmento de tinta permanece com a sua superfície plana, o que é um ponto crítico durante a análise. Quando da visualização da amostra através do microscópio (a ser descrito), é necessário ajustar o foco conforme a distância do objeto à lente. Caso a amostra não estivesse com a sua superfície plana, seria necessário ajustar o foco a cada deslocamento lateral durante a sua visualização. Mantendo-se a superfície plana, o ajuste do foco foi feito uma única vez, a cada grupo de 20 amostras. Todas as amostras, após fixadas nas lâminas de vidro, foram higienizadas com etanol 70% v/v, para que as impurezas fossem removidas.

Figura 4- Exemplo de lâmina de vidro preparada com 19 amostras de tintas automotivas, aderidas por fita dupla face e devidamente identificadas, preparadas para as análises com espectroscopia de infravermelho.

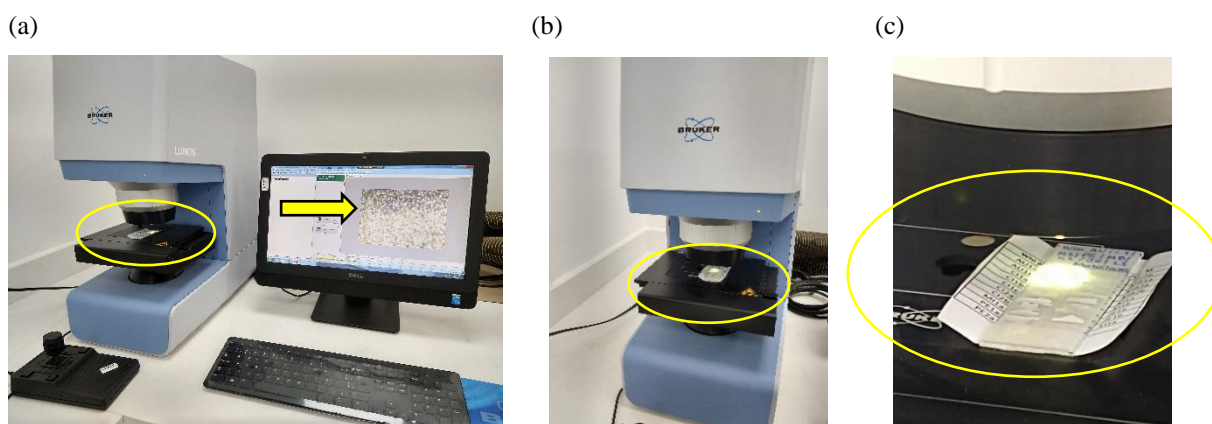


Fonte: Autoria Própria

Para este trabalho, os espectros de absorção no infravermelho com acessório de ATR das amostras de tintas automotivas foram coletados utilizando o equipamento LUMOS® com microscópio acoplado, equipado com um cristal ATR de germânio, da empresa Bruker Optics (conforme ilustram a Figura 5 a b c), na região de 4000 a 600 cm^{-1} , com resolução de 4 cm^{-1} e 64 varreduras. O cristal ATR foi limpo com metanol PA (59) antes de cada análise. As análises

foram feitas a partir da superfície exposta da tinta, correspondente à sua camada mais superficial, o verniz. Foram selecionados, para a coleta espectral, cinco pontos diferentes em cada amostra (portanto, em quintuplicata), com auxílio do *joystick* do equipamento. Algumas amostras apresentaram avarias, entretanto, seu espectro foi coletado da mesma forma, evitando-se, ao máximo, as regiões mais avariadas. Como foram analisadas 143 amostras e feitas cinco replicatas em cada, foram obtidos 715 espectros.

Figura 5- a, b, c- Ilustram o equipamento LUMOS® Bruker Optics, com uma lâmina de amostras de tintas automotivas em sua mesa de trabalho (elipses), durante a aquisição dos espectros infravermelho. A imagem ampliada da amostra em análise, pelo uso da microscopia, é vista no monitor de trabalho (seta).



Fonte: Autoria própria

3.2 TRATAMENTO DOS DADOS

Todos os cálculos foram realizados usando o software MATLAB (versão R2021b, MathWorks) e o PLS Toolbox (versão 881, Eigenvector). Uma vez que o programa gera os gráficos com as legendas dos eixos em língua inglesa, algumas das figuras terão numerações com a separação decimal feita por ponto, e não por vírgula, como é o correto da língua portuguesa. Devido a esses campos não serem editáveis, nos desculpamos com o leitor devido a essa limitação encontrada.

3.2.1 Pré-processamentos e PCA

As representações numéricas de cada espectro foram adicionadas a uma matriz, sendo que cada linha corresponde a uma amostra e cada coluna corresponde a um número de onda.

Primeiramente procedeu-se com uma avaliação visual dos 715 espectros, para a remoção daqueles que tinham comportamento anômalo, como aqueles que apresentavam diferentes bandas, ruído ou grandes desvios de linha de base, quando comparados aos demais espectros da mesma amostra. Essa avaliação visual foi realizada por comparação entre os cinco

espectros de uma mesma amostra, através da sua visualização simultânea (em uma mesma plotagem). Com os espectros remanescentes, procedeu-se ao pré-processamento dos dados, através de transformação de transmitância para absorvância, padronização normal de sinal (*standard normal variate* – SNV) e centrar na média (*Mean Centering*). Foi realizada a supressão, nos espectros, dos trechos referentes 1) às extremidades, que apresentam ruídos característicos do equipamento utilizado (de 4000 a 3350 cm^{-1} e de 690 a 600 cm^{-1}) e 2) à faixa do espectro relacionada à concentração de dióxido de carbono (CO_2) no ambiente (de 2720 a 1950 cm^{-1}). Tais regiões do espectro serão aqui denominadas “faixas não úteis”.

Foi realizada a análise exploratória dos dados, não supervisionada, para reconhecimento de padrões. Para tal análise, foi utilizada a PCA. Foram feitas análises com duas e com três componentes principais. O detalhamento dessa etapa será realizado quando da apresentação dos resultados.

3.2.2 PLS-DA

Para a construção do modelo PLS-DA, deve-se fazer a definição das classes de amostras conhecidas. É uma abordagem muito adequada quando se tem bem definidas as classes que serão separadas, como é o caso dos grupos de tintas automotivas, que são definidas por uma cor e tom de cor da sua respectiva marca. Em linhas gerais, para o desenvolvimento do modelo, foram realizadas três etapas. A primeira etapa foi a utilização de amostras conhecidas e representativas de classes conhecidas, denominadas conjunto de treinamento (também chamado de conjunto de calibração, ou grupo de treinamento ou grupo de calibração), para a construção do modelo. Na segunda etapa, foram utilizadas outras amostras das classes conhecidas, que não foram usadas no primeiro grupo, denominadas conjunto de validação (ou grupo de validação), para validar o modelo proposto. E na terceira etapa, foram utilizadas amostras desconhecidas, denominadas conjunto teste (também chamado, nesse caso, de conjunto de teste externo, ou grupo de teste externo), para a predição pelo modelo criado e validado. Nessa terceira etapa (do teste externo), avalia-se a capacidade do modelo de detectar amostras desses grupos como não pertencentes às classes já conhecidas e serem, portanto, consideradas como *outliers* (já que esse é o comportamento esperado, nesse caso).

As classes foram divididas entre os grupos (de treinamento, validação e teste) como consta na Tabela 2.

Tabela 2- Relação e número das amostras nomeadas conforme as respectivas classes de tintas automotivas, utilizadas nas análises, após separadas dentre os grupos de treinamento, de validação e de teste.

	Classes	Treinamento (número de amostras)	Validação (número de amostras)	Teste (número de amostras)	Total de amostras
1	Fiat Branco Banchisa	25	12		37
2	Fiat Branco Perolizado			1	1
3	Ford Branco Ártico	12	6		18
4	Ford Branco Pérola			2	2
5	GM Branco Mahler	8	4		12
6	Honda Branco Taffeta	4	2		6
7	Honda Branco Perolizado			3	3
8	Hyundai Branco Creamy	6	3		9
9	Hyundai Branco Perolizado			1	1
10	Nissan Branco Polar			3	3
11	Nissan Branco Diamond			1	1
12	Renault Branco Glacier	6	3		9
13	Toyota Branco Polar			3	3
14	Audi Branco Geleira			2	2
15	Citroen Branco Banquise	3	2		5
16	Jeep Branco Perolizado			2	2
17	Kia Branco Clear			1	1
18	Lifan Branco			1	1
19	Mercedes Benz Branco			1	1
20	Mercedes Benz CalcitWeiss			2	2
21	Mitsubishi Fairy White			1	1
22	Peugeot Blanc Nacré			3	3
23	VW Branco Cristal	12	5		17
24	VW Branco Puro			3	3
	Total de amostras	76	37	30	143

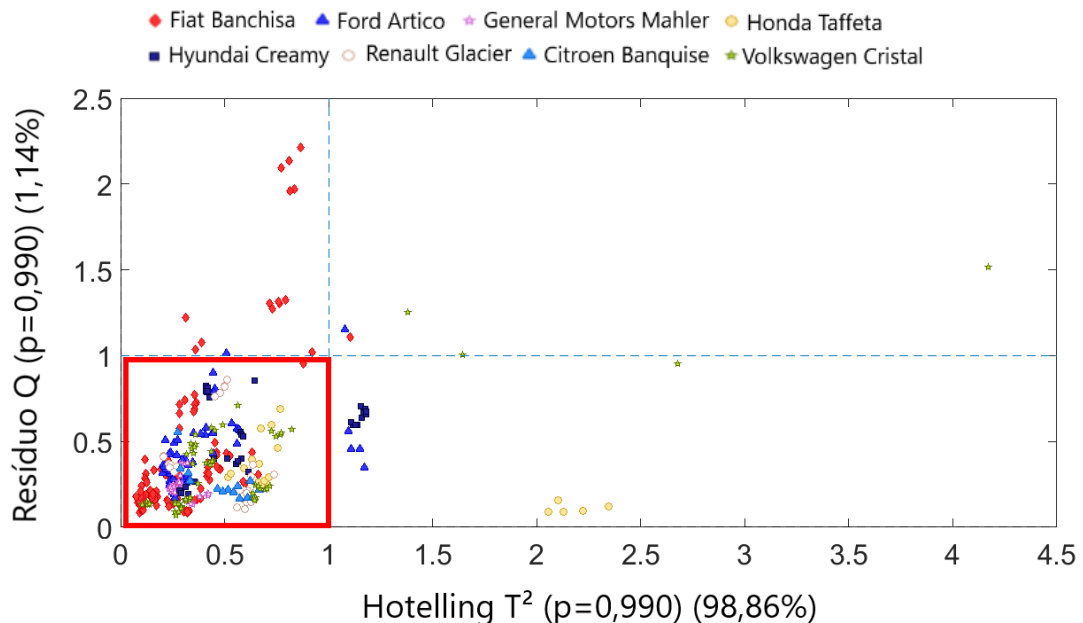
Fonte: autoria própria.

3.2.2.1 Construção do modelo

Foi realizada a definição do número de variáveis latentes a ser utilizada no modelo, conforme o cálculo de RMSECV, que será detalhado no tópico “Resultados”.

Para a definição de outliers, foram calculados os valores de *Hotelling T²* e Q residual, conforme exemplifica a Figura 6. O nível de confiança estabelecido para esse modelo foi de 0,99. A amostra é considerada não representativa do modelo (*outlier*) e deve ser excluída quando: 1) ambos se apresentam com valores altos; ou 2) quando pelo menos um deles se apresenta com valor alto. Essa decisão é discricionária ao pesquisador. Decidimos utilizar o segundo caso, ou seja, eliminando, portanto, tanto os espectros que apresentaram alto valor de *Hotelling T²* quanto os que apresentaram alto valor de Q residual. A análise se torna, dessa forma, mais rigorosa. São consideradas viáveis, portanto, apenas as amostras posicionadas no primeiro quadrante (delimitado na Figura 6 com um destaque vermelho), com baixos valores tanto de *Hotelling T²* quanto do resíduo Q.

Figura 6- Gráfico de *Hotelling T²* versus o resíduo Q da plotagem inicial dos dados, quando da construção do modelo. A área circundada pela linha vermelha destaca os espectros que tiveram baixos valores tanto de *T²* quanto Q, e, portanto, bem adaptados ao modelo. Os demais espectros são considerados *outliers*.

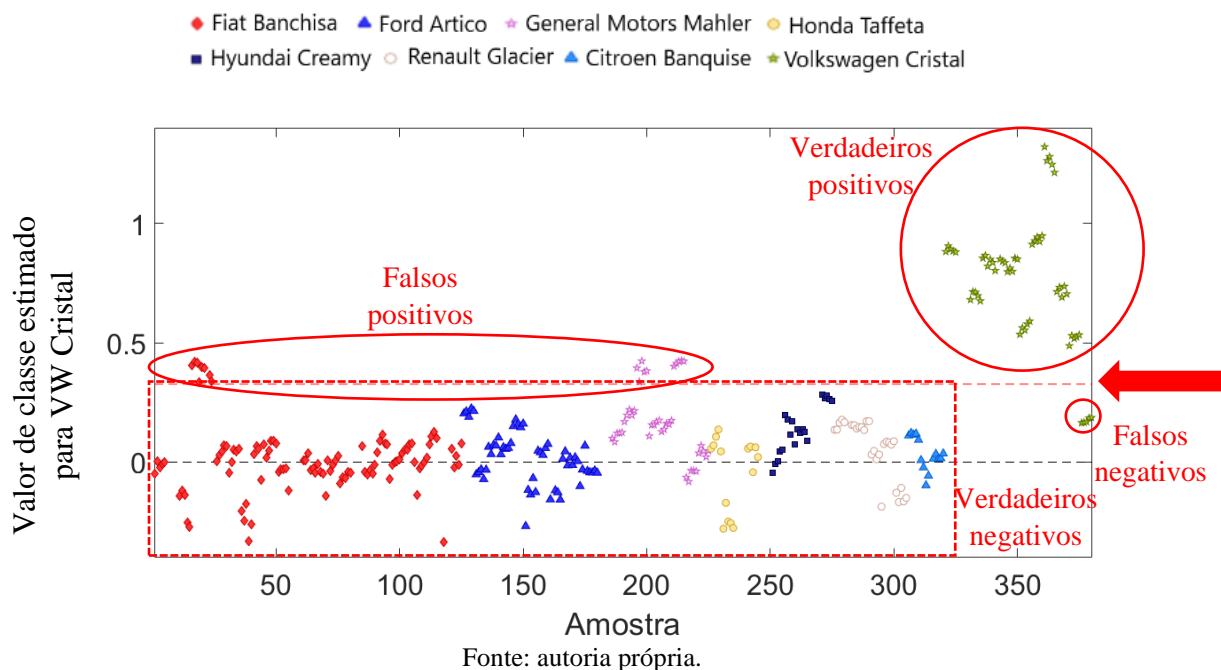


Fonte: autoria própria.

Para a discriminação das amostras, a partir dos valores estimados para o conjunto de treinamento, foi estimada, para a análise de cada classe, a probabilidade de cada espectro analisado de pertencer à classe analisada, e a probabilidade do espectro de não pertencer a ela. Foi feita, dessa forma, a determinação dos limites de discriminação de cada classe. A partir

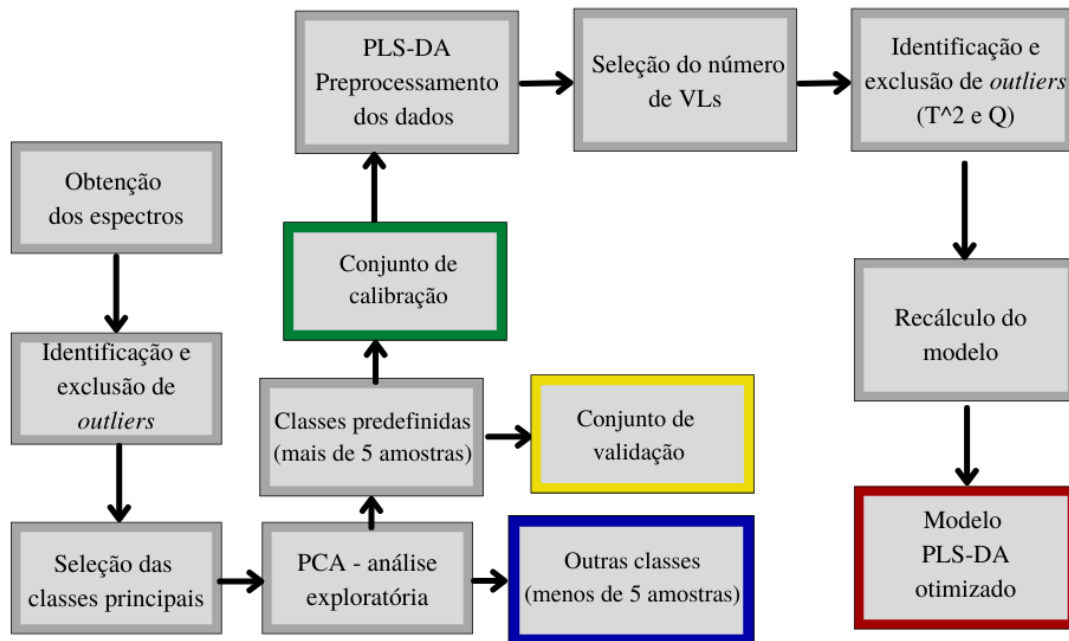
desses parâmetros, foi feita a comparação dos valores, de forma a classificar a amostra como pertencente ou não à classe analisada. Tal procedimento é repetido para cada classe, gerando-se um gráfico para cada classe definida. O gráfico possibilita a visualização dos espectros classificados corretamente como pertencentes à classe em questão (os verdadeiros positivos), dos espectros classificados erroneamente como não pertencentes à classe (os falsos negativos), dos espectros classificados corretamente como não pertencentes à classe (os verdadeiros negativos) e dos espectros classificados erroneamente como pertencentes à classe (os falsos positivos). A Figura 7 ilustra uma análise (de VW Cristal, a ser detalhada, no item de apresentação de resultados), aqui posicionada a título exemplificativo, de forma a facilitar a compreensão da análise de cada uma das classes.

Figura 7- Análise PLS-DA da classe Volkswagen Cristal. A seta destaca a linha de discriminação entre o grupo de espectros classificados como Volkswagen Cristal (espectros posicionados acima da linha) e o grupo de espectros classificados como não pertencentes à classe Volkswagen Cristal (posicionados abaixo da linha). Os espectros Volkswagen Cristal posicionados acima da linha são verdadeiros positivos. Os posicionados abaixo são falsos negativos. Os espectros de outras classes posicionados acima da linha são falsos positivos e as abaixo da linha são verdadeiros negativos.



A sequência de procedimentos para a construção do modelo está ilustrada na Figura 8. Apenas o grupo de calibração é utilizado para a construção do modelo, utilizando-se o número definido de variáveis latentes. Após a construção do modelo (utilizando-se os valores de T^2 e Q para a identificação e exclusão de *outliers*), o modelo é recalculado, chegando-se, por fim, ao modelo definitivo de calibração.

Figura 8- Sequência para a criação do modelo de calibração definitivo PLS-DA. Após a aquisição dos espectros, a primeira identificação de *outliers* foi feita de forma visual, pela inspeção simultânea dos cinco espectros da mesma amostra. As classes principais foram selecionadas conforme o número de amostras (classes mais numerosas). Para a análise PCA, foram utilizadas todas as classes. Para a construção do modelo, foi utilizado o conjunto de calibração, pertencente às classes principais. Após a seleção do número de variáveis latentes (aqui denominadas VLs), a identificação e exclusão de *outliers* foi feita baseado em *Hotelling T²* e *Q* residual. Após a exclusão dos *outliers*, foi feito o recálculo do modelo, gerando-se, dessa forma, o modelo PLS-DA otimizado.



Fonte: autoria própria.

Para se fazer a definição das classes principais (aquelas que serão usadas para a construção do modelo e para a primeira validação), o critério utilizado foi o uso de classes mais numerosas. Foram utilizadas apenas as classes com 5 ou mais amostras para comporem os conjuntos de treinamento e de validação. Ou seja, as amostras contidas na Tabela 3. O grupo de classes principais totalizou oito classes, a saber: Fiat Branco Banchisa, Ford Branco Ártico, GM Branco Mahler, Honda Branco Taffeta, Hyundai Creamy White, Renault Branco Glacier, Citroen Branco Banquise e VW Branco Cristal, que juntas somaram 113 amostras. Foi dividido da seguinte forma: dois terços das amostras foram utilizadas para comporem o grupo de calibração (76 amostras) e um terço foi utilizado para compor o grupo de validação (37 amostras). As classes menos numerosas (aquelas com quatro amostras ou menos), totalizando 16 classes, com 30 amostras em seu total, compuseram o grupo teste externo, e não entram nessa etapa (veja Tabela 2).

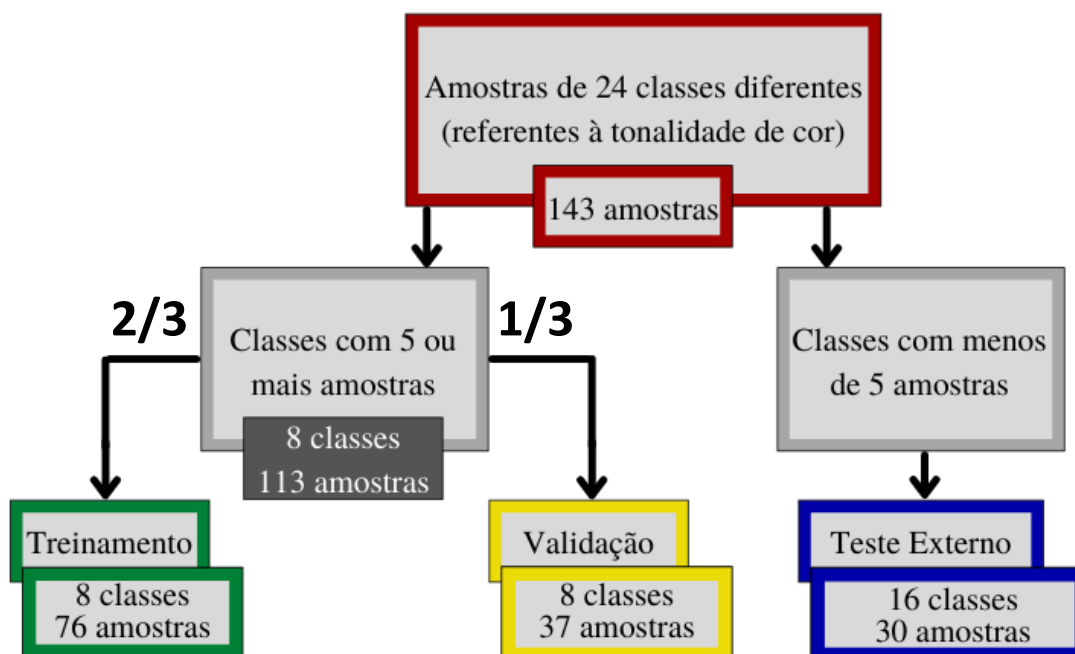
Tabela 3 – Relação e número de amostras das classes principais (aquelas mais numerosas, contendo cinco ou mais amostras) – conforme marcas e tonalidades de cores dos fabricantes (classes), divididas dentre os conjuntos de calibração e de validação.

	Classes utilizadas no Grupo de Treinamento e no Grupo de Validação	Treinamento (número de amostras)	Validação (número de amostras)	Total de amostras
1	Fiat Branco Banchisa	25	12	37
2	Ford Branco Ártico	12	6	18
3	GM Branco Mahler	8	4	12
4	Honda Branco Taffeta	4	2	6
5	Hyundai Creamy White	6	3	9
6	Renault Branco Glacier	6	3	9
7	Citroen Branco Banquise	3	2	5
8	VW Branco Cristal	12	5	17
	Total de amostras	76	37	113

Fonte: autoria própria.

A Figura 9 ilustra a divisão dos grupos amostrais.

Figura 9- Divisão das classes de amostras em grupos de calibração, de validação e de teste externo a partir das 24 classes de tintas automotivas, definidas previamente conforme as marcas e tonalidades de cor dos veículos.



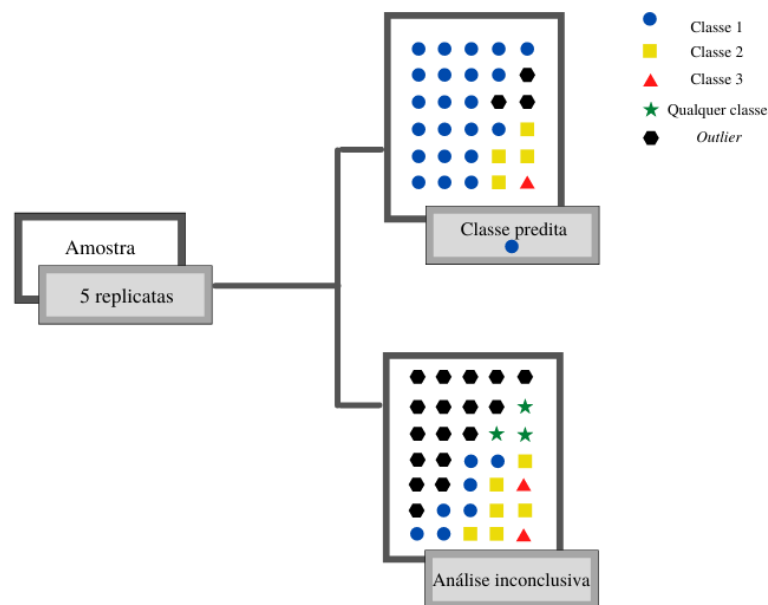
Fonte: autoria própria.

Como para cada amostra foram coletados espectros em quintuplicata, ou seja, cinco espectros ATR/FTIR, foi necessário adotar-se um critério para se definir como cada amostra seria classificada (dentre as possíveis oito classes principais), conforme o resultado da análise de cada um dos seus cinco espectros. Foi definido como critério que é necessária a classificação de no mínimo três espectros de uma mesma amostra classificada em uma determinada classe para a sua caracterização como pertencente à classe em questão. Caso a amostra não tenha tido pelo menos três espectros classificados em uma determinada classe, a análise daquele espectro foi considerada inconclusiva e a amostra foi eliminada como *outlier*. A amostra é considerada inconclusiva nas seguintes situações:

- Três, quatro ou cinco replicatas são consideradas *outliers*;
- Uma ou duas replicatas são consideradas *outliers* e as demais são consideradas pertencentes a pelo menos duas classes diferentes;
- A amostra não teve *outliers*, mas teve no máximo duas replicatas consideradas pertencentes a uma mesma classe.

Tais critérios estão ilustrados na Figura 10.

Figura 10- Critérios para a definição da classe da amostra (definida pela marca e tonalidade de cor do veículo), conforme a classificação de suas cinco replicatas. Caso três ou mais das cinco replicatas tenham sido classificadas para uma determinada classe, então a amostra é positiva para aquela classe. Nas demais situações, a sua classificação é dada como inconclusiva).



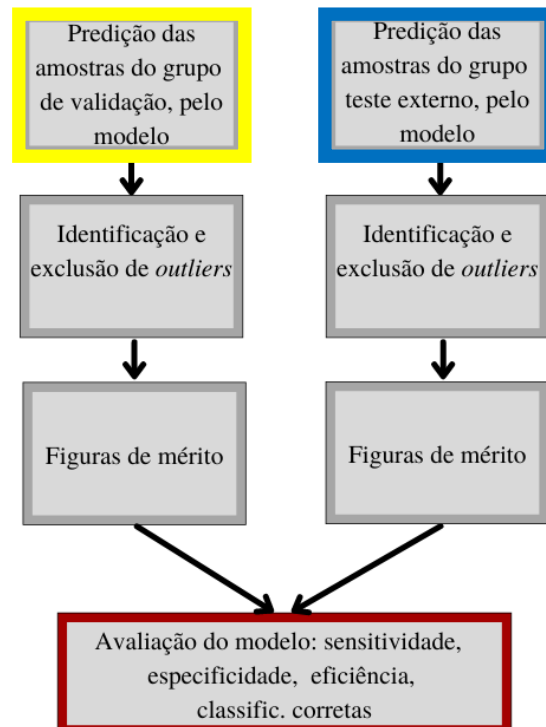
Fonte: autoria própria.

3.2.2.2 Validação do modelo

A validação do modelo ocorreu em duas etapas. Na primeira etapa, foi utilizado o grupo de validação (aquele contendo um terço das amostras das oito classes principais). Na segunda

etapa de validação, foi utilizado o grupo teste externo (aquele contendo as amostras pertencentes às 16 classes menos numerosas e não treinadas). A validação ocorreu como ilustrada na Figura 11. Cada um desses grupos (validação e teste externo) tiveram seus espectros preditos pelo modelo, em etapas distintas. Para cada etapa, após a predição inicial, foi feita a identificação e exclusão de *outliers*, e então foi feita uma nova predição. Com o resultado obtido, foram calculadas as figuras de mérito, que são os parâmetros que avaliam a qualidade do modelo, e serão detalhados adiante, oportunamente. A intenção da predição com o grupo de validação é avaliar se o modelo é capaz de identificar essas amostras corretamente, nas suas respectivas classes, já que elas pertencem ao mesmo grupo das poito classes utilizadas para a construção do modelo. A intenção da predição com o uso do grupo teste (composto por amostras das 16 classes não treinadas) foi avaliar se o modelo é capaz de identificar essas amostras como *outliers*, pois esse é o comportamento esperado de amostras cujas classes não tenham participado da construção do modelo.

Figura 11- Sequência para a validação do modelo PLS-DA de análise de tintas automotivas, usando-se os grupos de validação e de teste externo. Após a predição das amostras (em cada uma das etapas ilustradas), foram calculadas as figuras de mérito, para a avaliação da qualidade do modelo construído.



Fonte: autoria própria.

Na etapa da validação, deve ser feita um esclarecimento a respeito do que se considera como uma classificação correta, o que é considerada uma classificação errônea e o que é considerada uma análise inconclusiva. A classificação correta ocorre quando uma amostra é considerada, na sua própria classe, como positiva (e, portanto, verdadeira positiva), e ao mesmo tempo considerada como negativa em todas as demais classes (e, portanto, verdadeira negativa). A classificação errônea ocorre quando uma amostra é considerada como negativa em sua própria classe (e, portanto, falso negativa) e ao mesmo tempo é considerada como positiva em apenas uma classe diferente da correta (e, portanto, falso positiva, em apenas uma classe). A análise inconclusiva ocorre em duas situações: 1) quando uma amostra é considerada como positiva na sua própria classe (e, portanto, verdadeira positiva) mas ao mesmo tempo é considerada como positiva em uma ou mais classes diferentes (e, portanto, falsa positiva). Nesse caso, a amostra fica tecnicamente considerada como positiva em mais de uma classe diferente, e, portanto, a análise é inconclusiva; e, 2) quando uma amostra é considerada como negativa na sua própria classe (e, portanto, falso negativa) e, ao mesmo tempo, é considerada como negativa em todas as demais classes (e, portanto, verdadeira negativa). Nesse caso, essa amostra, tecnicamente, não está classificada em classe alguma, e, portanto, a análise é inconclusiva. Essa interpretação pode ser visualizada no Quadro 1.

Quadro 1 – Critérios para a definição de classificações corretas, de classificações errôneas ou de análises inconclusivas das amostras, realizadas pelo modelo, levando-se em consideração a análise da amostra na sua própria classe e a análise da amostra nas demais classes. FP=falso positivo, VN=verdadeiro negativo, VP=verdadeiro positivo e FN=falso negativo.

Análise em outras classes		FP	VN
Análise na própria classe	VP	Análise inconclusiva (amostra classificada em duas classes)	Classificação correta
	FN	Classificação errônea	Análise inconclusiva (amostra não classificada em qualquer classe)

Fonte: autoria própria.

3.2.2.3 Figuras de mérito

Após essa análise, são somados, nas análises do grupo de treinamento e do grupo de validação, os números correspondentes aos verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. Para o grupo de teste externo, a única figura aplicável é o falso positivo, já que não existe análise dentro da própria classe da amostra, tendo em vista que as classes desse grupo não foram categorizadas durante a confecção do modelo. A partir da soma dessas figuras, analisam-se a taxa de sensibilidade, a taxa de especificidade, a taxa de eficiência, a taxa de classificação correta do modelo e a taxa de resultados indefinidos, conforme as fórmulas ilustradas na Tabela 4 (98,99), cada uma com a sua respectiva aplicabilidade. A análise dessas figuras de mérito é o que possibilita a avaliação da qualidade do modelo.

Tabela 4 – Equações matemáticas para se calcular as figuras de mérito do modelo e seus grupos de aplicabilidade. FP = número de falsos positivos; FN = número de falsos negativos; VP = número de verdadeiros positivos; VN = número de verdadeiros negativos, N = número de amostras do conjunto analisado e I = número de amostras testadas.

Figura de mérito	Equação	Grupo de aplicabilidade
Taxa de falso negativo	$TFN = \frac{FN}{FN + VP} 100$	Equação 14 Calibração e validação
Taxa de falso positivo	$TFP = \frac{FP}{FP + VN} 100$	Equação 15 Calibração, validação e teste
Taxa de sensibilidade	$TST = \frac{VP}{FN + VP} 100$	Equação 16 Calibração e validação
Taxa de especificidade	$TSP = \frac{VN}{VN + FP} 100$	Equação 17 Calibração e validação
Taxa de eficiência	$TEF = \sqrt{\frac{VP * VN}{(VP + FN) * (VN * FP)}} 100$	Equação 18 Calibração e validação
Taxa de classificação correta	$TCC = \frac{(N - FP - FN)}{N} 100$	Equação 19 Validação e teste
Taxa de resultados indefinidos	$TINDEF = \frac{outliers + incon}{I} 100$	Equação 20 Calibração, validação e teste

Fonte: autoria própria.

Para a denominação das amostras, doravante serão omitidos os trechos “Whi” (relativo a branco) e a letra “M” (relativo a metálico), apenas para facilitar a escrita, uma vez que todas as amostras são brancas e foram removidas de partes metálicas dos veículos. Para exemplificar, a amostra denominada WhiFI16aM passará a ser denominada apenas como FI16a.

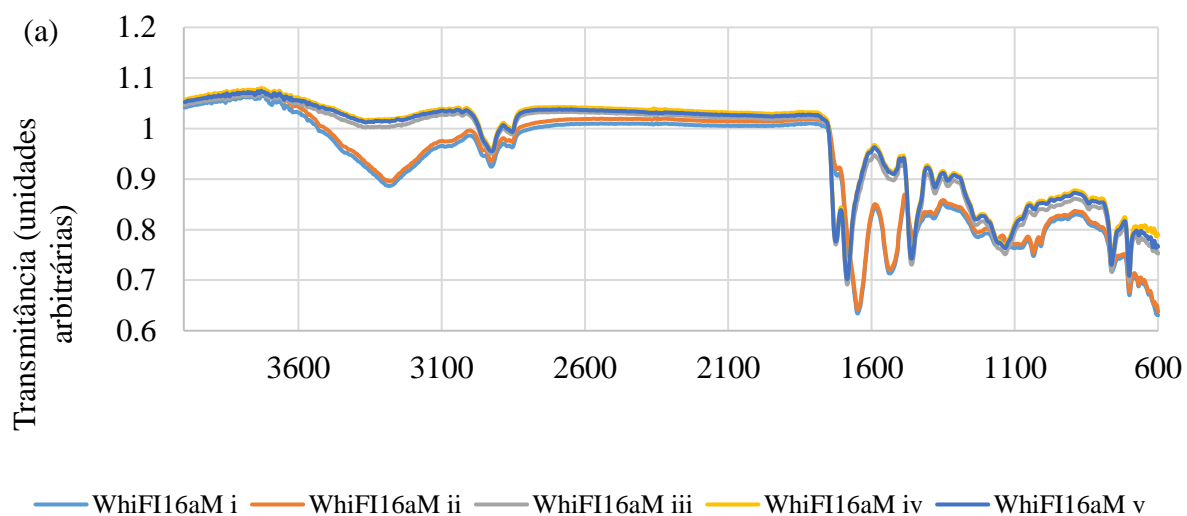
4 RESULTADOS E DISCUSSÃO

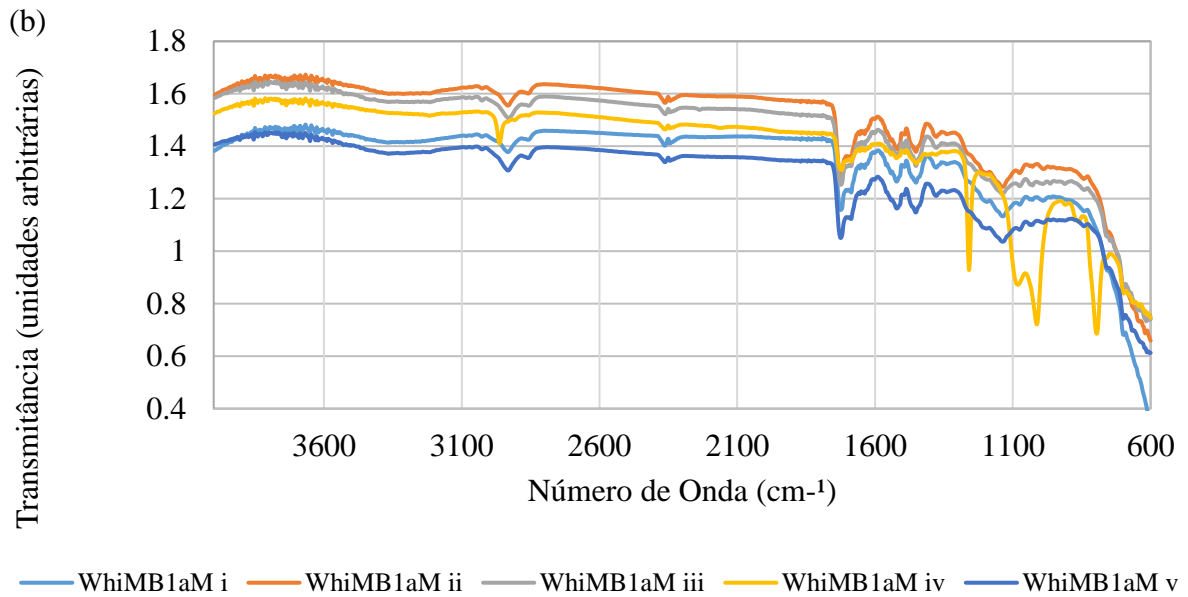
As amostras que apresentaram avarias em inspeção visual macroscópica foram: FI16a, FI22a, FI25b, FI26a, GM3a, GM4a, GM9b, GM10a, HO2a, HO9a, HY8a, RE1b, RE1c, CI1a, CI1b, MB1a, VW1a, VW2a, VW6a, VW7a, VW12a e VW14a. Tiveram mesmo assim os seus espectros obtidos com o uso do microscópio, o qual permitiu selecionar a região menos danificada da lasca de tinta.

Na etapa de comparação visual simultânea dos cinco espectros de uma mesma amostra, para a eliminação de espectros com o comportamento anômalo, foram eliminados 22 espectros, a saber: FI3b (1 espectro), FI4b (2), FI5c (1), FI15b (2), FI16a (2), FI22a (2), FI26a (2), FO3b (1), RE1b (2), AU1b (1), CI1b (1), CI1c (1), KI1a (1), MB1a (1), VW2a (1), VW6a (1). A

Figura 12 ilustra o exemplo de duas amostras: FI16a (que teve duas replicatas excluídas) e MB1a (que teve uma replicata excluída). Não houve eliminação de mais de dois espectros por amostra. Todas as amostras, nessa etapa, se mantiveram com pelo menos três replicatas válidas. Portanto, todas as amostras foram mantidas como amostras úteis.

Figura 12- Avaliação visual das cinco replicatas de espectros ATR/FTIR das amostras de tintas automotivas (a) WhiFI16aM: replicata WhiFI16aM ii (de cor laranja) e replicata WhiFI16aM v (de cor azul escura) apresentam comportamento anômalo, quando comparadas às demais replicatas da mesma amostra e (b) WhiMB1aM: replicata WhiMB1aM iv (de cor amarela) apresenta comportamento anômalo, quando comparada às demais replicatas da mesma amostra.





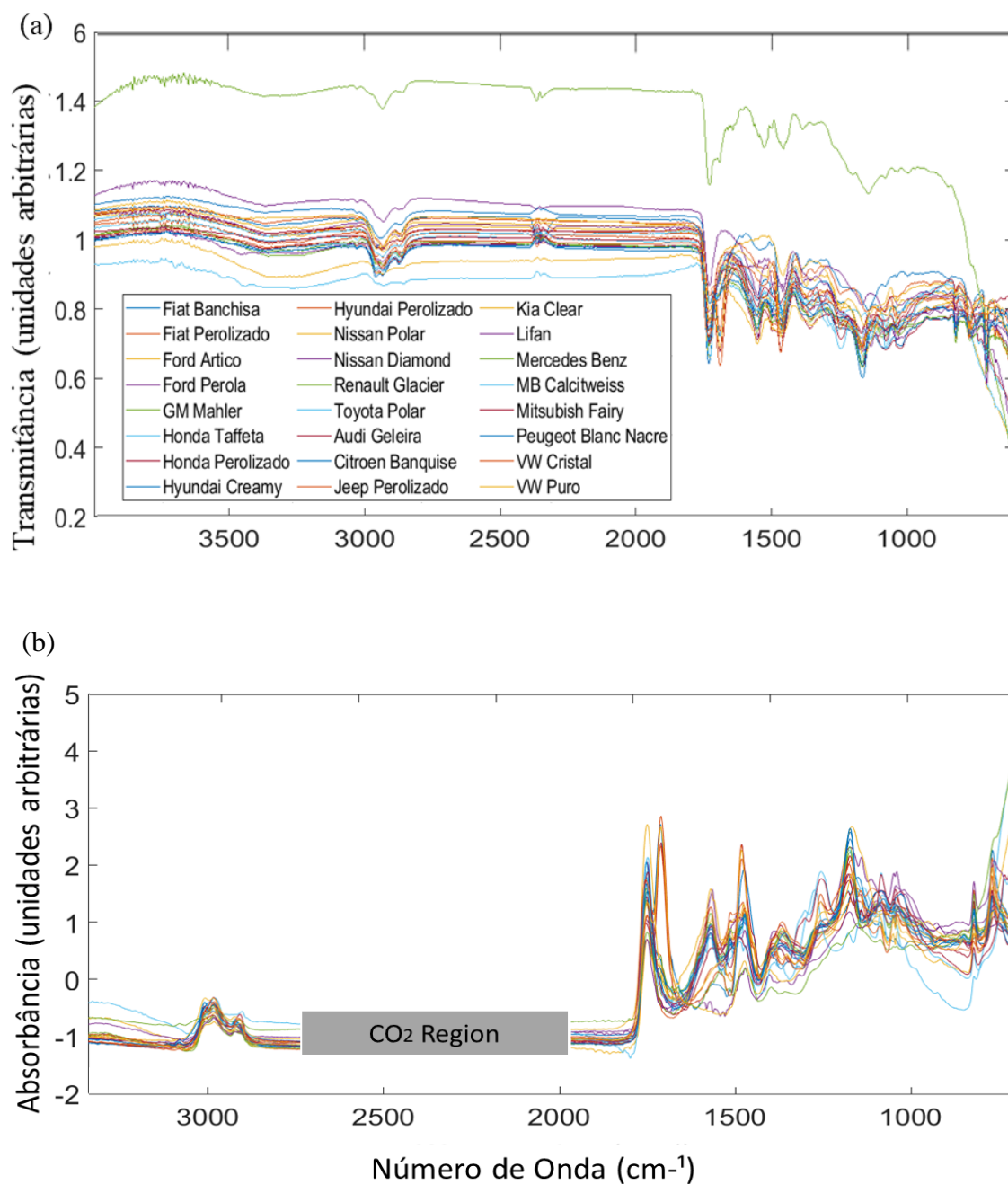
Fonte: autoria própria

Dos 715 espectros iniciais, 693 foram mantidos. A

Figura 13-a ilustra espectros brutos coletados, sendo um espectro representante de cada uma das 24 diferentes classes, apenas para visualização. A

Figura 13-b ilustra os mesmos espectros após a aplicação dos pré-processamentos: transformação de transmitância para absorbância, padronização normal de sinal (standard normal variate – SNV) e após a remoção das faixas não úteis (ou seja, após a remoção das extremidades, que apresentam ruídos característicos do equipamento e a remoção da faixa do espectro relacionada à concentração de dióxido de carbono (CO₂) no ambiente). Para o tratamento dos dados, foi também aplicado o pré-processamento centrar na média (*mean centering*). Entretanto, o gráfico foi gerado sem ele, apenas para facilitar a visualização dos espectros.

Figura 13– a) Espectros brutos ATR-FTIR de uma amostra de cada uma das 24 classes das tintas automotivas, b) espectros após os pré-processamentos SNV (standard normal variate) e transformação de transmitância para absorbância e remoção das faixas não úteis dos espectros.

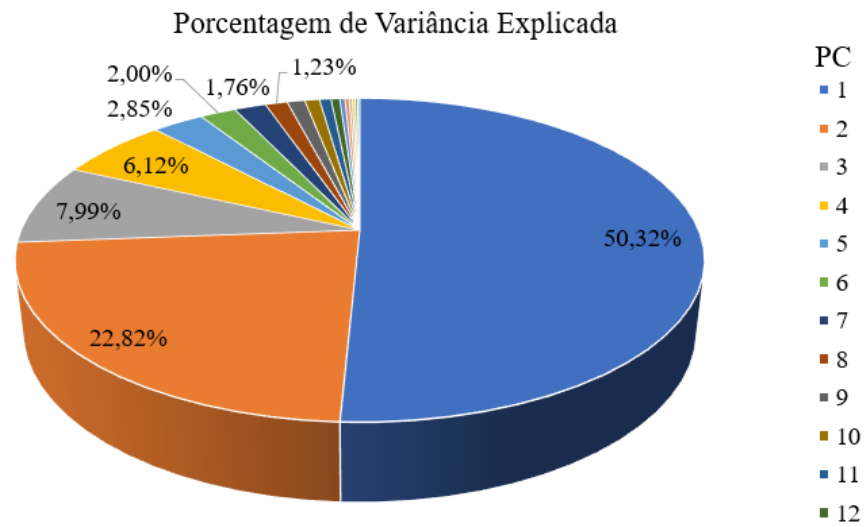


Fonte: autoria própria

4.1 RESULTADO DA ANÁLISE DE PCA

A primeira componente principal explica 50,32% da variância. A segunda explica 20,92% e a terceira, 7,99%, conforme ilustra a Figura 14. As três componentes principais totalizaram conjuntamente 79,23% de variância explicada.

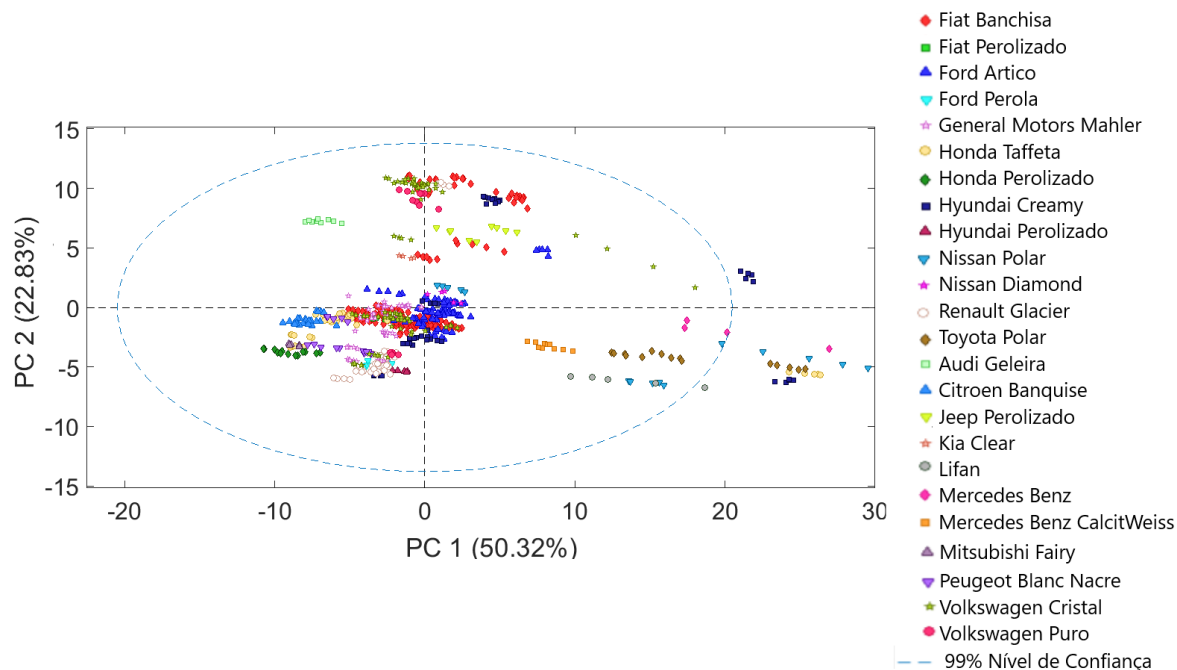
Figura 14 – Porcentagens de variância explicada por cada uma das componentes principais na análise de PCA.



Fonte: autoria própria.

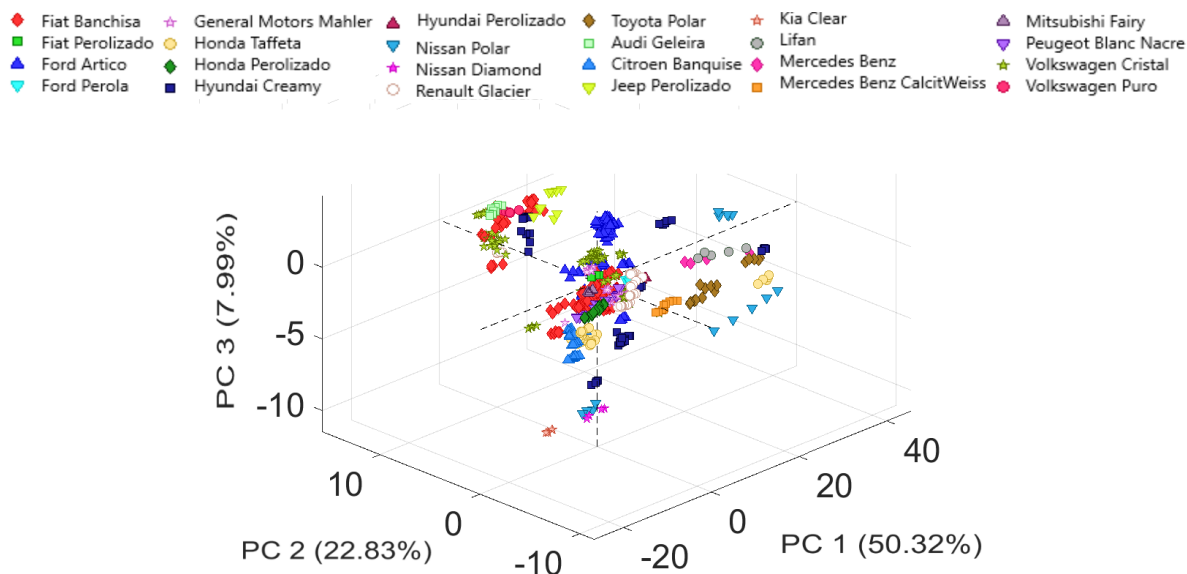
A Figura 15 ilustra a distribuição dos escores (ou seja, das coordenadas das amostras) em representação bidimensional de duas componentes principais (PC1 X PC2) e a Figura 16 ilustra a distribuição dos escores em representação tridimensional das três componentes principais (PC1 X PC2 X PC3).

Figura 15- Plotagem dos escores em sistema bidimensional da análise de componentes principais de tintas automotivas, com duas componentes principais.



Fonte: autoria própria.

Figura 16- Plotagem dos escores em sistema tridimensional de análise das componentes principais de tintas automotivas, com três componentes principais.



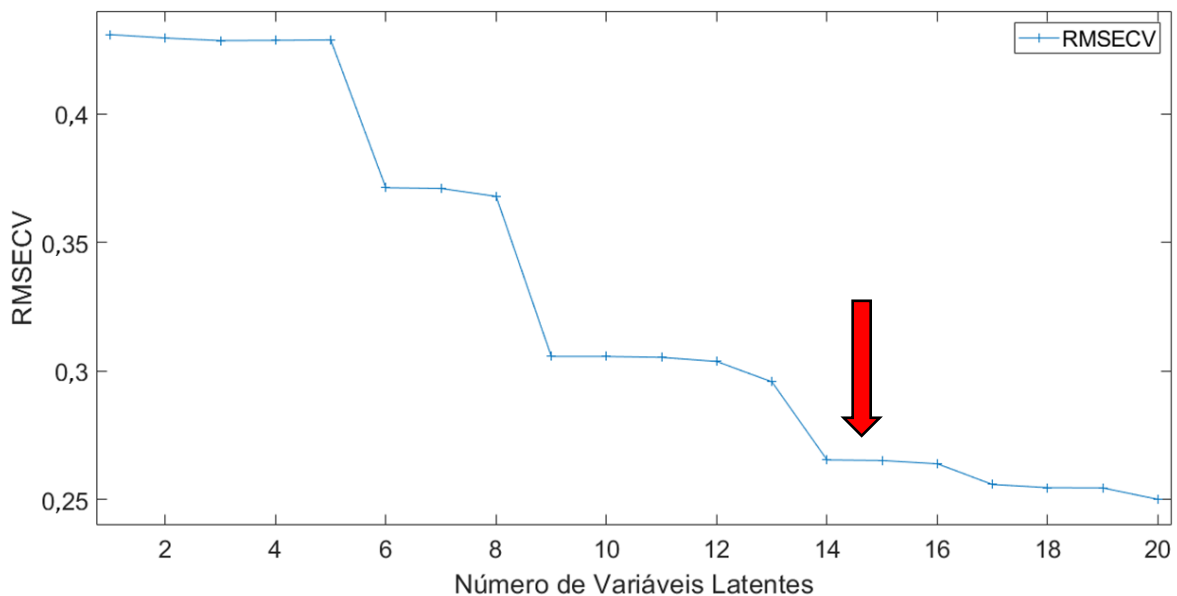
Fonte: autoria própria.

Não se observou a separação clara das classes nas plotagens usando-se duas componentes principais, nem com a utilização de três componentes principais, conforme ilustram os gráficos das figuras Figura 15 e Figura 16. Entretanto, essa separação parcial das amostras nos forneceu indícios de que os modelos de classificação supervisionados poderiam apresentar bons resultados.

4.2 RESULTADO DA ANÁLISE DE PLS-DA

Para a escolha do número de variáveis latentes, foi analisado o gráfico da raiz quadrada do erro médio quadrático de validação cruzada. A estabilização dos erros ocorre na 14ª variável latente, como ilustra o gráfico da Figura 17. Utilizamos 15 variáveis latentes nas análises. Embora isso possa ser considerado um alto número de variáveis latentes em algumas aplicações, é consistente com o número de diferentes classes predefinidas (oito) e as outras fontes de variação nas amostras (ou seja, mudanças ao longo do tempo devido às diferentes condições ambientais a que os veículos foram expostos e diferentes possibilidades de tratamentos recebidos pelos veículos, como polimentos e aplicações de películas).

Figura 17 – Raiz quadrada do erro médio quadrático de validação para a determinação do número de variáveis latentes a serem utilizadas no modelo PLS-DA, das classes principais. Os erros tendem a se estabilizar após 14 variáveis latentes.



Fonte: autoria própria.

Executando-se o modelo, foi realizada a remoção dos *outliers*, baseado no gráfico de *Hotelling T²* vs. resíduos *Q*, como ilustra a Figura 6. Os espectros contidos no quadrante assinalado, por apresentarem valores baixos de *T²* e *Q*, são espectros representativos do modelo. Optamos por excluir todos os demais como *outliers* (os contidos nos outros três quadrantes). Em etapa anterior, haviam sido excluídas 22 replicatas através de avaliação visual dos espectros. Somando-se os espectros removidos nessas duas etapas, houve a exclusão de 10 amostras, a saber: FI1b, FI3b, FI10a, FI16a, FI26a, FO16a, HO4a, HY1a, HY6a, VW2a. Como inicialmente havia 76 amostras no grupo de treinamento, restaram 66 amostras (330 espectros). Como houve uma amostra que teve apenas um espectro removido (amostra FO11a) e foi, portanto, mantida, restaram 329 espectros no grupo de calibração. Não ocorreu a exclusão de

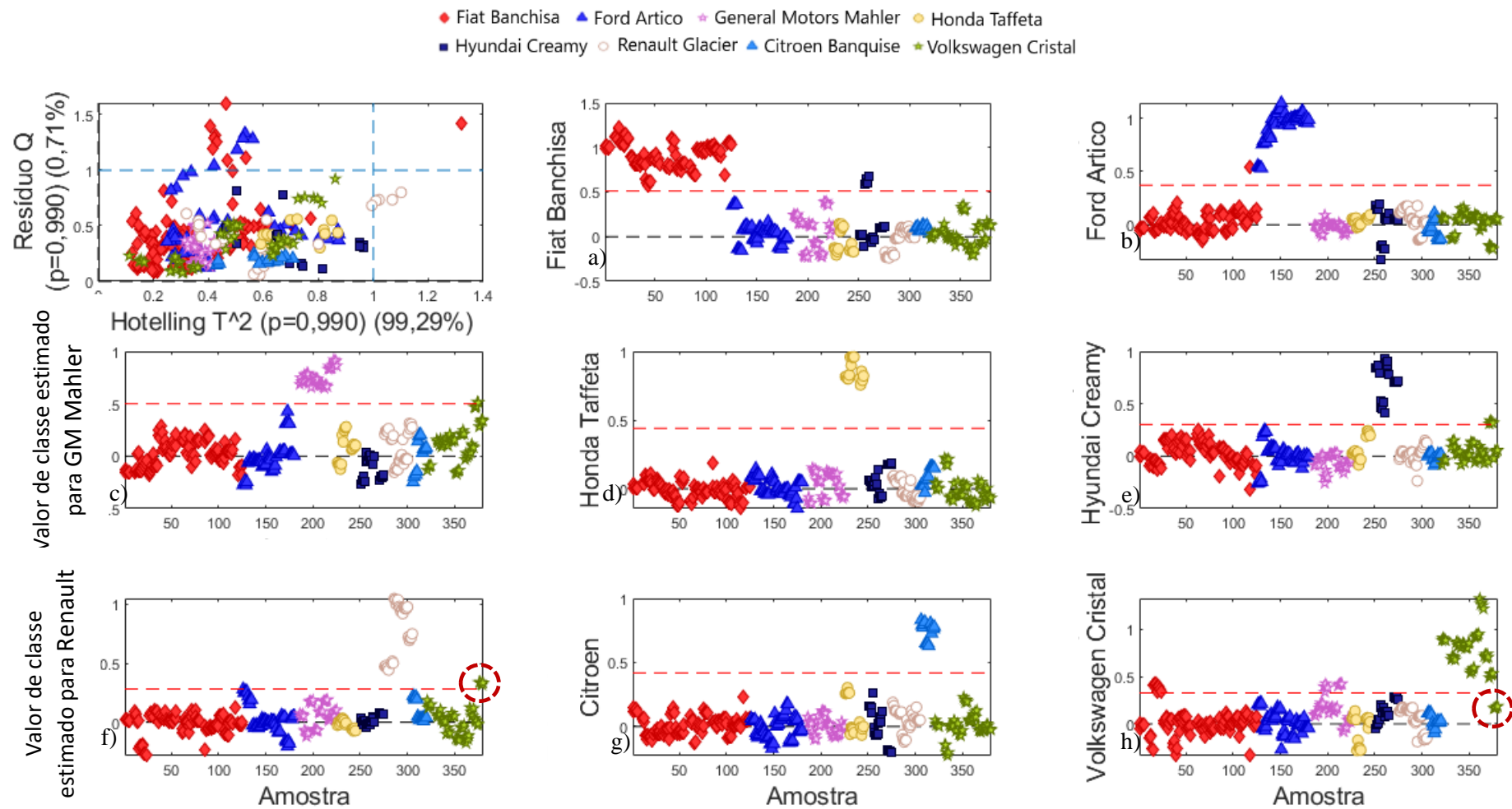
todas as amostras de uma mesma classe. Portanto, todas as classes predefinidas se mantiveram representadas.

A análise de cada classe, em separado, foi feita como ilustrada na Figura 7, relativa à análise da Classe Volkswagen Cristal. A seta ressalta a linha de discriminação entre a Classe Volkswagen Cristal e as demais classes, sendo que os espectros plotados acima da linha de discriminação são considerados da Classe Volkswagen Cristal e os espectros plotados abaixo da linha são considerados como não pertencentes à Classe Volkswagen Cristal. Os espectros sabidamente da Classe Volkswagen que aparecem acima da linha de discriminação são os verdadeiros positivos. Os espectros sabidamente da Classe Volkswagen Cristal que aparecem abaixo da linha de discriminação são os falsos negativos. Os espectros sabidamente não pertencentes à classe Volkswagen plotados acima da linha de discriminação são falsos positivos e os espectros sabidamente não pertencentes à classe Volkswagen plotados abaixo da linha de discriminação são verdadeiros negativos.

As análises de todas as classes estão ilustradas na Figura 18.

De acordo com a análise, as amostras que são tecnicamente classificadas em duas classes ou em nenhuma classe são análises inconclusivas. Seis amostras, nessa etapa, foram consideradas inconclusivas: FI3a, FI3b, GM4a, GM8a, HY4a, VW23a. Apenas uma amostra foi classificada erroneamente: a amostra VW25a, da classe Volkswagen Cristal, que foi considerada negativa para a sua própria classe (e, portanto, falso negativa) e foi classificada como pertencente à classe Renault Glacier. Está destacada na Figura 18 com círculos pontilhados. Tal amostra foi coletada de um veículo VW/Saveiro, Branco Geada, ano de fabricação 2017/ ano modelo 2018, e, portanto, não se tratava de um veículo muito velho. Ela não estava dentre as amostras que apresentavam avarias em inspeção visual macroscópica. Tampouco teve espectros excluídos por avaliação visual. Portanto, não foi possível atribuir o seu erro de classificação a algum motivo previsível. Por ter apresentado comportamento químico similar à classe Renault Glacier, apenas podemos aventar a possibilidade de ter passado por algum tipo de repintura ou camada extra de verniz, por exemplo, com formulação química semelhante.

Figura 18 – Calibração do modelo PLS-DA: *Hotelling T*² versus o resíduo Q após eliminação de *ouliers*. Análise de discriminação de espectros, para cada uma das classes principais na construção modelo PLS-DA: (a): Fiat Banchisa, (b): Ford Artico, (c): General Motors, (d): Honda Taffeta, (e): Hyundai Creamy, (f): Renault Glacier, (g): Citroen Banquise, (h): Volkswagen Cristal. Os círculos pontilhados destacam a amostra VW25a, que resultou como falso negativo para a própria classe e falso positivo para a classe Renault Glacier.

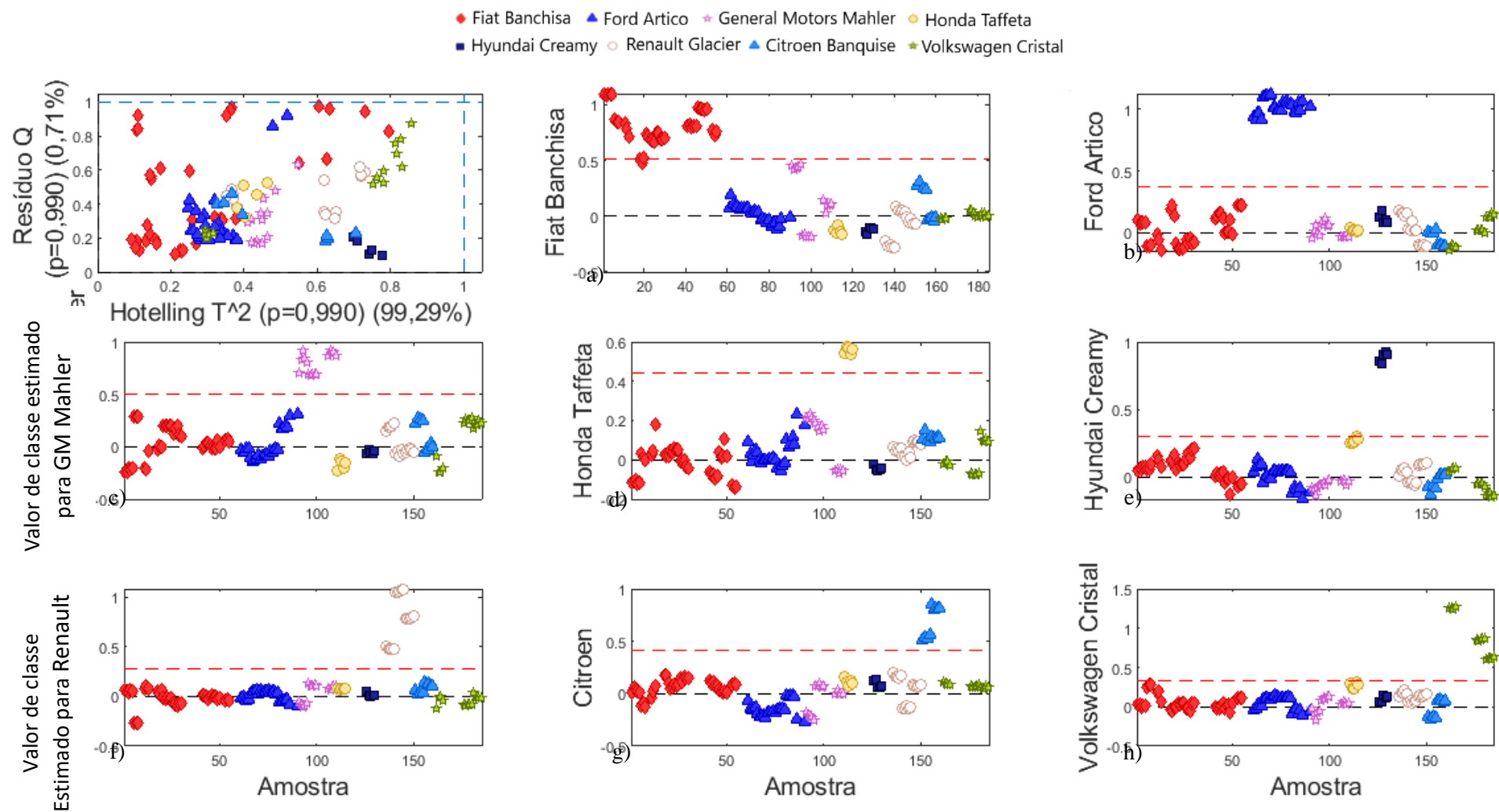


Fonte: autoria própria.

4.2.1 Validação do modelo

Após feita a otimização do modelo PLS-DA, as amostras de validação foram nele inseridas, para que fosse feita a predição dessas amostras que não participaram da etapa de calibração. O mesmo procedimento de remoção de *outliers* foi repetido, e a predição foi realizada novamente. Nessa etapa, nove amostras foram excluídas como *outliers*: FI13a, FI15b, GM9b, HO9a, HY3a, HY8a, VW7a, VW12a, FO17a. Não houve amostras classificadas erroneamente nessa etapa de validação. O resultado é ilustrado na Figura 19.

Figura 19 – Etapa de validação do modelo PLS-DA: *Hotelling T*² versus o resíduo Q após eliminação de *ouliers*. Análise de discriminação de espectros, para cada uma das classes principais na validação do modelo PLS-DA: (a): Fiat Banchisa, (b): Ford Artico, (c): General Motors, (d): Honda Taffeta, (e): Hyundai Creamy, (f): Renault Glacier, (g): Citroen Banquise, (h): Volkswagen Cristal.

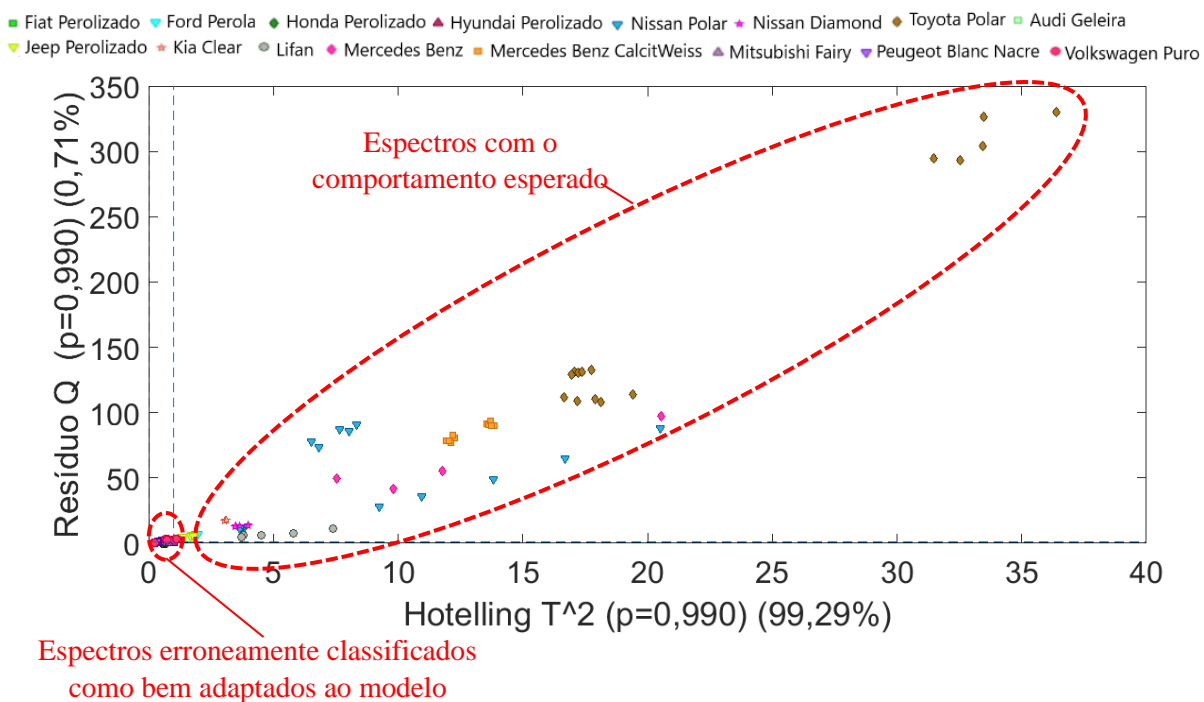


Fonte: autoria própria.

A etapa seguinte foi a predição das amostras separadas para teste. Nessa fase, como as amostras separadas são de classes diferentes das classes dos grupos de treinamento e de validação, o resultado esperado (e correto) é que todas se comportassem como *outliers*. As que não se comportaram como *outliers* é que são consideradas anômalas, nessa fase.

A imagem ampliada do gráfico de T^2 X Q permite observar que a maioria dos espectros apresentou o comportamento esperado (**Erro! Autoreferência de indicador não válida.**). Foram removidos, nessa etapa, os espectros que tiveram o comportamento esperado, de forma a avaliarmos apenas os demais.

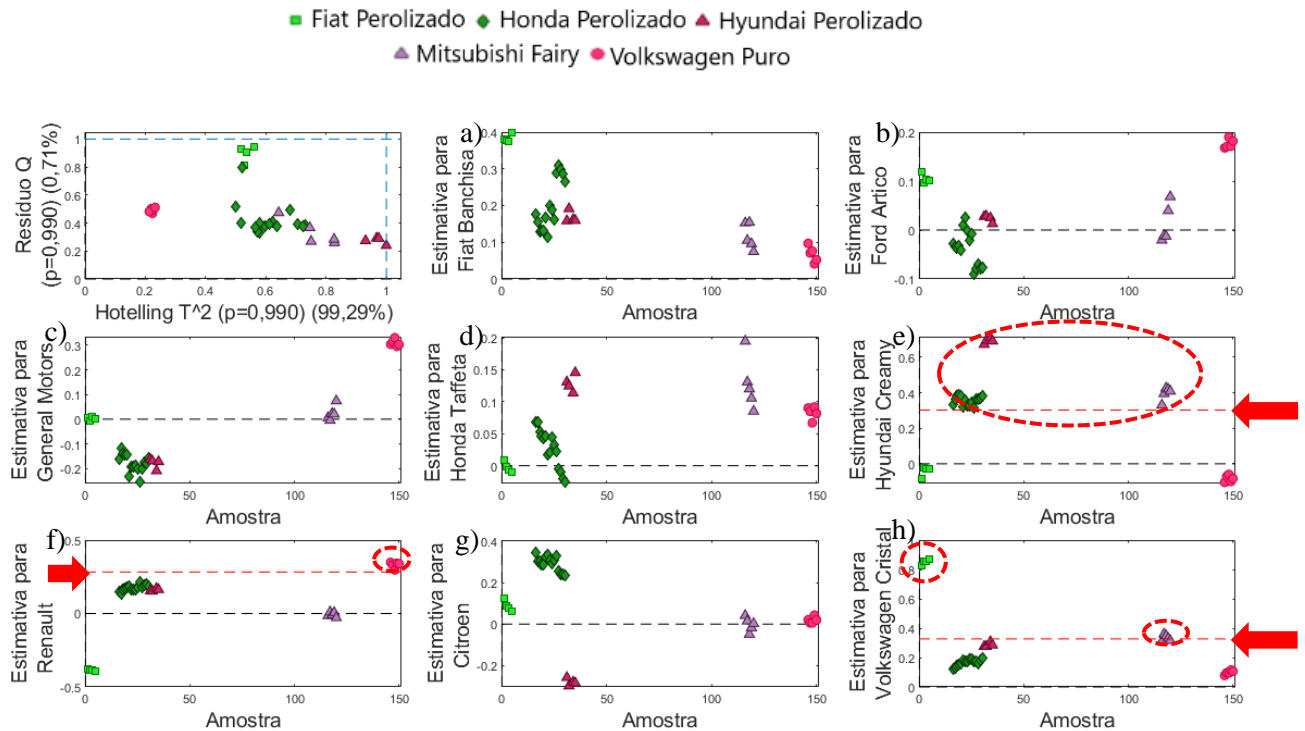
Figura 20 - Gráfico de *Hotelling* T^2 versus o resíduo Q da predição do grupo de teste externo pelo modelo PLS-DA. O espectros com valores elevados de T^2 e Q são considerados *outliers* (o comportamento esperado para esta etapa), e os com baixos valores, podem estar erroneamente classificados como bem adaptados ao modelo.



Fonte: autoria própria.

Após a remoção desses espectros e após a nova predição, procedeu-se à mesma avaliação feita na etapa anterior. Nessa fase, os espectros posicionados acima do limite de discriminação (linhas vermelhas, destacadas por setas) representam os espectros erroneamente classificados como positivos (falsos positivos) para a classe em análise. Tais espectros estão, na figura, destacados por linhas tracejadas). Os demais espectros (não destacados) encontram-se abaixo do limite de discriminação das classes e, portanto, não são considerados como falsos positivos. O resultado após a predição dos espectros do grupo teste no modelo é ilustrado na Figura 21.

Figura 21 – Etapa de predição dos espectros do grupo teste externo, no modelo PLS-DA: *Hotelling T²* versus o resíduo Q após eliminação de *ouliers*. Análise de discriminação de espectros do grupo teste externo para cada uma das classes principais, no modelo PLS-DA: (a): Fiat Banchisa, (b): Ford Ártico, (c): General Motors, (d): Honda Taffeta, (e): Hyundai Creamy, (f): Renault Glacier, (g): Citroen Banquise, (h): Volkswagen Cristal. As setas evidenciam as linhas de discriminação entre os espectros pertencentes ou não à respectiva classe e as elipses destacam os espectros posicionados acima das linhas de discriminação, e, portanto, passíveis de análise. Os demais tiveram o comportamento esperado.



Fonte: autoria própria.

Nessa etapa, uma amostra (MI1a) foi classificada em duas classes e, portanto, a sua análise foi inconclusiva. Seis amostras foram consideradas falsos positivos: a amostras FI23a, HO6a, HO6b, HO6c, HY5a, VW24a. A amostra FI23a (Fiat Perolizado) foi erroneamente classificada como Volkswagen Cristal. As amostras HO6a, HO6b, HO6c (Honda Perolizado) e HY5a (Hyundai Perolizado) foram erroneamente classificadas como pertencentes à classe Hyundai Creamy. A amostra VW24a (Volkswagen Gead) foi erroneamente classificada como Renault Glacier.

Devem-se fazer algumas considerações a respeito desses resultados. Das seis amostras de resultado errado, três delas vieram de um mesmo veículo (HO6a, HO6b e HO6c), um Honda Branco Perolizado. A consideração a ser feita é sobre o tipo de amostra com que trabalhamos. Os veículos examinados, por terem as mais diversas origens, possuem, cada um, diferentes características em relação ao desgaste, exposição a intempéries, graus de avarias, histórico de

limpeza, de polimentos da pintura e de aplicações de camadas protetoras. Portanto, é natural se supor que por exemplo esse veículo (HO6) possua alguma característica que o individualizou em relação aos demais. E as análises das três amostras relacionadas a esse veículo contribuíram em 50% para o resultado da taxa de classificações corretas dessa etapa, já que representou três amostras, dentre as seis erroneamente classificadas.

Ainda, dentre as 6 amostras erroneamente classificadas no grupo teste externo, uma delas (HY5a) tratava-se de um veículo de marca Hyundai, de cor “Branco Perolizado” (*Pearly White*). Ela se apresentou como falso positivo para a classe de Hyundai “Branco Creamy” (*Creamy White*). Como ocorreu em apenas uma amostra, não consideramos essa condição representativa. Mas isso nos faz ficarmos atentos, para os próximos estudos, para a ocorrência de falsos positivos entre veículos de marcas iguais. A análise espectral é realizada a partir da superfície da amostra. O feixe de radiação penetra aproximadamente entre 0,5 a 10 μm na amostra. A camada de verniz geralmente possui cerca de 40 μm de espessura (considerando-se a pintura nova). Pode-se inferir, portanto, que a camada de verniz é a que mais influencia na formação do espectro. Caso esse resultado passe a ser visto com maior frequência em estudos futuros (erros envolvendo veículos de diferentes tonalidades, mas de mesma marca), pode-se considerar a possibilidade de uso de formulação semelhante de verniz para diferentes tons de cores de uma mesma marca automotiva.

4.2.2 Figuras de mérito

Os resultados das figuras de mérito são apresentados na Tabela 5 (figuras de mérito referentes à etapa de calibração), Tabela 6 (figuras de mérito referentes à etapa de validação) e Tabela 7 (figuras de mérito referentes à etapa de teste). As tabelas de figuras de mérito das etapas de calibração e validação mostram, para a análise de cada uma das oito principais classes, o número de amostras, o número de outliers, o número de falsos positivos, de falsos negativos, de verdadeiros positivos, de verdadeiros negativos, as taxas de falsos positivos, de verdadeiros negativos, de sensibilidade, de especificidade e de eficiência do modelo. No caso da etapa de validação, há ainda a taxa de classificações corretas (não aplicável para a etapa de calibração).

Obteve-se para o modelo (levando-se em conta apenas os dados do grupo de calibração), taxa de sensibilidade de 98,9%, taxa de especificidade de 98,4% e taxa de eficiência de 98,6% (veja Tabela 5).

Tabela 5 – Figuras de mérito da calibração. NA = número de amostras; FP = número de falsos positivos; FN = número de falsos negativos; VP = número de verdadeiros positivos; VN = número de verdadeiros negativos; TFP = taxa de falsos positivos; TFN = taxa de falsos negativos; TST = taxa de sensibilidade; TSP = taxa de especificidade; TEF = taxa de eficiência. As taxas estão expressas em porcentagens.

	Fiat Banchisa	Ford Artico	GM Mahler	Honda Taffeta	Hyundai Creamy	Renault Glacier	Citroen Banquise	VW Cristal	Média
NA	25	12	8	4	6	6	3	12	9,5
Outlier	5	1	0	1	2	0	0	1	1,2
FP	1	0	1	0	1	1	0	3	0,8
FN	0	0	0	0	0	0	0	1	0,1
VP	20	11	8	3	4	6	3	10	8,1
VN	40	54	57	62	59	59	63	51	55,6
TFP	2,4	0	1,7	0	1,7	1,7	0	5,6	1,6
TFN	0	0	0	0	0	0	0	7,7	0,9
TSB	100	100	100	100	100	100	100	90,9	98,9
TSP	97,6	100	98,3	100	98,3	98,3	100	94,4	98,4
TEF	98,8	100	99,1	100	99,1	99,1	100	92,7	98,6

Fonte: autoria própria.

A taxa de resultados indefinidos foi de 13%, que é a taxa que leva em conta todas as amostras que se comportaram como *outliers* ou que foram consideradas inconclusivas na construção do modelo.

Obeve-se, para a predição das amostras de validação pelo modelo (levando-se em conta apenas os dados do grupo de validação), taxa de sensibilidade de 100%, taxa de especificidade de 100%, taxa de eficiência de 100%. Nessa etapa, existe também a figura de mérito da taxa de classificações corretas, já que é possível quantificar o percentual de acertos pelo modelo, das amostras que não participaram da construção do modelo (mas que são pertencentes às classes treinadas). A taxa de classificações corretas foi de 100% (veja Tabela 6).

A taxa de resultados indefinidos foi de 24% (*outliers* mais análises inconclusivas), durante a etapa de predição do grupo de validação.

Tabela 6 - Figuras de mérito da validação. NA = número de amostras; FP = número de falsos positivos; FN = número de falsos negativos; VP = número de verdadeiros positivos; VN = número de verdadeiros negativos; TFP = taxa de falsos positivos; TFN = taxa de falsos negativos; TST = taxa de sensibilidade; TSP = taxa de especificidade; TEF = taxa de eficiência; TCC = taxa de classificação correta. As taxas estão expressas em porcentagens.

	Fiat Banchisa	Ford Artico	GM Mahler	Honda Taffeta	Hyundai Creamy	Renault Glacier	Citroen Banquise	VW Cristal	Média
NA	12	6	4	2	3	3	2	5	4,6
Outlier	2	1	1	1	2	0	0	2	1,1
FP	0	0	0	0	0	0	0	0	0
FN	0	0	0	0	0	0	0	0	0
VP	10	5	3	1	1	3	2	3	3,5
VN	16	22	24	26	25	25	26	23	23,4
TFP	0	0	0	0	0	0	0	0	0
TFN	0	0	0	0	0	0	0	0	0
TST	100	100	100	100	100	100	100	100	100
TSP	100	100	100	100	100	100	100	100	100
TEF	100	100	100	100	100	100	100	100	100
TCC	100	100	100	100	100	100	100	100	100

Fonte: autoria própria.

O modelo se mostrou muito confiável para a predição de amostras pertencentes às oito classes principais, já que não apresentou quaisquer análises errôneas. Os resultados apresentados evidenciam que as oito classes principais, por serem as mais numerosas, estão melhor representadas e, portanto, puderam ser bem caracterizadas de forma a permitirem a criação de um modelo robusto.

Obteve-se, para predição das amostras de teste externo pelo modelo (levando-se em conta apenas os dados do grupo de teste), taxa de classificações corretas de 79,1%. As demais figuras de mérito não se aplicam a essa etapa, uma vez que se tratam de amostras cujas classes não foram treinadas. Portanto, não é possível haver verdadeiros positivos e nem falsos negativos (veja Tabela 7).

Tabela 7 – Figuras de mérito do teste externo. NA = número de amostras; FP = número de falsos positivos; TCC = taxa de classificação correta, expressa em porcentagem.

	Fiat Perol.	Ford Perola	Honda Perol.	Hyund. Perol.	Nissan Polar	Nissan Diam.	Toyota Polar	Audi Geleira	Jeep Perol.	Kia Clear	Lifan	M. Benz	M. Benz Calc.	Mitsub. Fairy	Peug. BNacre	VW Puro	Média
NA	1	2	3	1	3	1	3	2	2	1	1	1	2	1	3	3	1,9
FP	1	0	3	1	0	0	0	0	0	0	0	0	0	0	0	1	0,4
TCC	0	100	0	0	100	100	100	100	100	100	100	100	100	100	100	66,7	79,1

Fonte: autoria própria.

A taxa de classificações corretas da etapa de teste externo, de 79,1%, foi muito último inferior à mesma taxa obtida na etapa anterior (100%, na etapa de validação). Entretanto, há que se observar que essa etapa (a de teste externo) se refere às análises de amostras das classes sub representadas numericamente. Tal resultado nos mostrou que as análises podem ser muito aprimoradas, quando for possível aumentar o número amostral das classes menos numerosas. O aumento do número de amostras ou mesmo do número de classes poderá ser feito a qualquer tempo, de forma a tornar o modelo mais completo e, portanto, mais representativos de todas as classes.

A taxa de resultados indefinidos foi de 23% (*outliers* mais análises inconclusivas), durante a etapa de predição do grupo teste.

4.3 AVALIAÇÃO GERAL

Não foi observada uma relação entre as amostras que foram consideradas como avariadas, por inspeção visual (no início das análises) com as amostras *outliers*, inconclusivas ou erroneamente classificadas. Isso pode ter ocorrido pela utilização do microscópio acoplado ao equipamento de espectroscopia do infravermelho. A utilização deste acessório permite com que o pesquisador escolha os pontos de menor ocorrência de avarias na amostra, como marcas de fricção na pintura, ou mesmo áreas de visíveis desgastes naturais. Permite, desta forma, com que se escolham as áreas mais preservadas e, portanto, as coletas terão espectros de melhor qualidade, para as análises. Nossos resultados corroboram que o uso de tal acessório seja tão frequentemente relatado na literatura (17,39,46,47,100,101), para exemplificar, estando o seu uso inclusive listado em norma técnica (102) que padroniza as condutas de comparações e análises de tintas automotivas.

Não observamos uma correlação direta entre a idade dos veículos e o aumento do número de amostras *outliers*, inconclusivas ou erroneamente classificadas. Isso nos evidenciou que o intervalo de diferença de idade entre os veículos que escolhemos, de no máximo 10 anos, foi um intervalo que não potencializou os possíveis erros por degradação das amostras. Essa era uma preocupação, uma vez que a literatura traz evidências de haver uma mudança na composição química das amostras devido à exposição ambiental (16,44,80,81). Maric et al. (52) relatam inclusive migração de componentes entre as camadas de tintas, ao longo do tempo. Entretanto, o limite de idade das amostras selecionado se mostrou efetivo para evitar resultados inesperados com relação a esse fator. O intervalo de 10 anos nos deixa confortáveis em relação a estarmos trabalhando com cerca de metade da frota circulante no Distrito Federal, conforme informação do Detran DF, já relatada e referenciada neste trabalho.

A taxa de resultados indefinidos do grupo de calibração foi de 13%. A do grupo de validação foi de 24% e a do grupo teste foi de 23%. A média ponderada das taxas de resultados indefinidos (levando-se em consideração o número de amostras de cada grupo) foi de 18%, o que será, portanto, considerado como a taxa média de resultados indefinidos. Dos 380 espectros iniciais do grupo de calibração, restaram 329 espectros após a remoção de *outliers*, o que está dentro de valores estabelecidos por norma técnica (103). Ainda assim faremos uma consideração com relação a essa taxa. A taxa de análises indefinidas está relacionada aos critérios adotados para a definição do que é considerado *outlier*, baseados na estatística de *Hotelling T²* e resíduos *Q*, conforme já descrito. Decidimos definir como *outlier* aqueles espectros que tiveram alto valor de *Hotelling T²* ou alto valor de resíduos *Q* (e não “e”). Isso gera uma tendência de haver realmente mais análises inconclusivas, ao longo da criação e da validação do modelo. Por outro lado, a tendência de haver erros de análise passa a ser menor. Precisamos ter em mente que em Ciências Forenses, uma análise errada pode significar uma prisão injusta, de uma pessoa inocente, ou a absolvição errônea de um verdadeiro culpado. Portanto, privilegiamos o menor índice de erros em detrimento de uma taxa mais alta de análises indeterminadas.

Em análise de tintas automotivas, Lavine et al. (22) usaram análise semelhante à utilizada neste trabalho, com excelentes resultados. Entretanto, a comparação fica prejudicada pelo fato de terem se utilizado de outras ferramentas associadas, como algoritmos genéticos e coeficientes *wavelet*. Neste trabalho, as figuras de mérito do modelo, nas etapas de calibração e validação interna foram favoráveis, já que obtivemos taxa de sensibilidade de 98,9% para calibração e 100% para validação, taxa de especificidade de 98,4% para calibração e 100% para validação, taxa de eficiência de 98,6% para calibração e 100% para validação e taxa de classificações corretas de 100% para validação. Para a validação com o teste externo, a avaliação teve resultados inferiores. A taxa de classificações corretas foi de 79,1%. Faremos uma comparação com estudos forenses, que se utilizaram de método muito semelhante ou quase idêntico ao que utilizamos. Gautam et al. (104), em análise para discriminação de tintas de canetas de cor preta, com PLS-DA, conseguiram 100% de classificações corretas de 50 canetas analisadas. Martins et al. (97) relatam eficiência de 93,1% na discriminação de marcas de *whisky*, com o uso de PLS-DA. Em análise de falsificações, tiveram taxas de classificações corretas de 93% a 100%, dependendo da marca. Pereira et al. (105) construíram um modelo usando PLS-DA para a discriminação de cinco tipos de drogas psicoativas que geralmente possuem apresentações semelhantes, em comprimidos. O método, portanto, demonstra bons resultados para análises forenses, de maneira geral. Com relação à utilização do método de PLS-

DA para análise de tintas automotivas, não encontramos artigos que tivessem utilizado método idêntico ao que usamos. A publicação do presente trabalho seria, então, pioneira. Com relação aos nossos resultados, consideramos que foram favoráveis, já que obtivemos 100% de classificações corretas na primeira etapa de validação (a das amostras pertencentes às classes treinadas) e 79,1% na segunda etapa de validação (a de teste externo, que pode ser aprimorado pelo aumento do número amostral, quando houver a possibilidade).

Fazendo-se um **resumo dos resultados**, obtivemos êxito na confecção do modelo matemático de classificação baseado nos dados de espectroscopia por infravermelho (ATR/FTIR) das 143 amostras de tintas automotivas analisadas, de cor branca, de diferentes marcas, modelos e tons de cores, de veículos circulantes no Distrito Federal, com a utilização de análise de componentes principais (PCA) e de análise discriminante por mínimos quadrados parciais (PLS-DA). As amostras foram analisadas diretamente em sua superfície, sem a realização de preparos (exceto a limpeza com álcool), de forma rápida, limpa e não destrutiva. O intervalo de 10 anos de diferença de idade das amostras analisadas se mostrou suficiente para se evitar erros de classificações relacionados ao desgaste natural dos veículos. O microscópio acoplado se mostrou um acessório importante para a seleção de locais menos avariados das amostras, para se proceder às coletas dos espectros. O rigor nos critérios para a definição de *outliers* garante resultados mais confiáveis, uma vez que é preferível uma análise inconclusiva a uma errada, em se tratando de Ciências Forenses. A análise errada, nesse contexto, pode representar uma prisão ou uma soltura injusta. As taxas de sensibilidade, especificidade e eficiência do modelo de calibração foram, respectivamente, 98,9%, 98,4% e 98,6%, enquanto as do grupo de validação foram todas de 100%. A taxa de classificações corretas do grupo de validação interna foi de 100% e o do teste externo, 79,1%, sendo que essa última poderá ser aprimorada assim que for possível se fazer a inserção de novas amostras, para que haja melhor caracterização das respectivas classes.

4.4 PERSPECTIVAS FUTURAS

Uma vez que não se observou relação direta entre a idade dos veículos e os erros de classificação, o intervalo de idade das amostras utilizado (de no máximo 10 anos) se mostrou efetivo para evitar erros nas análises devido a mudanças na composição química das tintas. Em trabalhos futuros, iremos considerar a possibilidade de ampliarmos a diferença de idade dos veículos para 15 anos de forma a tentarmos abarcar um maior percentual da frota circulante nos novos modelos a serem construídos.

Deve ser feita uma consideração em relação ao número amostral utilizado e uma possível perspectiva futura. Nas etapas de calibração e validação, foram utilizadas as classes mais numerosas e, portanto, mais bem representadas numericamente. Na etapa de validação pelo teste externo, foram usadas as classes sub representadas numericamente, uma vez que eram classes com no máximo três amostras. Esse resultado evidenciou claramente a importância do alto número amostral nesse tipo de análise. Quanto maior o número de amostras pertencentes à mesma classe, utilizadas na etapa de calibração, maior a possibilidade de a modelagem matemática caracterizar bem aquela classe. Classes numerosas reúnem melhor as diferentes informações que porventura existam dentro da classe. Pequenas variações entre as amostras de uma mesma classe existem e são aceitáveis. O conhecimento dessas pequenas variações é necessário para a caracterização da classe de forma abrangente. Isso nos deixa confiantes de que quando for possível se elevar o número amostral de todas as classes, será possível aumentar o número de classes treinadas e, portanto, o exame de amostras desconhecidas ocorrerá de forma muito mais confiável.

O aumento do número amostral, entretanto, está vinculado à coleta de mais fragmentos de tinta. Todas as coletas de amostras utilizadas nesse trabalho, conforme já descrito, foram feitas quando da realização de perícias em veículos que sofreram colisões. Para o aumento do número amostral, essa coleta precisa ser intensificada, e preferencialmente acrescentada em protocolos de exames em veículos, ou seja, de forma que elas sejam feitas por padrão, pelo perito que faça o exame pericial veicular, sempre que o veículo se enquadrar nos critérios de interesse.

Consideramos, para estudos futuros, análises envolvendo tecnologias portáteis análogas ou similares à que foi utilizada neste trabalho. O emprego de equipamento portátil, se usado no local do crime, representa a dinamização de resultados das análises, ainda que preliminares, o que certamente traz um aumento imensurável das chances de se concluir com sucesso uma investigação de acidente com veículo evasor. Mas existe uma aplicação muito importante com relação às técnicas portáteis, que merecem uma consideração. Uma vez que fosse possível repetir a análise feita neste trabalho com o emprego de um equipamento portátil, de forma a se validar a metodologia para a construção do modelo com o uso de equipamento portátil, o pesquisador/operador poderia acumular, de forma muito mais rápida, informações espectrais de novos veículos, já que isso permitiria a aquisição de espectros de veículos íntegros. Não seria necessário que as amostras estivessem em formato de fragmentos (que é o que atualmente permite que elas sejam analisadas nos equipamentos de mesa). As coletas poderiam ser feitas

em estacionamentos públicos, em concessionárias, em oficinas mecânicas, por exemplo, desde que a técnica selecionada não fosse invasiva ou destrutiva.

A ideia, para estudos futuros, de utilização de tecnologias portáteis, portanto, além de dar mais celeridade às investigações criminais, permitiria o aumento do número amostral de forma rápida, já que permite a coleta de espectros a partir de veículos íntegros, e não apenas de fragmentos coletados em locais de acidentes de trânsito.

5 CONCLUSÃO

Foi construído um modelo matemático a partir de espectros de amostras de fragmentos de tintas automotivas de cor branca, adquiridos por espectroscopia infravermelho, que 1) se mostrou capaz de classificar o primeiro grupo de amostras, levando em consideração a marca de cada veículo e a tonalidade da sua cor, 2) permitiu a predição correta de 100% de novas amostras, quando aplicadas ao modelo, amostras tais pertencentes às mesmas classes (marcas e tonalidades) inicialmente analisadas, e 3) foi capaz de identificar corretamente como *outliers* 79,1% das amostras de um terceiro grupo, amostras tais não pertencentes às classes utilizadas para a construção do modelo. Trata-se de um modelo promissor para a utilização forense, para se iniciar a identificação de um veículo evasor de acidente de trânsito.

REFERÊNCIAS

1. Brasil. Lei no9.503/1997. Código de Trânsito Brasileiro, Brasília, 23 de setembro de 1997.
2. Brasil. Supremo Tribunal Federal. Recurso Extraordinário nº 971959/RS – Rio Grande do Sul. Relator: Ministro Luiz Fux. 31 de julho de 2020. <https://jurisprudencia.stf.jus.br/pages/search/sjur428870/false>
3. Beccaria C. Dos Delitos e das Penas. 12th ed. Rio de Janeiro: Ediouro; 1999. 154 p.
4. Wright DM. Sourcing paint smears: A hate crime highlights the utility of the paint data query (PDQ) database. *J Can Soc Forensic Sci.* 2012; 45(2): 79–88. <https://doi.org/10.1080/00085030.2012.10757185>.
5. Jastee. Standard Guide for Using Infrared Spectroscopy in Forensic Paint Examinations Scientific Working Group on Materials Analysis (SWGMA). 2000; 2(1): 73–87. <http://www.fbi.gov/about->.
6. The Royal Canadian Mounted Police [homepage na internet]. Paint Data Query. 2016. p. 32. https://www.nist.gov/system/files/documents/2016/11/04/hodgkins_-_pdq_-_intl_forensic_automotive_db.pdf
7. FBI. Laboratory Services [homepage na internet]. <https://www.fbi.gov/services>.
8. Lavine BK, White CG, Ding T. Library Search Prefilters for Vehicle Manufacturers to Assist in the Forensic Examination of Automotive Paints. *Appl Spectrosc.* 2018 Mar; 72(3): 476–88. <https://doi.org/10.1177/0003702817737787>.
9. Lavine BK, White CG, Allen MD. Forensic analysis of automotive paints using a pattern recognition assisted infrared library searching system: Ford (2000–2006). *Microchem J.* 2016 Nov; 129: 173–83. <https://doi.org/10.1016/j.microc.2016.06.024>.
10. Muehlethaler C, Gueissaz L, Massonnet G. Forensic Paint Analysis. In: *Encyclopedia of Forensic Sciences: Second Edition.* Université de Lausanne, Lausanne, Switzerland: Elsevier Inc.; 2012. p. 265–72. doi:10.1016/B978-0-12-382165-2.00109-4.
11. Khandasammy SR, Fikiet MA, Mistek E, Ahmed Y, Halamkova L, Bueno J, et al. Bloodstains, paintings, and drugs: Raman spectroscopy applications in forensic science. *Forensic chem.* 2018; 8: 111–33. <https://doi.org/10.1016/j.forc.2018.02.002>.
12. Doty KC, Muro CK, Bueno J, Halamkova L, Lednev IK. What can Raman spectroscopy do for criminalistics? *J Raman Spectrosc.* 2016; 47(1): 39–50. <https://doi.org/10.1002/jrs.4826>.
13. Edelman GJ, Gaston E, van Leeuwen TG, Cullen PJ, Aalders MCG. Hyperspectral imaging for non-contact analysis of forensic traces. *Forensic Sci Int.* 2012; 223(1): 28–39. <https://doi.org/10.1016/j.forsciint.2012.09.012>.
14. Manley M. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chem Soc Rev.* 2014; 43(24): 8200–14. <https://doi.org/10.1039/C4CS00062E>.

15. Maric M, van Bronswijk W, Lewis SW, Pitts K. Rapid characterisation and classification of automotive clear coats by attenuated total reflectance infrared spectroscopy. *Anal Methods*. 2012; 4(9): 2687–93. <https://doi.org/10.1039/C2AY25419K>.
16. van der Pal KJ, Sauzier G, Maric M, van Bronswijk W, Pitts K, Lewis SW. The effect of environmental degradation on the characterisation of automotive clear coats by infrared spectroscopy. *Talanta*. 2016; 148: 715–20. <https://doi.org/10.1016/j.talanta.2015.08.058>.
17. Martyna A, Michalska A, Zadora G. Interpretation of FTIR spectra of polymers and Raman spectra of car paints by means of likelihood ratio approach supported by wavelet transform for reducing data dimensionality. *Anal Bioanal Chem*. 2015; 407(12): 3357–76. <https://doi.org/10.1007/s00216-015-8558-9>.
18. Martyna A, Zadora G, Neocleous T, Michalska A, Dean N. Hybrid approach combining chemometrics and likelihood ratio framework for reporting the evidential value of spectra. *Anal Chim Acta*. 2016; 931: 34–46. <https://doi.org/10.1016/j.aca.2016.05.016>.
19. Lavine BK, Fasasi A, Mirjankar N, Sandercock MP. Development of search prefilters for infrared library searching of clear coat paint smears. *Talanta*. 2014; 119: 331–40. <https://doi.org/10.1016/j.talanta.2013.10.066>.
20. Lavine BK, Fasasi A, Mirjankar N, White C, Mehta J. Search prefilters for library matching of infrared spectra in the PDQ database using the autocorrelation transformation. *Microchem J*. 2014 Mar; 113: 30–5. <https://doi.org/10.1016/j.microc.2013.11.001>.
21. Lavine BK, Fasasi A, Sandercock MP. Improving PDQ database search strategies to enhance investigative lead information for automotive paints. *Microchem J*. 2014 Nov; 117: 133–7. <https://doi.org/10.1016/j.microc.2014.06.007>.
22. Lavine BK, Fasasi A, Mirjankar N, Sandercock M, Brown SD. Search prefilters for mid-infrared absorbance spectra of clear coat automotive paint smears using stacked and linear classifiers. *J Chemom*. 2014 May; 28(5): 385–94. <https://doi.org/10.1002/cem.2598>.
23. Ferreira KB, Oliveira AGG, Goncalves AS, Gomes JA. Evaluation of Hyperspectral Imaging Visible/Near Infrared Spectroscopy as a forensic tool for automotive paint distinction. *Forensic Chem*. 2017; 5: 46–52. <https://doi.org/10.1016/j.forc.2017.06.001>.
24. Maric M, van Bronswijk W, Pitts K, Lewis SW. Characterisation and classification of automotive clear coats with Raman spectroscopy and chemometrics for forensic purposes. *J Raman Spectrosc*. 2016; 47(8): 948–55. <https://doi.org/10.1002/jrs.4925>.
25. Lambert D, Muehlethaler C, Gueissaz L, Massonnet G. Raman analysis of multilayer automotive paints in forensic science: measurement variability and depth profile. *J Raman Spectrosc*. 2014 Nov; 45(11–12): 1285–92. <https://doi.org/10.1002/jrs.4490>.

26. Ferreira KB, Oliveira AGG, Gomes JA. Raman spectroscopy of automotive paints: Forensic analysis of variability and spectral quality. *Spectrosc Lett.* 2017; 50(2): 102–10. <https://doi.org/10.1039/C7AY02684F>.
27. Maric M, Marano J, Cody RB, Bridge C. DART-MS: A New Analytical Technique for Forensic Paint Analysis. *Anal Chem.* 2018 Jun; 90(11): 6877–84. <https://doi.org/10.1021/acs.analchem.8b01067>.
28. Muramoto S, Gillen G, Windsor ES. Chemical discrimination of multilayered paint cross sections for potential forensic applications using time-of-flight secondary ion mass spectrometry. *Surf Interface Anal.* 2018; 50(9): 889–96. <https://doi.org/10.1002/sia.6509>.
29. Asfaw A, Wibetoe G, Beauchemin D. Solid sampling electrothermal vaporization inductively coupled plasma optical emission spectrometry for discrimination of automotive paint samples in forensic analysis. *J Anal at Spectrom.* 2012 Nov; 27(11): 1928–34. <https://doi.org/10.1039/C2JA30193H>.
30. Yang S-H, Shen JY, Chang MS, Wu GJ. Differentiation of vehicle top coating paints using pyrolysis-gas chromatography/mass spectrometry and multivariate chemometrics with statistical comparisons. *Anal Methods.* 2015; 7(4): 1527–34. <https://doi.org/10.1039/C4AY02730B>.
31. Michalska A, Martyna A, Zadora G. Investigation of various factors influencing Raman spectra interpretation with the use of likelihood ratio approach. *Forensic Sci Int.* 2018 Jan; 282: 60–73. <https://doi.org/10.1016/j.forsciint.2017.10.034>.
32. Kumar R, Sharma V. Chemometrics in forensic science. *TrAC - Trends Anal Chem.* 2018 May; 105: 191–201. <https://doi.org/10.1016/j.trac.2018.05.010>.
33. Vandenaabeele P, Edwards HGM, Jehlicka J. The role of mobile instrumentation in novel applications of Raman spectroscopy: Archaeometry, geosciences, and forensics. *Chem Soc Rev.* 2014; 43(8): 2628–49. <https://doi.org/10.1039/C3CS60263J>.
34. Lavine BK, Fasasi A, Mirjankar N, White C, Sandercock MP. Search prefilters to assist in library searching of infrared spectra of automotive clear coats. *Talanta.* 2015 Jan; 132: 182–90. <https://doi.org/10.1016/j.talanta.2014.08.061>.
35. Lavine BK, White CG, Allen MD, Weakley A. Pattern Recognition-Assisted Infrared Library Searching of the Paint Data Query Database to Enhance Lead Information from Automotive Paint Trace Evidence. *Appl Spectrosc.* 2017 Mar; 71(3): 480–95. <https://doi.org/10.1177/0003702816666287>.
36. Lavine BK, White CG, Allen MD, Fasasi A. Improving Investigative Lead Information in the Forensic Examination of Automotive Paints. In: Lavine, BK and Brown, SD and Booksh K. 40 Years of Chemometrics - From Bruce Kowalski To the Future. 1155 Sixteenth St Nw, Washington, Dc 20036 USA: Amer Chemical Soc; 2015. P. 195–220.
37. Lavine BK, White CG, Allen MD, Fasasi A, Weakley A. Evidential significance of automotive paint trace evidence using a pattern recognition based infrared library

- search engine for the Paint Data Query Forensic Database. *Talanta*. 2016; 159: 317–29. <https://doi.org/10.1016/j.talanta.2016.06.035>.
38. Perera UDN, Nishikida K, Lavine BK. Development of Infrared Library Search Prefilters for Automotive Clear Coats from Simulated Attenuated Total Reflection (ATR) Spectra. *Appl Spectrosc*. 2018 Jun; 72(6): 886–95. <https://doi.org/10.1177/0003702818759664>.
 39. Maric M, van Bronswijk W, Lewis SW, Pitts K. Synchrotron FTIR characterisation of automotive primer surfacer paint coatings for forensic purposes. *Talanta*. 2014 Jan; 118: 156–61. <https://doi.org/10.1016/j.talanta.2013.10.016>.
 40. Streitberger HJ, Dossel KF. *Automotive paints and coatings*. Vol. 1002. Wiley Online Library; 2008.
 41. Bell S. *Forensic chemistry*. Pearson Prentice Hall New Jersey; 2006.
 42. Bentley J, Willis S, McCullough J, McDermott S. In *Forensic Examination of Glass and Paint*. Taylor and Francis: London; 2001.
 43. Suzuki EM. Infrared Spectra of U.S. Automobile Original Finishes (1998-2000). IX. Identification of Bismuth Oxychloride and Silver/White Mica Pearlescent Pigments Using Extended Range FT-IR Spectroscopy, XRF Spectrometry, and SEM/EDS Analysis. *J Forensic Sci*. 2014 Sep; 59(5): 1205–25. <https://doi.org/10.1111/1556-4029.12414>.
 44. Oliveira AGG, Wiercigroch E, Gomes JA, Malek K. Infrared and Raman spectroscopy of automotive paints for forensic identification of natural weathering. *Anal Methods*. 2018 Mar; 10(10): 1203–12. <https://doi.org/10.1039/C7AY02684F>.
 45. Duarte JM, Sales NGS, Sousa MH, Bridge C, Maric M, Gomes J de A. Automotive paint analysis: How far has science advanced in the last ten years? *Trends Analyt Chem*. 2020 Nov; 132: 116061. <https://doi.org/10.1016/j.trac.2020.116061>.
 46. Kwofie F, Perera UDN, Allen MD, Lavine BK. Transmission infrared imaging microscopy and multivariate curve resolution applied to the forensic examination of automotive paints. *Talanta*. 2018; 186: 662–9. <https://doi.org/10.1016/j.talanta.2018.02.025>.
 47. Zieba-Palus J, Trzcinska BM. Application of infrared and Raman spectroscopy in paint trace examination. *J Forensic Sci*. 2013 Sep; 58(5): 1359–63. <https://doi.org/10.1111/1556-4029.12183>.
 48. Lv J, Liu S, Feng J, Liu Y, Zhou S, Chen R. Effective identification of paints pigments in hit-and-run cases with confocal Raman microscope. *Pigment Resin Technol*. 2016; 45(4): 294–300. <https://doi.org/10.1108/PRT-05-2015-0044>.
 49. Lv J, Zhang W, Liu S, Chen R, Feng J, Zhou S, et al. Analysis of 52 automotive coating samples for forensic purposes with Fourier transform infrared spectroscopy (FTIR) and Raman microscopy. *Environ Forensics*. 2016 Jan 2; 17(1): 59–67. <https://doi.org/10.1080/15275922.2015.1091403>.

50. Yang S-H, Shen JY, Chang MS, Wu GJ. Quantification of vehicle paint components containing polystyrene using pyrolysis-gas chromatography/mass spectrometry. *Anal Methods*. 2012 Jul; 4(7): 1989–95. <https://doi.org/10.1039/C2AY05809J>.
51. Chen TH, Wu SP. Forensic applications of direct analysis in real time (DART) coupled to Q-orbitrap tandem mass spectrometry for the in situ analysis of pigments from paint evidence. *Forensic Sci Int*. 2017; 277: 179–87. <https://doi.org/10.1016/j.forsciint.2017.06.001>.
52. Maric M, van Bronswijk W, Lewis SW, Pitts K, Martin DE. Characterisation of chemical component migration in automotive paint by synchrotron infrared imaging. *Forensic Sci Int*. 2013; 228(1–3): 165–9. <https://doi.org/10.1016/j.forsciint.2013.01.032>.
53. Malek MA, Nakazawa T, Kang HW, Tsuji K, Ro CU. Multi-Modal Compositional Analysis of Layered Paint Chips of Automobiles by the Combined Application of ATR-FTIR Imaging, Raman Microspectrometry, and SEM/EDX. *Molecules*. 2019; 24(7). <https://doi.org/10.3390/molecules24071381>.
54. Suzuki EM. Infrared Spectra of U. S. Automobile Original Finishes (Post-1989). VIII: In Situ Identification of Bismuth Vanadate Using Extended Range FT-IR Spectroscopy, Raman Spectroscopy, and X-Ray Fluorescence Spectrometry. *J Forensic Sci*. 2014 Mar; 59 (2): 344–63. <https://doi.org/10.1111/1556-4029.12318>.
55. Palenik CS, Groves E, Insana J, Palenik S. Locating, Identifying, and Comparing Sub-visible Paint Particles. *J Forensic Sci*. 2019 Nov 22; 64(6): 1851–8. <https://doi.org/10.1111/1556-4029.14062>.
56. Kruglak KJ, Dubnicka M, Kammrath B, Maxwell V, Reffner JA. The Evidentiary Significance of Automotive Paint from the Northeast: A Study of Red Paint. *J Forensic Sci*. 2019 Sep 21; 64(5): 1345–58. <https://doi.org/10.1111/1556-4029.14007>.
57. Lv J, Feng J, Liu Y, Wang Z, Zhao M, Shi R. Discriminating amino resin paints from alkyd resin paints with two kinds of pigments in automotive coatings in forensic analysis by FTIR spectroscopy. *Anal Lett*. 2012; 45(9): 1079–85. <https://doi.org/10.1080/00032719.2012.670795>.
58. Trzcinska BM. A microtrace of great importance - Case study. *Z Zagadnien Nauk Sadowych*. 2013; 95: 670–7.
59. Lavine BK, Fasasi A, Mirjankar N, Nishikida K, Campbell J. Simulation of Attenuated Total Reflection Infrared Absorbance Spectra: Applications to Automotive Clear Coat Forensic Analysis. *Appl Spectrosc*. 2014 May; 68(5): 608–15. <https://doi.org/10.1366/13-07266>.
60. Zieba-Palus J. Examination of the variation of chemical composition and structure of paint within a car body by FT-IR and Raman spectroscopies. *J Mol Struct*. 2020; 1219: 128558. <https://doi.org/10.1016/j.molstruc.2020.128558>.
61. Kwofie F, Perera UDN, Allen MD, Lavine BK. Application of infrared microscopy and alternating least squares to the forensic analysis of automotive paint chips. *J Chemom*. 2021; 35(1): 3277. <https://doi.org/10.1002/cem.3277>.

62. Zieba-Palus J, Was-Gubala J. An investigation into the use of micro-Raman spectroscopy for the analysis of car paints and single textile fibres. *J Mol Struct.* 2011; 993(1–3): 127–33. <https://doi.org/10.1016/j.molstruc.2010.12.041>.
63. Zieba-Palus J, Michalska A. Characterization of Blue Pigments Used in Automotive Paints by Raman Spectroscopy. *J Forensic Sci.* 2014 Jul; 59(4): 943–9. <https://doi.org/10.1111/1556-4029.12499>.
64. Zieba-Palus J, Michalska A. Photobleaching as a useful technique in reducing of fluorescence in Raman spectra of blue automobile paint samples. *Vib Spectrosc.* 2014; 74: 6–12. <https://doi.org/10.1016/j.vibspec.2014.06.007>.
65. Michalska A, Martyna A, Zieba-Palus J, Zadora G. Application of a likelihood ratio approach in solving a comparison problem of Raman spectra recorded for blue automotive paints. *J Raman Spectrosc.* 2015; 46(9): 772–83. <https://doi.org/10.1002/jrs.4719>.
66. Gutierrez-Fallas D, Montero-Zeledon E, Murillo-Quiros N, Rojas-Rojas L. Analysis of automotive paint fragments used as forensic evidence by Raman spectroscopy. *Tecnol en Marcha.* 2016 Nov; 29(SI): 57–66. <https://doi.org/10.18845/tm.v29i8.2985>.
67. Rojas-Rojas L, Murillo-Quiros N, Gutierrez-Fallas D, Montero-Zeledon E. Raman spectroscopy analysis of automotive paint for forensic purposes. *Tecnol en Marcha.* 2017; 30(2): 3–13. <https://doi.org/10.18845/tm.v30i2.3191>.
68. Bell SEJ, Stewart SP, Armstrong WJ. Raman Spectroscopy for Forensic Analysis of Household and Automotive Paints. In: *Infrared and Raman Spectroscopy in Forensic Science*. School of Chemistry and Chemical Engineering, Queen’s University, Belfast BT9 5AG, United Kingdom: John Wiley and Sons; 2012: 121–35.
69. Buzzini P, Suzuki E. Forensic applications of Raman spectroscopy for the in situ analyses of pigments and dyes in ink and paint evidence. *J Raman Spectrosc.* 2016 Jan; 47(1, SI): 16–27. <https://doi.org/10.1002/jrs.4818>.
70. Sabaradin RZR, Osman R. Evaluation of evidence value of car primer using pyrolysis-gas chromatography-mass spectrometry (Py-GC-MS) and chemometrics. *Sci Lett.* 2021; 15(1): 43–57. <https://doi.org/10.24191/sl.v15i1.11796>.
71. Nakai I, Ono Y, Li Q, Nishiwaki Y, Tanaka K, Nakayama S. Capability of a TES microcalorimeter SEM-EDS system for forensic analysis of automotive paint and gunshot residue. *Surf Interface Anal.* 2010; 42(5, SI): 402–10. <https://doi.org/10.1002/sia.3113>.
72. Corrêa RS, Melo VF, Abreu GGF, Sousa MH, Chaker JA, Gomes JA. Soil forensics: How far can soil clay analysis distinguish between soil vestiges? *Sci Justice.* 2018; 58(2): 138–44. <https://doi.org/10.1016/j.scijus.2017.09.003>.
73. Fasasi A, Mirjankar N, Stoian R-I, White CG, Allen MD, Sandercock MP, et al. Pattern Recognition-Assisted Infrared Library Searching of Automotive Clear Coats. *Appl Spectrosc.* 2015 Jan; 69(1): 84–94. <https://doi.org/10.1366/14-07578>.

74. Payne G, Wallace C, Reedy B, Lennard C, Schuler R, Exline D, et al. Visible and near-infrared chemical imaging methods for the analysis of selected forensic samples. *Talanta*. 2005; 67(2): 334–44. <https://doi.org/10.1016/j.talanta.2005.03.042>.
75. Trzcinska BM, Zieba-Palus J, Koscielniak P. Examination of Car Paint Samples Using Visible Microspectrometry for Forensic Purposes. *Anal Lett*. 2013; 46(8): 1267–77. <https://doi.org/10.1080/00032719.2012.760099>.
76. Zieba-Palus J, Trzcinska BM, Koscielniak P. Comparative analysis of car paint traces in terms of color by VIS microspectrometry for forensic needs. *Anal Lett*. 2010; 43(3): 436–45. <https://doi.org/10.1080/00032710903402358>.
77. Massonnet G, Stoecklein W. Identification of organic pigments in coatings: applications to red automotive topcoats: Part III: Raman spectroscopy (NIR FT-Raman). *Sci Justice*. 1999; 39(3): 181–7. [https://doi.org/10.1016/S1355-0306\(99\)72045-9](https://doi.org/10.1016/S1355-0306(99)72045-9).
78. Smalldon KW, Moffat AC. The Calculation of Discriminating Power for a Series of Correlated Attributes. *J Forensic Sci Soc*. 1973 Oct; 13(4): 291–5. [https://doi.org/10.1016/S0015-7368\(73\)70828-8](https://doi.org/10.1016/S0015-7368(73)70828-8).
79. Grant B, Sauzier G, Lewis SW. Discrimination of automotive window tint using ATR-FTIR spectroscopy and chemometrics. *Forensic Sci Int*. 2020; 313: 110338. <https://doi.org/10.1016/j.forsciint.2020.110338>.
80. Thoonen G, Nys B, Vander Haeghen Y, De Roy G, Scheunders P. Automatic forensic analysis of automotive paints using optical microscopy. *Forensic Sci Int*. 2016; 259: 210–20. <https://doi.org/10.1016/j.forsciint.2015.12.040>.
81. Huang L, Beauchemin D. Forensic analysis of automotive paint chips for the identification of the vehicle manufacturer, colour and year of production using electrothermal vaporization coupled to inductively coupled plasma optical emission spectrometry. *J Anal at Spectrom*. 2017; 32(8): 1601–7. <https://doi.org/10.1039/C7JA00196G>.
82. He J, Lv J, Ji Y, Feng J, Liu Y. Multiple Characterizations of Automotive Coatings in Forensic Analysis. *Spectrosc Lett*. 2013 Nov; 46(8): 555–60. <https://doi.org/10.1080/00387010.2013.764901>.
83. Lavine BK, Almirall J, Muehlethaler C, Neumann C, Workman Jr J. Criteria for comparing infrared spectra: A review of the forensic and analytical chemistry literature. *Forensic Chem*. 2020; 18: 100224.
84. Sauzier G, van Bronswijk W, Lewis SW. Chemometrics in forensic science: approaches and applications. *Analyst*. 2021.
85. Wei C, Wang J. A rapid and nondestructive approach for forensic identification of car bumper splinters using attenuated total reflectance Fourier transform infrared spectroscopy and chemometrics. *J Forensic Sci*. 2021; 66(2): 583–93.

86. Qiu W, Li W. Non-destructive characterization and discrimination of vehicle bumpers fragments in forensic science using molecular spectral fusion analysis and chemometrics. *Microchem J.* 2021; 106299.
87. Martinis BS de, Oliveira MF de. *Química Forense Experimental*. São Paulo: Cengage Learning; 2015.
88. Bruni AT, Velho JA, Oliveira MF de. *Fundamentos de Química Forense: Uma Análise Prática da Química que Soluciona Crimes*. Campinas: Millennium Editora; 2012.
89. Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl Spectrosc.* 1989;43(5):772–7.
90. Brereton RG. *Chemometrics*. John Wiley & Sons, Ltd. Chichester, UK: John Wiley & Sons, Ltd; 2003.
91. Ferreira M, Antunes AM, Melgo MS, Volpe PLO. Quimiometria I: calibração multivariada, um tutorial. *Química Nov.* 1999; 22(5): 724–31.
92. Miller JN, Miller JC. *Estadística y quimiometría para química analítica*. 2002.
93. Borin A, Poppi RJ. Multivariate quality control of lubricating oils using fourier transform infrared spectroscopy. *J Braz Chem Soc.* 2004;15(4):570–6.
94. Braga JWB, Junior AADS, Martins IS. Determination of viscosity index in lubricant oils by infrared spectroscopy and PLSR. *Fuel.* 2014;120:171–8.
95. Soares LF, Silva DC da, Bergo MCJ, Coradin VTR, Braga JWB, Pastore T. Avaliação de espectrômetro NIR portátil e PLS-DA para a discriminação de seis espécies similares de madeiras amazônicas. *Quim Nov.* 2017; 40(4): 418–26. <https://doi.org/10.21577/0100-4042.20170014>.
96. Custódio MF, Magalhães LO, Arantes LC, Braga JWB. Identification of synthetic drugs on seized blotter papers using ATR-FTIR and PLS-DA: Routine application in a forensic laboratory. *J Braz Chem Soc.* 2021; 32(3): 513–22. <https://doi.org/10.21577/0103-5053.20200205>.
97. Martins AR, Talhavini M, Vieira ML, Zacca JJ, Braga JWB. Discrimination of whisky brands and counterfeit identification by UV–Vis spectroscopy and multivariate data analysis. *Food Chem.* 2017; 229: 142–51. <https://doi.org/10.1016/j.foodchem.2017.02.024>.
98. Martins AR. Desenvolvimento de método para discriminação entre marcas de uísque e identificação de adulteração por espectroscopia UV-Vis e PLS-DA. 2015.
99. Kunze DCG da C. Correção do efeito da umidade em espectros NIR empregando EPO para o desenvolvimento de modelos PLS-DA para discriminação de madeiras. 2020.
100. Muehlethaler C, Massonnet G, Hicks T. Evaluation of infrared spectra analyses using a likelihood ratio approach: A practical example of spray paint examination. *Sci Justice.* 2016 Mar; 56(2): 61–72. <https://doi.org/10.1016/j.scijus.2015.12.001>.
101. Jost C, Muehlethaler C, Massonnet G. Forensic aspects of the weathering and ageing of

spray paints. *Forensic Sci Int.* 2016 Jan; 258: 32–40.
<https://doi.org/10.1016/j.forsciint.2015.11.001>.

102. ASTM Standard E1610-18. Standard Guide for Forensic Paint Analysis and Comparison. *Astm Int.* 2018; i: 1–13.
103. ASTM E1655-05. Standard Practices for Infrared Multivariate Quantitative Analysis. *Astm Int.* 2012; 05(Reapproved 2012):29.
104. Gautam R, Chauhan R, Kumar R, Sharma V. PLS-DA and infrared spectroscopy based rapid and non-destructive discrimination of black ball and gel pen inks for forensic application. *Forensic Sci Int Reports.* 2021; 3:100162.
<https://doi.org/10.1016/j.fsir.2020.100162>.
105. Pereira LSA, Lisboa FLC, Coelho Neto J, Valladão FN, Sena MM. Screening method for rapid classification of psychoactive substances in illicit tablets using mid infrared spectroscopy and PLS-DA. *Forensic Sci Int.* 2018;288:227–35.
<https://doi.org/10.1016/j.forsciint.2018.05.001>.

ANEXO – FROTA CIRCULANTE NO DISTRITO FEDERAL



GOVERNO DO DISTRITO FEDERAL
SECRETARIA DE ESTADO DE SEGURANÇA PÚBLICA
DEPARTAMENTO DE TRÂNSITO DO DISTRITO FEDERAL
GABINETE DA DIREÇÃO-GERAL
GERÊNCIA DE ESTATÍSTICA DE ACIDENTE DE TRÂNSITO



Número da Frota de veículos por ano de fabricação, de 2008 a 13/01/2019 Distrito Federal

	Número de veículos	Porcentagem do total da frota
Veículos fabricados até 2008	768.836	43,3169%
Veículos fabricados em 2009	95.978	5,4075%
Veículos fabricados em 2010	113.305	6,3837%
Veículos fabricados em 2011	118.639	6,6842%
Veículos fabricados em 2012	118.413	6,6715%
Veículos fabricados em 2013	128.262	7,2264%
Veículos fabricados em 2014	115.670	6,5169%
Veículos fabricados em 2015	90.957	5,1246%
Veículos fabricados em 2016	70.172	3,9536%
Veículos fabricados em 2017	79.131	4,4583%
Veículos fabricados em 2018	75.543	4,2562%
Veículos fabricados em 2019	5	0,0003%
Total	1.774.911	100,0000%

Data de referência: 13/01/2019

Fonte: GDF/SSP/DETRAN/DG/DIRTEC