
Uso de Espectroscopia de Emissão Ótica com Plasma Induzido por Laser (LIBS) Associada à Aprendizagem de Máquinas para Análise de Resíduos de Disparos de Armas de Fogo(GSR)

Gabriel da Silva Souza



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS GRADUAÇÃO EM FÍSICA**

Gabriel da Silva Souza

**Uso de Espectroscopia de Emissão Ótica com
Plasma Induzido por Laser (LIBS) Associada à
Aprendizagem de Máquinas para Análise de
Resíduos de Disparos de Armas de Fogo(GSR)**

Dissertação de Mestrado apresentada ao
Programa de Pós Graduação do Instituto de
Física da Universidade Federal de Uberlândia
como parte dos requisitos para a obtenção do
título de Mestre em Física.

Área de concentração: Física

Orientador: Prof. Dr. Jader de Souza Cabral

Uberlândia
2025

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

S729u Souza, Gabriel da Silva, 2000-
2025 Uso de Espectroscopia de Emissão Ótica com Plasma Induzido por
Laser (LIBS) associada à Aprendizagem de Máquinas para Análise de
Resíduos de Disparos de Armas de Fogo (GSR) [recurso eletrônico] /
Gabriel da Silva Souza. - 2025.

Orientador: Jader de Souza Cabral.
Dissertação (Mestrado) - Universidade Federal de Uberlândia,
Programa de Pós-graduação em Física.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.di.2025.5090>
Inclui bibliografia.
Inclui ilustrações.

1. Física. I. Cabral, Jader de Souza, 1985-, (Orient.). II. Universidade
Federal de Uberlândia. Programa de Pós-graduação em Física. III. Título.

CDU: 53

André Carlos Francisco
Bibliotecário-Documentalista - CRB-6/3408



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Física
Av. João Naves de Ávila, 2121, Bloco 1A, Sala 213 - Bairro Santa Mônica, Uberlândia-MG,
CEP 38400-902
Telefone: (34) 3239-4309 - www.infis.ufu.br - cpgfisica@ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Física				
Defesa de:	Dissertação de Mestrado				
Data:	25 de fevereiro de 2025	Hora de início:	09:00	Hora de encerramento:	11:01
Matrícula do Discente:	12312FIS002				
Nome do Discente:	Gabriel da Silva Souza				
Título do Trabalho:	Uso de Espectroscopia de Emissão Ótica com Plasma Induzido por Laser (LIBS) associada a Aprendizagem de Máquinas para análise de resíduos de disparo de armas de fogo (GSR)				
Área de concentração:	Física				
Linha de pesquisa:	Espectroscopia Óptica associada a Machine learning para reconhecimento de padrões				
Projeto de Pesquisa de vinculação:	N.A				

Reuniu-se, por meio de Videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Física -UFU, assim composta: Professores Doutores: José de los Santos Guerra - INFIS/UFU, Bruno Spolon Marangoni - UFMS e Jader de Souza Cabral - INFIS/UFU, orientador do discente.

Iniciando os trabalhos o presidente da mesa, Prof. Jader de Souza Cabral, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao discente a palavra para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(as) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Bruno Spolon Marangoni, Usuário Externo**, em 25/02/2025, às 11:56, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Jader de Souza Cabral, Professor(a) do Magistério Superior**, em 25/02/2025, às 12:05, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **José de los Santos Guerra, Professor(a) do Magistério Superior**, em 25/02/2025, às 12:06, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6127933** e o código CRC **FB2B4181**.

Referência: Processo nº 23117.012058/2025-51

SEI nº 6127933

Agradecimentos

Gostaria de agradecer aos meus pais Manoel e Maura, por me incentivar a crescer como ser humano, independente dos obstáculos da vida, e a Deus me guiando em meio a sua infinita sabedoria.

Ao meu orientador Dr. Jader de Souza Cabral, foi um prazer aprender e testemunhar o que significa ser um pesquisador de verdade.

Também gostaria de agradecer ao grupo de pesquisa em ciências de dados Boitatá, mais precisamente ao seu diretor Mardônio Jó Carvalho de França pelos ensinamentos na área.

E obrigado a CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior e a PROCAD - Programa de Cooperação Acadêmica em Defesa Nacional pelo financiamento.

A todos os professores do programa de pós graduação em Física da UFU que contribuíram para o meu aprendizado, muito obrigado.

*A busca pela verdade consistirá no exercício da justiça - Gabriel S.Souza
Bem-aventurados os que têm fome e sede de justiça, pois serão satisfeitos. Mateus 5:6*

Resumo

Em uma situação na qual envolva o acidente ou agressão com uma arma de fogo, é possível identificar o responsável pelos disparos através de marcadores químicos. Isso é feito identificando elementos já catalogados pertencentes a munições convencionais, como por exemplo o Ba - bário, Pb - chumbo e Sb - antimônio. No entanto, munições como a NTA (*Non-toxic ammunition*) não possuem marcadores químicos tão claros como as munições convencionais necessitando de métodos alternativos na identificação de um possível atirador. Logo, este trabalho propõe um protocolo experimental capaz de fornecer tais marcadores químicos, e ainda fornece uma margem de precisão na identificação evitando falsos positivos. Para isso, é utilizado a espectroscopia óptica de emissão com plasma induzido por laser (LIBS) fornecendo a assinatura elementar das amostras, que por sua vez, serve de informação para treino do modelo de classificação *Random Forest*(RF) que é baseado em métricas estatísticas de informação. Com este método de aprendizagem de máquina supervisionado, é possível analisar quais parâmetros são utilizados ao distinguir um atirador de um não atirador tornando-o explicável. As amostras foram coletadas em um estande de tiro, sob protocolos estabelecidos pela SENASP(Secretaria Nacional de Segurança Pública), consistindo de 2, 6 e 8 disparos com munição NTA, e 1 e 6 disparos com munições convencionais. E para comparação foram coletadas amostras de 0 disparo (amostras da mão de um indivíduo que não entrou em contato com a arma de fogo). Os resultados apresentam marcadores químicos identificados a partir de transições atômicas de Cu I, Cu II, Ti I e Mo I, o modelo de classificação foi capaz de atingir uma precisão acima de 90% e possui um poder de generalização alto o suficiente para distinguir um atirador de um não atirador independentemente do tipo de munição, convencional ou NTA.

Palavras-chave: Munição NTA. LIBS. Random Forest.

Abstract

In a situation involving an accident or assault with a firearm, it is possible to identify the person responsible for the shots using chemical markers. This is done by identifying elements already catalogued in conventional ammunition, such as Ba - barium, Pb - lead and Sb - antimony. However, ammunition such as NTA (non-toxic ammunition) does not have chemical markers as clear as conventional ammunition, requiring alternative methods to identify a possible shooter. Therefore, this work proposes an experimental protocol capable of providing such chemical markers, while still providing a margin of accuracy in identification and avoiding false positives. To this end, laser-induced breakdown spectroscopy (LIBS) is used to provide the elemental signature of the samples, which in turn serves as information for training the *Random Forest*(RF) classification model, which is based on statistical information metrics. With this supervised machine learning method, it is possible to analyze which parameters are used to distinguish a shooter from a non-shooter, making it explainable. The samples were collected at a shooting range, under protocols established by SENASP (National Secretariat for Public Security), consisting of 2, 6 and 8 shots with NTA ammunition, and 1 and 6 shots with conventional ammunition. For comparison, samples were taken from 0 shots (samples from the hand of an individual who did not come into contact with the firearm). The results show chemical markers identified from Cu I, Cu II, Ti I and Mo I atomic transitions, the classification model was able to achieve an accuracy of over 90% and has a generalization power high enough to distinguish a shooter from a non-shooter regardless of the type of ammunition, conventional or NTA.

Keywords: NTA ammunition. LIBS. Random Forest.

Listas de ilustrações

Figura 1 – Ilustração dos mecanismos Físicos do método LIBS. A) o feixe laser é absorvido pela matéria desencadeando o processo de ablação. B) O aumento de temperatura e expansão abrupta da pluma gasosa comprime o gás atmosférico produzindo ondas de choque, onde ao mesmo tempo é ionizada. C) A medida que o plasma arrefece, linhas de emissão atômica podem ser observadas junto a emissões contínuas devido o efeito <i>Bremsstrahlung</i> . Fonte: Elaborada pelo autor.	26
Figura 2 – O gráfico exemplifica o efeito de blindagem para um plasma com temperatura na ordem de 6500 K e uma densidade de elétrons na ordem de $1,09 \times 10^{17} \text{ cm}^{-3}$. Fonte: Elaborada pelo autor.	28
Figura 3 – Em meio ao fluxo de partículas no plasma, os elétrons podem ser capturados por íons emitindo fôtons durante a recombinação, também conhecido como emissão livre-livre. Fonte: Elaborada pelo autor.	29
Figura 4 – Arvore de decisão criada a partir de dados de espectroscopia LIBS no <i>Random Forest</i> . Fonte: Elaborada pelo autor.	30
Figura 5 – Representação do método <i>Cross Validation</i> na divisão da base de dados em um conjunto de validação e treino. Imagine que todo o conjunto de dados seja a barra em azul, a partir dele, dividiremos em um grupo de amostras usadas para treino do modelo de classificação (barra laranja) e outro grupo para testá-lo (barra em verde). O <i>Cross Validation</i> utiliza toda a base de dados como treino, selecionado um percentual para validação de forma que nenhuma amostra seja repetida. Fonte: Elaborada pelo autor.	33
Figura 6 – Cada ponto do mapa de score é um espectro representado em 2d, construído a partir dos componentes principais mais representativos da matriz de covariância. Fonte: Elaborada pelo autor.	35
Figura 7 – A Figura a) apresenta um deslocamento da base devido ao ruído branco, onde a Figura b) apresenta sua correção. Fonte: Elaborada pelo autor.	37

Figura 8 – Regiões da mão onde são coletadas os resquícios de arma de fogo. Fonte: Elaborada pelo autor.	39
Figura 9 – Fitas dupla-face contendo as amostras de resquícios de disparo. Fonte: Elaborada pelo autor.	40
Figura 10 – Diagrama esquemático do sistema experimental para medidas LIBS. Fonte: Elaborada pelo autor.	41
Figura 11 – Representação das regiões de disparo com os pulsos laser na fita dupla face. Fonte: Elaborado pelo autor.	41
Figura 12 – Espectro médio de 8 e 0 disparos apresentado no intervalo de 205 - 225 nm. Fonte: Elaborado pelo autor.	43
Figura 13 – Espectro médio de 2, 8 disparos NTA e 0 disparos, no intervalo de 205 - 225 nm. Mesmo com apenas dois disparo, é possível coletar amostras em quantidades o suficiente para ter uma resolução clara das linhas de transição. Fonte: Elaborado pelo autor.	44
Figura 14 – Espectro médio de 0, 2 e 8 disparos no intervalo de 260 - 274 nm destacando marcadores químicos Cu II e Ti I. Fonte: Elaborada pelo autor	45
Figura 15 – Espectro médio de 8 e 0 disparos no intervalo de 470 - 530 nm destan- cando a presença de Compostos químicos de aço inox. Fonte: Elabo- rada pelo autor	45
Figura 16 – Matriz de confusão das classes, 0 - não atirador, 1 - atirador. Fonte: Elaborada pelo autor	47
Figura 17 – O modelo de classificação estabelece uma medida estatística(%) de re- levância para cada uma das características dos dados, sendo essas des- critas em barras no gráfico. Fonte: Elaborada pelo autor.	48
Figura 18 – As 5 características mais relevantes na discriminação das classes (0 e 1) são os elementos químicos: Ti I(260,076 nm), Fe II(288,262 nm, 288,206 nm), Fe I(274,374 nm), Fe I(302,124 nm) e Fe II(234,421 nm). Fonte: Elaborada pelo autor	48
Figura 19 – Cada ponto é um espectro representado em 2d, onde em vermelho está a classe - 0 e em verde a classe - 1. Comparando esta figura com a matriz de confusão (Figura 17), o falso - positivo e o falso - negativo se destacam como um ponto verde em meio ao grupo da esquerda e um ponto vermelho em meio ao grupo da direita. Fonte: Elaborada pelo autor.	49
Figura 20 – Matriz de confusão do modelo <i>Random Forest</i> treinado a partir de 138 espectros de 0 e 2 disparos. Onde, as classes são definidas aqui como, 0 - nenhum disparo e 1 - apenas dois disparos. Fonte: Elaborada pelo autor.	50

Figura 21 – A esquerda em vermelho temos os espectros classificados como 0 disparo, e a direita em azul os espectros classificados como 2 disparos. Fonte: Elaborada pelo autor.	51
Figura 22 – Espectro médio das amostras de 8 disparos - NTA e 6 disparos - convencional, em uma das regiões onde se encontram os marcadores químicos da munição convencional, Sb I (antimônio) e Pb I (chumbo). Fonte: Elaborada pelo autor.	52
Figura 23 – Espectro médio das amostras de 8 disparos - NTA e 6 disparos - convencional, em uma das regiões onde se encontram os marcadores químicos de NTA. Fonte: Elaborado pelo autor.	52
Figura 24 – A esquerda em azul, amostras classificadas como munição convencional, enquanto que a direita em amarelo, espectros classificados como NTA. Fonte: Elaborada pelo autor.	53
Figura 25 – PCA dos espectros classificados. Ao fundo em vermelho se encontra o grupo de amostras 0 disparo, e em azul e verde o grupo de atirados que se destacam em duas regiões ao longo da componente principal 1. Fonte: Elaborada pelo autor.	54

Lista de tabelas

Tabela 1 – Amostras coletadas	40
Tabela 2 – Número total de espectros restantes com a remoção de <i>outliers</i>	42
Tabela 3 – Linhas de emissão atômica medidas	46

Sumário

1	INTRODUÇÃO	21
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Física da Espectroscopia LIBS	25
2.2	Modelo de Classificação Random Forest	29
2.2.1	Métricas de validação do modelo de classificação	32
2.3	Análise de Componentes Principais (PCA)	34
2.4	Correção de Linhas de Base	36
2.5	Normalização dos Espectros	37
2.6	Eliminação de <i>Outliers</i>	37
3	METODOLOGIA	39
3.1	Produção e Coleta de Resíduos de Disparo de Armas de Fogo GSR	39
3.2	Medidas experimentais LIBS	40
3.3	Pré-Tratamento e Análise dos Espectros	42
4	ANÁLISE DISCURSAO DOS RESULTADOS	43
4.1	Identificação dos Marcadores Químicos	43
4.2	Resultados Random Forest	46
4.2.1	<i>Random Forest</i> aplicado a espectros de 0 e 2 disparos NTA	50
4.3	Distinção Entre Munições Convencionais e NTA	51
5	CONCLUSÃO	55
	REFERÊNCIAS	57
	APÊNDICES	61

CAPÍTULO 1

Introdução

A ciência forense se ampara nas múltiplas ferramentas fornecidas por métodos experimentais e analíticos da física, química e biologia, com o intuito de fornecer ao poder judiciário dados concretos, e de forma imparcial, dos eventos ocorridos em um delito, acidente, ou de qualquer situação na qual envolva uma vítima. Na área de balística, tanto na vítima quanto no agressor, são identificadas provas que levam à caracterização do ocorrido, mas neste trabalho, o foco consistirá na identificação de um agressor com arma de fogo a partir da técnica de espectroscopia óptica.

Uma arma de fogo consiste em um equipamento capaz de lançar projéteis em alta velocidade devido à expansão de gases, provocada pela liberação intensa de energia térmica por um propelente. A pólvora negra é o agente detonador mais antigo já registrado, datado do século IX, consistindo de uma mistura de carbono (C), nitrato de potássio (KNO_3) e enxofre (S). A primeira arma de fogo se assemelhava a um mini canhão de mão, onde era carregado com pólvora, chumaço e em seguida as munições, presas ao fundo do objeto no qual possuía um pequeno furo usado para dar ignição à pólvora com fogo (HEARD, 2011).

Em equipamentos mais modernos, buscou-se aprimorar a precisão da mira e diminuir a quantidade de fumaça liberada na detonação da pólvora modificando as munições, no entanto, tal fumaça ainda existe e é de grande relevância na aquisição de provas do disparo. A munição convencional é constituída de um núcleo do projétil que é inserida em um estojo, que por sua vez é preenchido pelo propelente. Na base da munição é acoplado a espoleta, responsável por dar ignição ao propelente quando atingida pelo percussor (peça da arma de fogo). De tais modificações, são fabricadas munições com diversas propriedades, cada uma delas adequadas a contextos específicos de uso. Toda vez que uma arma de fogo é disparada, os gases provenientes da ignição do propelente, da espoleta juntamente com fragmentos da bala e da própria arma são ejetados junto com o projétil (SEROL et al., 2023). Tais gases são definidos como resíduos do disparo GSR(*Gunshot Residues*), que por sua vez se depositarão na mão do atirador, em seu traje e, às vezes na vítima. Por essas características, os protocolos de investigação aplicados

na identificação de um atirador, partem da coleta de amostras nas mãos (MINZIÈRE et al., 2023).

A partir disso, em munições convencionais de revólveres e pistolas, é encontrado um padrão químico dos resíduos correspondendo a bário (Ba), chumbo (Pb) e antimônio (Sb), também definidos como marcadores químicos. Em se tratando de metais tóxicos, a presença dos três marcadores identifica o atirador. O componente responsável por inflamar o propelente é composto por substâncias inorgânicas de estifnato de chumbo ($C_6HN_3O_8Pb$), nitrato de bário ($Ba(NO_3)_2$) e trissulfeto de antimônio (Sb_2S_3) formando aglomerados de partículas após entrar no estado de fusão com os disparos (DALBY; BUTLER; BIRKETT, 2010).

As técnicas experimentais que englobam a investigação de espaços e objetos que estiveram em contato com GSR, relevam não só a estrutura morfológica dos resíduos mas também a sua identificação elementar (SHRIVASTAVA; JAIN; NAGPAL, 2021). A técnica de microscopia eletrônica de varredura (MEV), por exemplo, destaca a característica da microestrutura dos resquícios de disparo (ROMOLO; MARGOT, 2001), e pesquisas com a técnica de RBS(*Rutherford backscattering*) além da identificação elementar, são capazes de detectar até mesmo efeitos de contaminação cruzada, identificando chumbo em munições que não apresentam tal elemento químico (DUARTE et al., 2018). Uma outra alternativa experimental é usar a técnica de Espectroscopia de Emissão Óptica com Plasma Induzido por Laser (LIBS - *Laser - Induced Breakdown Spectroscopy*). Entre as vantagens de usar a espectroscopia LIBS podemos citar: (i) sua característica de não degradar completamente a amostra, permitindo que outras técnicas experimentais ainda sejam utilizadas mutuamente; (ii) possibilidade de adquirir informação o suficiente dos GSRs para análise multivariada e (iii) possibilidade de compor sistemas portáteis (LÓPEZ-LÓPEZ et al., 2017; DOÑA-FERNÁNDEZ et al., 2018; FAMBRO et al., 2017; CIOCCIA et al., 2024).

Todas essas técnicas definem de forma clara a presença de Ba, Pb e Sb, no entanto, no mercado armamentista há munições que não apresentam tais marcadores químicos. As munições definidas como ecológicas são as NTA (Non-toxic ammunition), onde a espoleta não contém metais pesados, assim, durante os disparos nem mesmo o atirador e a arma são contaminados. A justificativa de fabricar tais munições NTA se baseia na relação da presença de saturnismo através da contaminação por Chumbo - Pb encontrado em munições convencionais, como bem apresentado na ficha técnica da Companhia Brasileira de Cartuchos (CBC, 2022). Esse novo tipo de munição acarreta em um problema para a ciência forense, pelo fato de sua composição química poder ser encontrada em qualquer âmbito social, inclusive naqueles que não envolvam o porte de armas de fogo. Segundo o Anuário Brasileiro de Segurança Pública, estima-se que até 2025 cerca de 4,8 milhões de armas de fogo deverão ser gerenciadas e fiscalizadas pela Polícia Federal (PÚBLICA, 2024). Logo, independentemente dos incidentes que tais porte de arma possa vir acarretar

à segurança pública, se faz necessário técnicas experimentais capazes de investigar um maior número de cenários possíveis, garantindo imparcialidade e um alto grau de precisão ao identificar atiradores.

Sendo assim, este trabalho de mestrado tem como objetivo principal propor um protocolo experimental capaz de identificar e distinguir um atirador de arma de fogo portador de munições NTA de um não atirador a partir da técnica de espectroscopia LIBS. Será avaliada a capacidade da técnica, a partir da comparação entre espectros, para identificar atiradores e marcadores químicos em amostras NTAs. Além disso, associaremos à técnica LIBS o modelo de aprendizagem de máquina supervisionado *Random Forest* capaz de mapear padrões, servindo assim, como método de classificação. Desta forma, será possível obtermos uma margem de precisão ao identificar atiradores e não atiradores de amostras desconhecidas.

Como objetivo secundário, é proposto (i) destacar a distinção entre portadores de NTA e um não atirador através da análise de componentes principais (PCA); (ii) apresentar a diferença entre espectros LIBS de munições convencionais e NTA; (iii) testar o poder de generalização que o modelo de classificação é capaz de apresentar, analisando a precisão ao identificar não atirador de atiradores portadores tanto de munições convencionais quanto de NTA.

A fim de justificar os métodos utilizados neste trabalho capazes de alcançar os objetivos aqui proposto, o próximo capítulo abordará em detalhes a física necessária para que a técnica LIBS seja possível, e as ferramentas matemáticas úteis para manipulação dos espectros registrados.

CAPÍTULO 2

Fundamentação Teórica

2.1 Física da Espectroscopia LIBS

Na natureza, estados da matéria são alcançados durante a sua interação com alguma fonte de energia, alterando a dinâmica de seus constituintes, gerando propriedades distintas. Por exemplo, quando um gás em temperaturas na ordem de dezenas de milhares de kelvins é ionizado, o fluxo de íons e elétrons geram campos elétricos e magnéticos definindo tal estado da matéria como plasma (GOLDSTON, 2020). Plasmas possuem emissões atômicas devido às suas características, sendo assim, bastante úteis na identificação e caracterização de amostras sólidas, como na espectroscopia.

A técnica experimental *Laser-Induced Breakdown Spectroscopy* do acrônimo LIBS, busca, através do controle de produção e análise de plasmas induzidos por laser, obter informações da composição elementar da amostra sem causar deterioração completa da mesma. Com lasers pulsados e de alta fluência (GW/cm^2), é possível gerar ablação na região focal onde fica localizado o objeto de estudo. Por consequência das configurações de irradiância e tempo do pulso laser (ns), um vapor é formado e, em seguida, ocorre ignição do plasma onde nascem as linhas de emissão atômica (FORTES et al., 2013; NICOLODELLI et al., 2024; RIBEIRO, C et al., 2024).

A ablação da matéria, consiste na remoção de uma pequena porção da amostra (da ordem de 10^{-12} g) na forma de fusão, sublimação, ionização, erosão ou explosão do objeto. Na LIBS, os processos físicos desde a ablação, consistem na geração do plasma seguida por sua expansão e resfriamento e emissão óptica; tais fenômenos são regidos pelo tempo de exposição da matéria ao pulso laser e pela irradiância incidente (Figura 1). Sendo assim, as seções seguintes abordarão os aspectos físicos em um sistema que evolui numa escala temporal de nanosegundos (tempo de duração do pulso laser), até microsssegundos (tempo que acontece a detecção das linhas espetrais) (SINGH; THAKUR, 2020).

A matéria sólida quando exposta a uma irradiância da ordem de $10^2 - 10^{11}$ W/cm^2 , está sujeita a ablação por meio da sublimação que em seguida é ionizada. Sendo mecanismos térmicos de ablação preponderantes para irradiância $< 10^8 \text{ W}/\text{cm}^2$, isso significa

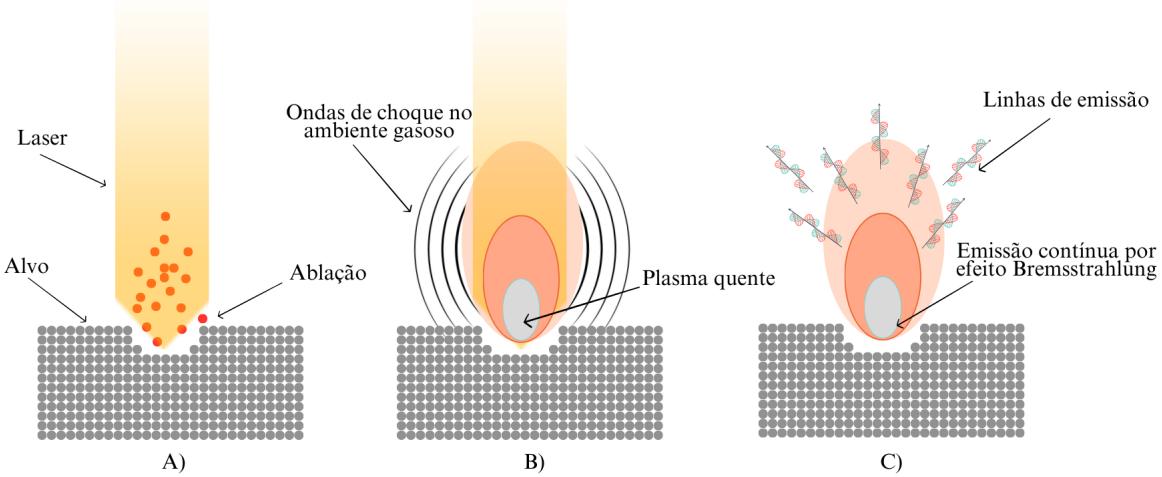


Figura 1 – Ilustração dos mecanismos Físicos do método LIBS. A) o feixe laser é absorvido pela matéria desencadeando o processo de ablação. B) O aumento de temperatura e expansão abrupta da pluma gasosa comprime o gás atmosférico produzido ondas de choque, onde ao mesmo tempo é ionizada. C) A medida que o plasma arrefece, linhas de emissão atômica podem ser observadas junto a emissões contínuas devido o efeito *Bremsstrahlung*. Fonte: Elaborada pelo autor.

que nessa ordem de grandeza é possível observar transições bem definidas de fase, como fusão ou vaporização (TAYLOR; FRANCIS, 1989). Durante o pulso laser da ordem de nanosegundos, os elétrons da camada de valência da matéria vaporizada são acelerados e ejetados formando uma pluma ionizada num efeito conhecido como *Bremsstrahlung* inverso (HEALD; MARION, 2012). Com a formação do plasma, o fluxo de elétrons e íons pode barrar parte do feixe laser caso ainda tenha incidência. Tal efeito é conhecido como bloqueio (*shielding*) do plasma (RYBICKI; LIGHTMAN, 1991). Partindo do fato de que um potencial elétrico ϕ (Eq. 2.1) se propaga em um meio que se encontra em equilíbrio térmico, e a densidade de carga ρ dependerá da densidade relativa de íons n_i e elétrons n_e .

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0} \quad (2.1)$$

$$\rho = e(n_i - n_e) \quad (2.2)$$

onde ϵ_0 é a constante de permissividade elétrica no vácuo e e é a carga elementar do elétron. Devido ao equilíbrio térmico, n_i e n_e dependerão apenas do fator de Boltzmann (GOLDSTON, 2020), como mostram Eq.(2.3) e Eq. (2.4).

$$n_e = n_0 \cdot \exp\left(\frac{-e\phi}{k_B T}\right) \quad (2.3)$$

$$n_i = n_0 \cdot \exp\left(\frac{e\phi}{k_B T}\right) \quad (2.4)$$

onde k_B é a constante de Boltzmann, T a temperatura do gás de íons e elétrons e n_0 a densidade de partículas dentro do volume delimitado pelo plasma definindo uma quase neutralidade em um sistema idealizado. Substituindo as equações acima na equação de Poisson (Eq. 2.1), temos:

$$\nabla^2 \phi = \frac{e(n_0 \exp\left(\frac{e\phi}{k_B T}\right) - n_0 \exp\left(\frac{-e\phi}{k_B T}\right))}{\epsilon_0} \quad (2.5)$$

Devido à alta temperatura do plasma, temos que os efeitos da energia potencial elétrica em relação à energia térmica será $e\phi \ll K_B T$, logo fazendo a expansão em série de Taylor e ignorando termos de ordem maior que 2, temos:

$$\nabla^2 \phi = \frac{en_0}{\epsilon_0} \left[1 + \frac{e\phi}{k_B T} - \left(1 - \frac{e\phi}{k_B T} \right) \right] \quad (2.6)$$

Reorganizando a Eq. (2.6) encontramos o comprimento de Debye, λ_D Eq. (2.8). A distribuição de cargas no plasma impede a penetração de um campo no meio, onde tal blindagem atua em uma escala de comprimento λ_D .

$$\nabla^2 \phi = \frac{2}{\lambda_D^2} \phi \quad (2.7)$$

$$\lambda_D = \sqrt{\frac{\epsilon_0 k_B T}{e^2 n_0}} \quad (2.8)$$

Supondo que o potencial elétrico ϕ tenha simetria radial Eq. (2.9), podemos reescrever a Eq. (2.7) em termos das derivadas radiais, Eq (2.10), cuja solução é dada pela Eq. (2.11):

$$\phi = \frac{\Phi(r)}{r} \quad (2.9)$$

$$\frac{1}{r^2} \frac{d}{dr} (r^2 \cdot \frac{d\phi}{dr}) = \frac{2}{\lambda_D^2} \phi \quad (2.10)$$

$$\phi(r) = \frac{\phi_0}{r} \cdot \exp\left(\frac{-\sqrt{2}}{\lambda_D} r\right) \quad (2.11)$$

O comprimento de Debye, λ_D , também pode ser interpretado como o raio de uma esfera imaginária onde ocorrerá tal blindagem. A proteção do plasma faz com que o potencial elétrico decaia exponencialmente (veja a Figura 2) para uma carga blindada no interior da esfera de Debye, fazendo com que as ondas eletromagnéticas incidente sofram amortecimento reduzindo a taxa de ablação da amostra alvo (RUSSO et al., 2007).

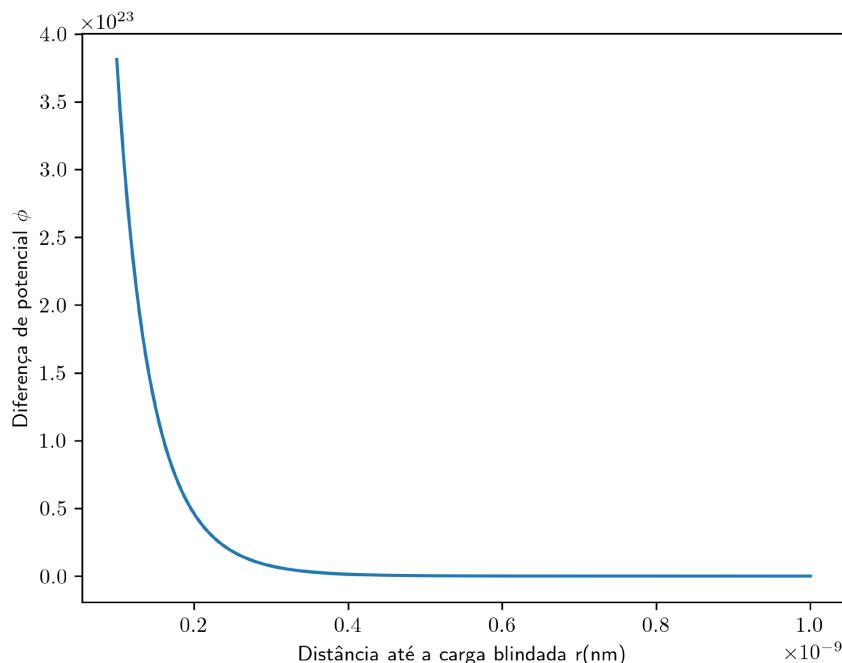


Figura 2 – O gráfico exemplifica o efeito de blindagem para um plasma com temperatura na ordem de 6500 K e uma densidade de elétrons na ordem de $1,09 \times 10^{17} \text{ cm}^{-3}$.
Fonte: Elaborada pelo autor.

Após tal ignição do plasma, no ar, a pluma ionizada em alta temperatura se expande em alta velocidade comprimindo os gases atmosféricos provocando ondas de choque e sua evolução dependerão do material ablacionado, intensidade do laser, região focal e do ambiente atmosférico.

Os mecanismos de arrefecimento do plasma estão ligados tanto à sua emissão espectral quanto a condução térmica para o ambiente atmosférico. A medida que o plasma se expande, o gás emite radiação de recombinação desacelerando elétrons com energia cinética ξ_e em íons de carga Z. Esse efeito *Bremsstrahlung* (Figura 3), resulta na emissão contínua de fôtons com energia $h\nu$ proporcional a energia de ionização χ^n do nível n Eq. (2.12) (RAI; THAKUR, 2007).

$$h\nu = \xi_e + \chi^n \quad (2.12)$$

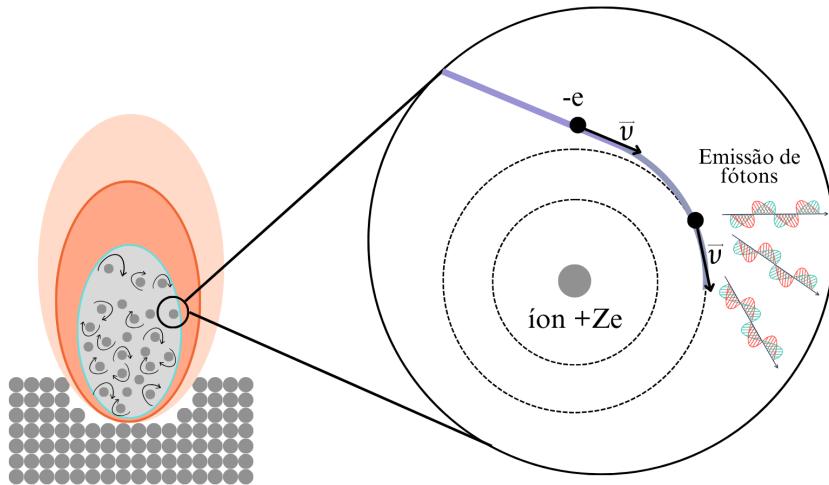


Figura 3 – Em meio ao fluxo de partículas no plasma, os elétrons podem ser capturados por íons emitindo fótons durante a recombinação, também conhecido como emissão livre-livre. Fonte: Elaborada pelo autor.

As linhas de emissão atômica necessárias na identificação elementar da amostra são sobrepostas às emissões contínuas próxima ao alvo. No entanto, como as emissões contínuas são provenientes de efeitos térmicos, elas decaem mais rapidamente a medida que a pluma de gás se arrefece, sendo as linhas de emissão por átomo registradas com mais clareza na ordem de microsegundo após o disparo do laser. Logo, a emissão de linha espectral I_{ij} [W/sr] registrada na transição entre os estados i e j , dependerá da probabilidade de emissão espontânea A_{ij} , energia do fóton emitido e densidade de átomos N_i no estado i (NOLL, 2012).

$$I_{ij} = \frac{A_{ij}}{4\pi} \frac{hc}{\lambda_{ij}} N_i \quad (2.13)$$

2.2 Modelo de Classificação Random Forest

O *Random Forest* é um algoritmo de aprendizagem de máquina supervisionado, com a finalidade de classificar ou gerar um modelo de regressão em múltiplas árvores de decisão. Usando-o como um método de classificação cada um dos dados de treino são previamente identificados com a sua respectiva classe. Durante o processo de construção das árvores de decisão o modelo basicamente mapeia padrões a partir das características de cada uma das amostras, calculando estatisticamente as variáveis mais representativas associadas à sua classe e as definindo como um ponto de decisão (BREIMAN et al., 2017).

A variável inicial de um árvore de decisão é denominada raiz, enquanto que as ramificações a partir dela são definidas como nós (filhos) (veja a Figura 4). A árvore cresce com uma distribuição hierárquica até um ponto onde não é mais possível agregar informação a uma classe ou até que todas as amostras de um subconjunto seja pertencente a apenas uma classe. Nesse estágio final, os nós de decisão são denominados de folhas.

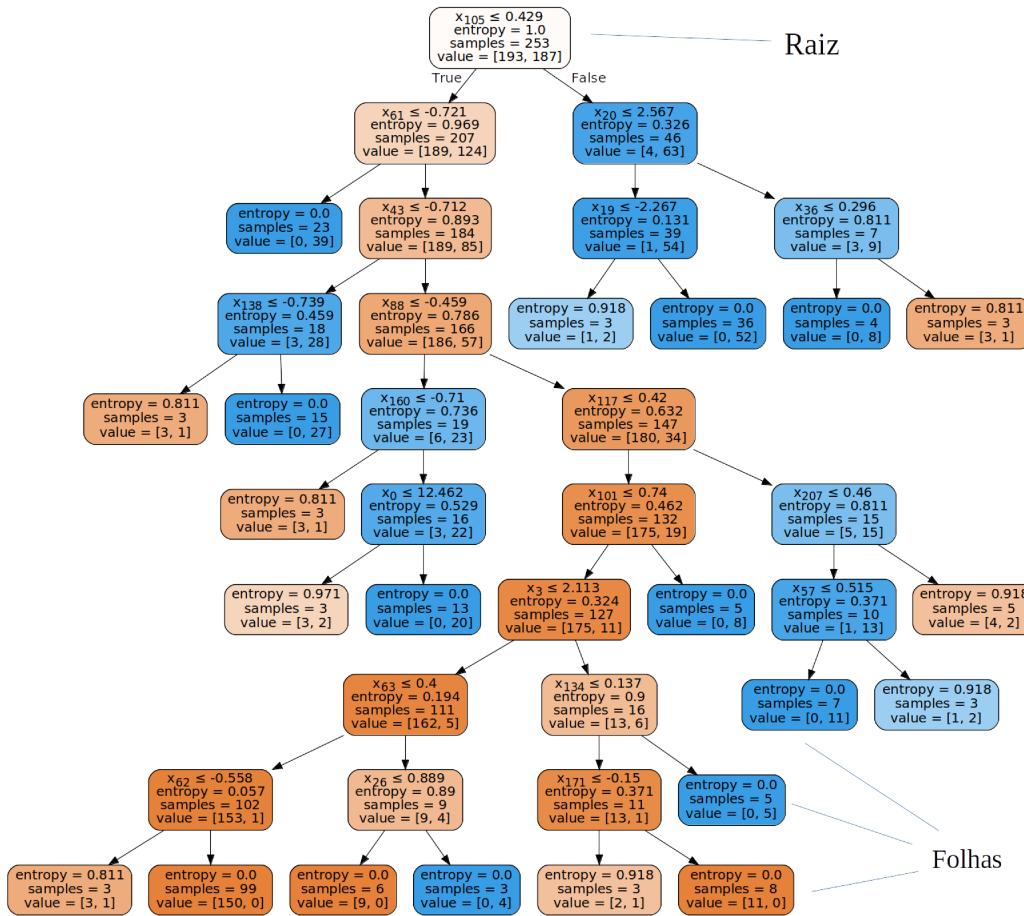


Figura 4 – Árvore de decisão criada a partir de dados de espectroscopia LIBS no *Random Forest*. Fonte: Elaborada pelo autor.

As bibliotecas da linguagem de programação *Python*, como por exemplo *Scikit learn*, utilizada neste trabalho, usam como métricas para seleção de variáveis a teoria de informação de Shanon. É possível quantificar o quanto de informação é obtida com a realização de um dado evento apenas se baseando na probabilidade de sua ocorrência. A informação de Shanon f também conhecida por surpresa é definida como (AN, 1961):

$$f = \log_2\left(\frac{1}{p(i)}\right) \quad (2.14)$$

onde $p(i)$ é a probabilidade de ocorrência do evento i. Partindo de um conjunto de possibilidades $X = x_1, x_2, \dots, x_n$ e do caráter aditivo que a expressão \log_2 possui, a informação e a incerteza pode ser quantificadas em um aspecto geral calculando sua entropia média H Eq.(2.15) (SHANNON, 1948).

$$\begin{aligned} H(X) &= p(x_1)f_1 + p(x_2)f_2 + \dots + p(x_n)f_n \\ &= -p(x_1)\log_2(p(x_1)) - p(x_2)\log_2(p(x_2)) - \dots - p(x_n)\log_2(p(x_n)) \\ H(X) &= -\sum_{i=1}^n p(x_i)\log_2(p(x_i)) \end{aligned} \quad (2.15)$$

Usando a Eq. (2.15), as probabilidades $p(x_i)$ são calculadas no *Random Forest* com base na frequência de ocorrência das n - classes, e em função deste resultado, o atributo utilizado em cada nó é definido pela expressão matemática ganho de informação G Eq. (2.16) (MICHIE et al., 1995).

$$G = H_{classe} - \sum_{i=1}^n \frac{Nf_i}{Np} H_i \quad (2.16)$$

Em que Nf_i e Np são o número de amostras pertencentes ao nó filho e ao pai, respectivamente, que H_i é a entropia do atributo selecionado. Supondo que na base de dados um dos atributos seja numérico como temperatura por exemplo, então, o algorítimo seleciona um valor A e usa o mesmo como um ponto de referência separando os demais valores do mesmo atributo em dois subconjuntos, maiores e menores que A . Em cada subconjunto é calculado sua entropia, onde no fim ambos são somados ponderando cada intervalo dividindo o número de amostras em seu respectivo subconjunto Nf_i pelo número total de amostras Np . O ganho de informação G é então calculado, subtraindo da entropia da coluna de classificação (H_{classe}) a entropia ponderada da variável temperatura.

O valor numérico selecionado com maior ganho de informação será um nó de decisão, que por sua vez, cada intervalo de amostras selecionados a partir dele será um subconjunto. De cada um desses subconjuntos outro nó filho deve ser calculado analisando outra variável pertencente a esses conjuntos de amostras, repetindo esse processo até que a entropia chegue a 0(todas as amostras de um subconjunto pertencem a mesma classe) ou a 1 (veja a Figura 4). Construindo N (estimadores) árvores de decisão, o *Random Forest* classificará uma amostra de teste em cada uma das classes, aquela com maior probabilidade de ocorrer é definida como resultado (HASTIE et al., 2009). A entropia é um dos métodos acessíveis com o *Random Forest*, podendo também serem utilizados outras métricas como a impureza *GINI*.

Na área de aprendizagem de máquina também é comumente utilizado o classificador KNN (k - Nearest Neighbors). Após classificar previamente cada amostra de treino no modelo, o algoritmo calcula uma distância euclidiana entre os espectros, mapeando os k vizinhos mais próximos. Assim, definido um valor de k, os espectros utilizados como treino são classificados a partir de uma votação, onde a classe com mais vizinhos próximos do espectro de treino, será então a classe atribuída (GOLDBERG, 2005).

2.2.1 Métricas de validação do modelo de classificação

Cada modelo treinado de *machine learning* se especializa em problemas individuais, sendo improvável generalizações de tarefas. Isso significa que um modelo utilizado para classificar imagens não será capaz de executar a mesma função ao tentar distinguir sinais sonoros, ou qualquer outra espécie de dados. Sendo assim, métricas de avaliação analisam o desempenho de um modelo em executar tarefas específicas que lhe foram atribuídas (WARDHANI et al., 2019).

Em se tratando de classificadores como o *Random Forest*, a métrica Acurácia calcula a proporção de amostras classificadas corretamente, dividindo o número de acertos pelo número total de possibilidades, Eq(2.17). Também é possível averiguar o quanto uma dada classe é beneficiada, ou se não há um enviesamento das decisões de classificação, analisando a proporção de verdadeiro positivos (TP), previsão correta de uma classificação positiva, e verdadeiro negativo (TN) que se trata de uma resposta correta de uma classificação negativa. Junto a isso, informações rotuladas previamente como positivas, mas, previstas como negativas são denominadas de falsos negativos (FN), enquanto que falsos positivos (FP), são dados rotulados como negativas, no entanto, previstas como positiva (RAINIO; TEUHO; KLÉN, 2024).

Os resultados de previsões binárias, podem ser representados em uma tabela 2x2 definida como matriz de confusão, como exemplo, rotulando de positivo para atirador e negativo para não atirador. De tais informações é possível calcular a Sensibilidade Eq.(2.18), a qual destaca o grau de assertividade do modelo em classes positivas. Da mesma forma, a qualidade do modelo em detectar amostras rotuladas como negativas pode ser calculada como Especificidade, Eq(2.19) (ZHU et al., 2010).

$$Acurácia = \frac{TP + TN}{TP + TN + FN + FP} \quad (2.17)$$

$$Sensibilidade = \frac{TP}{TP + FN} \quad (2.18)$$

$$Especificidade = \frac{TN}{TN + FP} \quad (2.19)$$

Em todo caso, tais resultados do grupo de dados de treino podem ser alterados de forma sistemática buscando um meio de aperfeiçoar o classificador. Fixando uma porcentagem

dos dados para teste, tal grupo pode ser alterado toda vez que o modelo é treinado e testado, repetindo esse processo até que todos os dados do conjunto sejam utilizados como conjunto de validação ao menos uma vez (Figura 5). Tal método *Cross Validation*, é um meio que busca evitar super ajustes em modelos supervisionados, sendo uma alternativa no lugar da seleção aleatória dos dados de validação (HANDELMAN et al., 2019).



Figura 5 – Representação do método *Cross Validation* na divisão da base de dados em um conjunto de validação e treino. Imagine que todo o conjunto de dados seja a barra em azul, a partir dele, dividiremos em um grupo de amostras usadas para treino do modelo de classificação (barra laranja) e outro grupo para testá-lo (barra em verde). O *Cross Validation* utiliza toda a base de dados como treino, selecionando um percentual para validação de forma que nenhuma amostra seja repetida. Fonte: Elaborada pelo autor.

Apesar de haver inúmeros meios de averiguar os resultados encontrados em *Machine Learning*, é um trabalho árduo em muitas situações, encontrar uma explicação plausível dos meios usados na manipulação das informações, podendo isso ser inacessível o que dificulta o entendimento do porquê o modelo chegou em tais resultados. Este impasse define os mecanismos da inteligência artificial como "caixa preta", por não ser precisamente claro justificar os fins atribuídos à base de dados. Em qualquer critério de classificação, os atributos são mapeados e associados a uma classe, definindo uma dada amostra. Dessa forma, uma alternativa, para dar mais clareza ao julgamento estatístico de decisão, é analisar o quanto impactante um atributo se torna para o resultado do modelo (BREIMAN, 2001).

A contribuição que o i -ésimo atributo proporciona ao modelo de classificação é definida como importância I_i . Em árvores de decisão, seu grau de relevância pode ser quantificado a partir das métricas de modelagem GINI ou Entropia, Eq.(2.20), ponderando estas por $p(n)$, que se trata do número de amostras em um nó n pelo número de amostras total (número de amostras na raiz) (??). Isso claro, desde que n pertença a arvore T .

$$I_i = \sum_{n \in T, i_n=i} p(n) \Delta_{GINI}(n) \quad (2.20)$$

A partir dos resultados encontrados com a Eq. (2.20), os valores I_i de cada variável são normalizados resultando em uma importância total igual a 1. Sendo assim, é possível atribuir uma porcentagem no poder de distinção das classes para cada variável.

2.3 Análise de Componentes Principais (PCA)

Ao analisar um conjunto de dados experimentais é conveniente expressá-los em gráficos para que se tenha uma noção mais intuitiva do que eles representam, ou até mesmo atribuir conclusões que não são de acesso imediato ao simplesmente averiguar uma tabela ou matriz numérica. Para isso, em meio a um conjunto de amostras, que contenha um número elevado de atributos(dimensões), é necessário encontrar um conjunto de dados que represente as amostras em dimensão menor, que facilite o tratamento de dados e possibilite uma visualização gráfica, mas que ainda preserve a informação do conjunto original. A análise de componentes principais (PCA - *Principal Component Analysis*) é um método de redução de dimensão que realiza tal papel, projetando os dados originais em eixos que apontam na direção de maior variância do conjunto de dados (GREENACRE et al., 2022).

Um dos métodos de construção da PCA é através de uma matriz de covariância. Em uma matriz de dados espectroscópicos, cujas linhas estão associadas às amostras e as colunas às variáveis de cada amostra o cálculo de covariância determina a relação linear entre dois atributos a partir de suas variâncias, expressa na equação seguinte Eq.(2.21):

$$Cov(m_j, m_k) = \sum_{i=1}^n \frac{(a_{ij} - \bar{a}_j)(a_{ik} - \bar{a}_k)}{n-1} \quad (2.21)$$

Onde na equação temos a_{ij} , a_{ik} , \bar{a}_j e \bar{a}_k valores numéricos do j-ésimo e k-ésimo atributo pertencentes a i-ésima amostra e os respectivos valores médios de cada variável. Os termos m_j e m_k representam os elementos Cov_{jk} da matriz de covariância e n o número total de amostras.

Em um conjunto de dados com 50 amostras (linhas) e 1000 atributos (colunas), a matriz de covariância terá uma dimensão de 1000x1000. Se tratando de uma matriz autoadjunto seus autovetores serão linearmente independentes entre si, podendo assim, serem usados como eixos ortogonais onde os dados originais serão projetados. Cada um dos autovetores é definido como uma PC (componente principal), que representa um percentual das amostras originais calculada a partir do seu autovalor. Tal parcela

de representatividade é denominada de variância explicada Eq. (2.22) (BRO; SMILDE, 2014).

$$Var_k = \frac{\lambda_k}{\sum_{i=1}^n \lambda_i} \quad (2.22)$$

onde o percentual de representatividade do k -ésimo componente principal é seu autovalor λ_k , dividido pelo soma de todos os autovalores pertencentes a matriz de covariância. Logo, a dimensão é reduzida fazendo o produto da matriz de dados original $M_{50 \times 1000}$ pela matriz de PCs(*loadings*) $P_{1000 \times 1000}$. O resultado de tal produto é denominado de *score*, onde cada coluna representa uma parcela dos dados originais.

Representando os espectros medidos em dois componentes principais (veja a Figura 6), temos que cada ponto do mapa de *score* é interpretado como um espectro. Devido a uma variância particular presente em cada tipo de amostra, surge a formação de grupos delimitado a partir do valor 0 da componente principal 1.

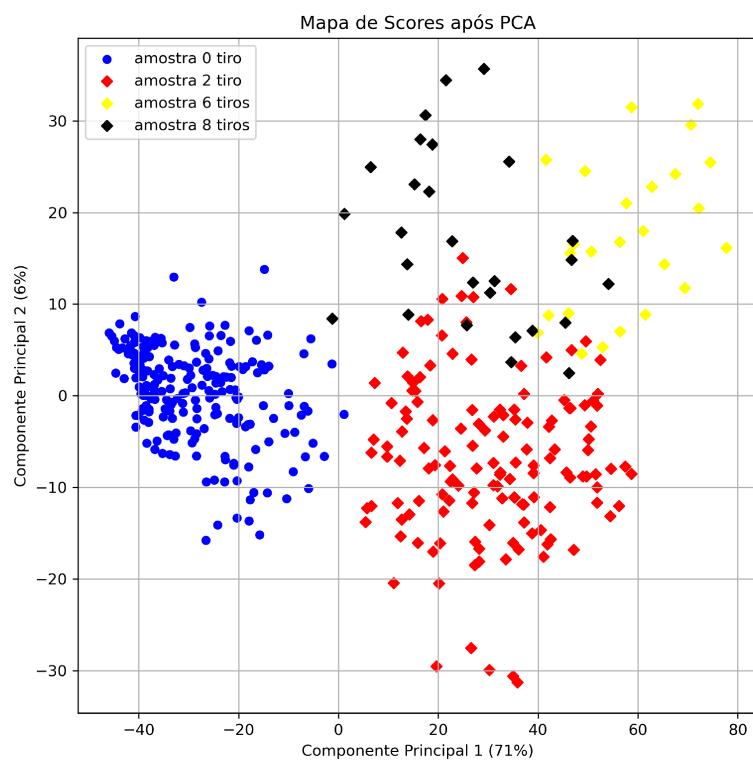


Figura 6 – Cada ponto do mapa de score é um espectro representado em 2d, construído a partir dos componentes principais mais representativos da matriz de covariância. Fonte: Elaborada pelo autor.

2.4 Correção de Linhas de Base

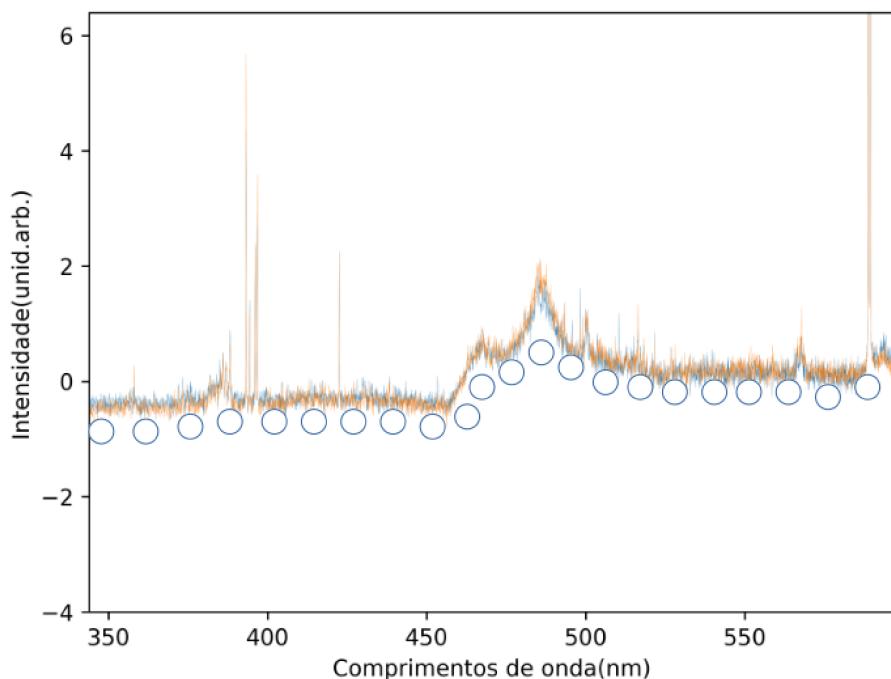
No processo de irradiação, as linhas de emissão atômica podem sofrer interferências, seja pelas condições em que o aparato experimental é condicionado, ou até mesmo por conta da alta temperatura do plasma. Tal ruído consiste de uma emissão contínua, deslocando a base das linhas espectrais, dificultando sua análise qualitativa. Uma forma de eliminar o ruído é corrigindo a base do espectro medido, sendo isso possível com o método *Rolling-balls* (SHVEDCHENKO; Suvorova, 2017).

$$f(x, y) = (g(x, y) + 1) - (C(x, y) - r) \quad (2.23)$$

O método *Rolling-balls* consiste de um algorítimo que desloca uma bola sobre a base do espectro Eq.(2.23). Os pontos tangente ao círculo formam um gráfico de fundo descritos a partir das coordenadas do seu centro $C(x, y)$ que é eliminada do espectro original $g(x, y)$, tendo como resultado uma base equivalente a cada um dos picos de emissão atômica $f(x, y)$. Manipulando o raio r do círculo modificamos a sensibilidade de exclusão (Rashed, 2016).

A Figura 7 - a apresenta o comportamento do *Rolling-balls* a partir de um valor de r igual a 2. Corrigindo todo o distanciamento da base em torno de 450 nm e 500 nm, as linhas de emissão atômica não são afetadas conservando a identidade do espectro (Figura 7 - b).

a)



b)

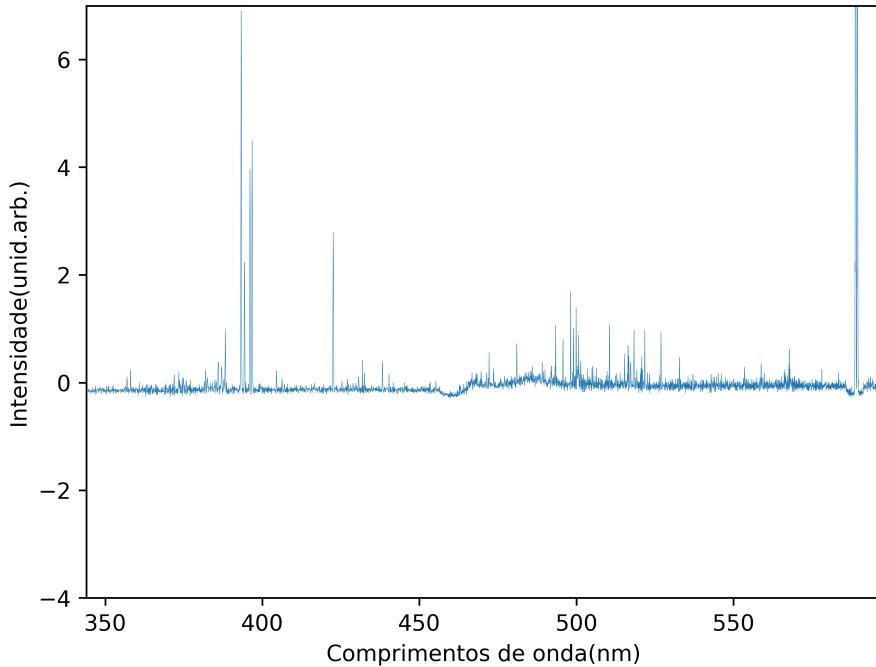


Figura 7 – A Figura a) apresenta um deslocamento da base devido ao ruído branco, onde a Figura b) apresenta sua correção. Fonte: Elaborada pelo autor.

2.5 Normalização dos Espectros

No tratamento de qualquer informação antes de sua aplicação, específica, independente de sua natureza, é conveniente manter as variáveis em mesma escala. Isso possui um apelo maior quando os dados numéricos constituem um significado que é interpretado graficamente. Desta forma, ao compararmos duas medidas, espectrais por exemplo, podemos afirmar que eventuais diferenças entre elas não são um efeito de escala, ou seja, a normalização traz mais confiabilidade na análise comparativa.

A normalização pode ser feita através de padronização em distribuições de probabilidade denominada de distribuição normal padrão (DHANOA et al., 1994):

$$Z_i = \frac{Z_r - \mu}{\sigma} \quad (2.24)$$

Z_i é o novo valor da variável real Z_r do espectro medido subtraído de sua média μ , e ambos divididos pelo desvio padrão σ . Após o cálculo envolvendo a Eq. (2.24), o novo espectro terá uma média centrada em $\mu = 0$ e desvio padrão $\sigma = 1$.

2.6 Eliminação de *Outliers*

Durante a aquisição experimental dos dados, cada medida está sujeita a um erro, seja na manipulação do equipamento, ou devido às características da amostra. Quando um

subconjunto de informações destoa significativamente das demais informações do conjunto, ele é definido como um *outlier*. No contexto da aquisição de espectros, um *outlier* pode ser, por exemplo, uma medida na qual muitos picos de emissão óptica não são registrados devido a pequenas flutuações na potência do pulso laser e irregularidades na superfície da amostra.

Uma forma de eliminar um *outlier* é através do método *Spectral Angle Mapper - SAM*. Assumindo que a média global dos espectros é um resultado representativo de todo o conjunto, e que cada medida se trata de um vetor multidimensional, podemos realizar um produto escalar entre a média global e cada espectro e eliminarmos os espectros que mais se diferenciam da média global (ZHANG; LI, 2014).

A partir da Eq.(2.25), os *outliers* podem ser identificados através do parâmetro C_{SAM} . Quantos mais próximo de 1 for $\cos \theta$, mais semelhante o espectro \vec{x}_i será da média \vec{X}_m do conjunto. Para as medidas LIBS, é conveniente ajustarmos $C_{SAM} = 0,8$, isso significa que todos aqueles espectros \vec{x}_i em que seu produto escalar com a média resultar em um valor $< 0,8$ será considerado um *outlier*.

$$\cos \theta = \frac{\vec{x}_i \cdot \vec{X}_m}{|\vec{x}_i| \cdot |\vec{X}_m|} \geq C_{SAM} \quad (2.25)$$

Com isso, tais métodos e suas aplicações no contexto experimental são apresentados no próximo capítulo.

CAPÍTULO 3

METODOLOGIA

Os procedimentos para execução do trabalho foram realizados em três etapas: (1) Produção e coleta de resíduos de disparo de armas de fogo GSR (*gunshot residue*) seguindo o protocolo operacional padrão de perícia criminal estabelecidos pela Secretaria Nacional de Segurança Pública (SENASA); (2) realização das medidas experimentais para coleta dos espectros LIBS das amostras de GSR e (3) análise e tratamento de dados espectroscópicos medidos via técnicas de quimiometria multivariada agregada a aprendizado de máquinas (machine learning).

3.1 Produção e Coleta de Resíduos de Disparo de Armas de Fogo GSR

Na primeira etapa, o material coletado do disparo de armas de fogo foi disponibilizado pela equipe adjunta ao Programa de Cooperação Acadêmica em Defesa Nacional (Procad), oriunda da Universidade Federal do Mato Grosso do Sul. A produção e coleta de amostras foi feita a partir de uma equipe de peritos criminais do Núcleo Regional de Criminalística (NRC/CR), seguindo os protocolos estabelecidos pela SENASP (Ministério da Justiça e Segurança Pública (Brasil), 2013) a fim de garantir que não ocorresse a contaminação das amostras. Tais resíduos foram coletados das regiões demarcadas na Figura 8, isso, a partir de fitas dupla - face.

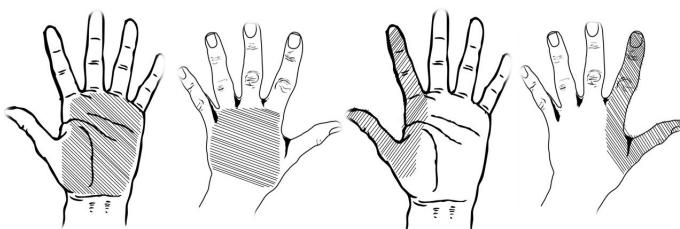


Figura 8 – Regiões da mão onde são coletadas os resquícios de arma de fogo. Fonte: Elaborada pelo autor.

As amostras de resquícios de disparo (GSR) coletadas no stande de tiro são provenientes de dois tipos de munição: Uma convencional nomeada aqui de munição - A, e as NTA (*Non-Toxic Ammunition*) definidas no mercado como ecológicas por serem livres de metais pesados (Ba, Pb, Sb). As fitas dupla - face foram armazenadas em placas de Petri, como ilustra a Figura 9. Cada amostra coletada, foi separada e nomeada como indicado na tabela 1.

Tabela 1 – Amostras coletadas. Fonte: Elaborado pelo autor.

Tipo de Amostra	Número de Disparos
NTA	2, 6 e 8
Munição - A	1 e 6
Sem contato com a arma de fogo	0

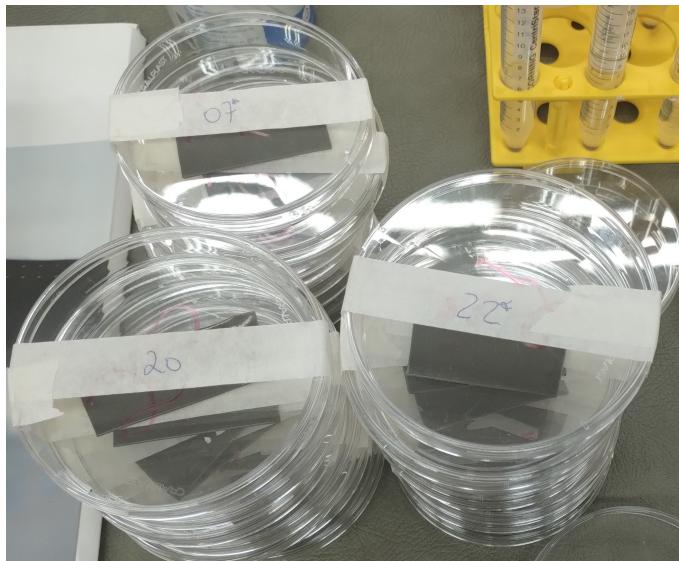


Figura 9 – Fitas dupla-face contendo as amostras de resquícios de disparo. Fonte: Elaborada pelo autor.

3.2 Medidas experimentais LIBS

Na segunda etapa, os procedimentos experimentais consistiram na aquisição de dados a partir da Espectroscopia de Emissão Ótica com Plasma Induzido por laser (LIBS). Tal técnica de espectroscopia LIBS utiliza o processo de ablação (deterioração explosiva) e excitação da amostra levando os átomos, moléculas, íons e elétrons a um estado de plasma, que por sua vez, emitirá luz. A partir da luz coletada via cabos ópticos e conduzida para um conjunto de espectrômetros é possível identificar as transições atômica e moleculares realizadas nesse processo de emissão ótica pelos elementos constituintes da amostra de GSR através das linhas espectrais de emissão do espectro LIBS (veja Figura

10). As medidas LIBS foram realizadas no equipamento LIBS de bancada, modelo J200 e fabricante *Applied Spectra* oriunda da Universidade Federal de São Carlos no grupo de pesquisa GAIA.

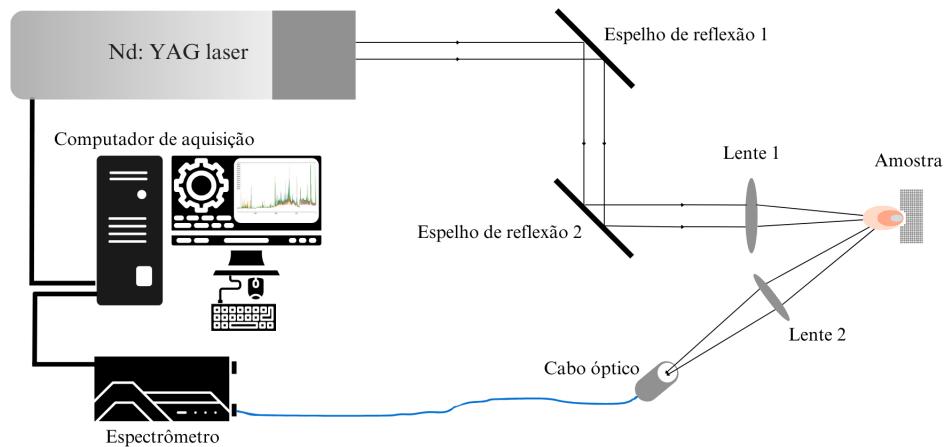


Figura 10 – Diagrama esquemático do sistema experimental para medidas LIBS. Fonte: Elaborada pelo autor.

Para execução experimental o sistema J200 é equipado com um laser Nd:YAG de comprimento de onda fundamental em 1064 nm e energia máxima na ordem de 50 mJ, no qual foi submetido às seguintes configurações: os disparos dos pulsos lasers foram realizados em uma grade 5x5 totalizando 25 disparos em cada fita (veja Figura 11), isso, em 100 % da potência máxima aplicada em um spot de 150 μm de diâmetro. Cada disparo laser foi espaçado em um intervalo de tempo de 3 s, com um atraso de 0,3 μs entre o pulso laser e a aquisição pelos espectrômetros (*delay time*).

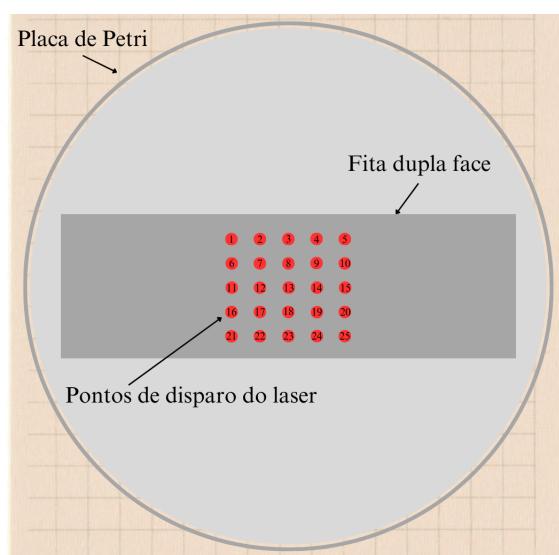


Figura 11 – Representação das regiões de disparo com os pulsos laser na fita dupla face. Fonte: Elaborado pelo autor.

3.3 Pré-Tratamento e Análise dos Espectros

A terceira etapa, consiste na análise e discriminação dos dados, com a utilização de rotinas computacionais desenvolvidas neste projeto em linguagem Python. Substâncias químicas encontradas em GSR podem fazer parte de outras situações que não envolvem armas de fogo, dificultando os meios em discernir um atirador de um não atirador. Logo, a identificação dos elementos químicos das amostras foi realizada a partir do banco de dados de transições atômicas de elementos químicos NIST (National Institute of Standards and Technology).

Na tentativa de contornar tal situação foi utilizada a análise multivariada de dados, que é capaz de correlacionar múltiplas variáveis pertinente a um dado evento e mostrar com clareza a correlação entre ambas. Assim, podemos ser capazes de excluir dados que forneçam falsos positivos e atribuir a um software de aprendizagem de máquina quais informações são cruciais na detecção de um atirador a partir dos dados em espectroscopia LIBS das amostras GSR.

Para efeito do mesmo, na tentativa de reduzir erros experimentais foram feitas médias a cada dois espectros LIBS, em cada tipo de amostra. Em seguida foi aplicado um pré-tratamento com os métodos: (i) *Rolling-Balls* em um raio de ajuste igual a 10; (ii) *Standard Normal Variate* para normalização e (iii) *Spectral Angle Mapper* para remoção de *outliers* com um limite de $\cos \theta \geq 0,95$ restando um total de amostras destacado na tabela 2. Feito isso, 70% dos espectros foram selecionados de forma aleatória e usados como treino nos modelos de classificação *Random Forest* e KNN e os 30% restantes foram usados para teste e plotados com sua dimensão reduzida a partir do PCA. O *Random Forest* foi aplicado com 184 estimadores e KNN usado com 7 vizinhos após um processo de otimização deses parâmetros.

Tabela 2 – Número total de espectros restantes com a remoção de *outliers*. Fonte: Elaborada pelo autor.

Tipo de Amostra	Número de Espectros Registrados	Pós-processamento
NTA - 2	625	138
NTA - 6	85	25
NTA - 8	85	27
Munição - A 1	50	20
Munição - A 6	50	25
Sem contato com a arma de fogo	650	218

Sendo assim, com os classificadores foi possível dimensionar o quanto distinto são as amostras pertinentes a um atirador e a um não atirador, e, a partir deles, averiguar o grau de precisão na identificação de um suspeito. Por fim, justificando os processos estatísticos usados nos modelos, foram calculadas os atributos (transições atômicas) que mais contribuem na distinção entre as amostras. No próximo capítulo são apresentadas tais implicações justificadas a partir de seus resultados.

CAPÍTULO 4

Análise dos Resultados

4.1 Identificação dos Marcadores Químicos

A partir dos pré-tratamentos abordados na metodologia, padronizando todos os espectros a fim de diminuir a margem de erro, encontramos os seguintes resultados: Os elementos químicos encontrados nas amostras de 2, 6 e 8 disparos NTA, que não estão presentes em nenhuma das amostras 0 disparo (indivíduo sem contato com a arma), ou seja, os marcadores químicos, são; Cu I, Cu II, Ti I e Mo I, todos identificados e apresentados nas figuras desta seção (veja a Figura 12).

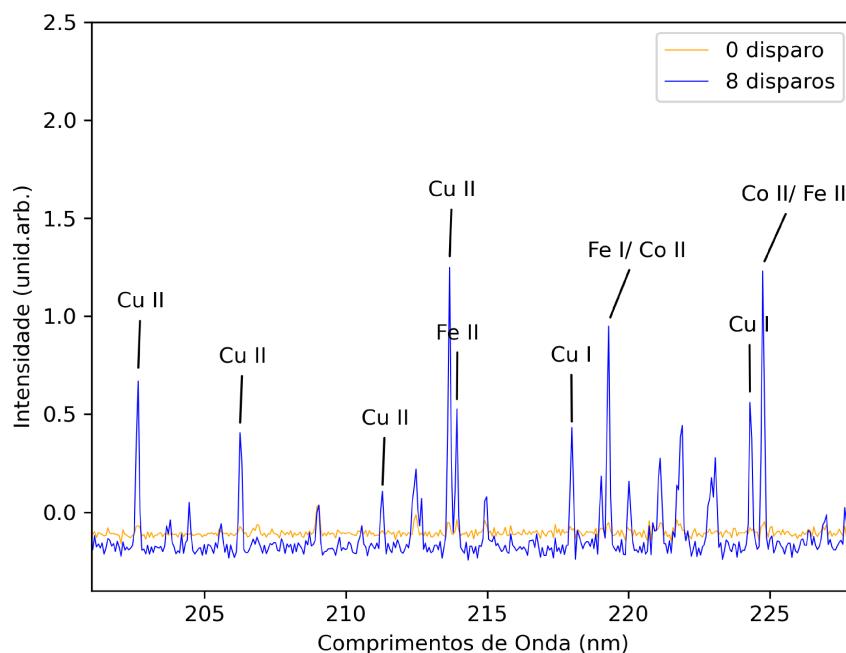


Figura 12 – Espectro médio de 8 e 0 disparos apresentado no intervalo de 205 - 225 nm.
Fonte: Elaborado pelo autor.

Analizando em detalhes o espectro médio no intervalo de 205 nm até 225 nm (veja a Figura 13) onde há uma concentração maior da linhas de emissão dos marcadores químicos Cu I (217,9 nm; 224,30 nm; 249,36 nm) e Cu II (202,64 nm; 206,25 nm; 211,29 nm; 213,66 nm; 270,46 nm), comparando os espectros de 2 e 8 disparos não é constatado uma diferença significativa na intensidade entre os picos de transição. Logo, com apenas dois disparo é possível identificar elementos marcadores da munição NTA com resolução suficientemente boa para deferir do espectro 0 disparo (Figura 13).

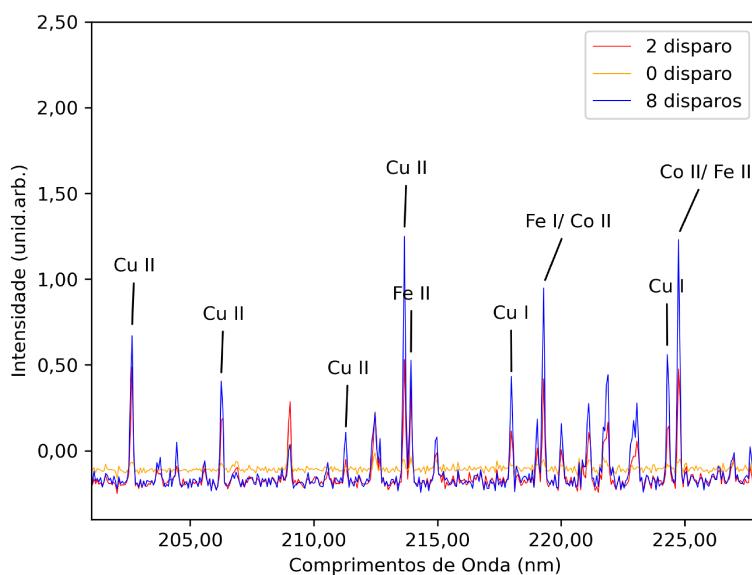


Figura 13 – Espectro médio de 2, 8 disparos NTA e 0 disparos, no intervalo de 205 - 225 nm. Mesmo com apenas dois disparo, é possível coletar amostras em quantidades o suficiente para ter uma resolução clara das linhas de transição. Fonte: Elaborado pelo autor.

Os elementos químicos utilizados na fabricação de aço inoxidável formam ligas metálicas junto a Titânio, Crômio e Molibdênio. O aço inox é bastante usual na fabricação de canos de armas de fogo, graças a sua resistência a corrosão e altas temperaturas como apresentado na ficha técnica do modelo Série 700 (TAURUS, 2016), no entanto, não é dado detalhes de qual tipo de aço é utilizado.

As Figuras 14 e 15 destacam linhas de emissão associadas a alguns de tais metais inerentes à fabricação de aço inox. Isso pressupõe, que de fato ao disparar uma arma de fogo, a presença de raiamento gera condições abrasivas, entre o cano e o projétil, grandes o suficiente para que resquícios do cano da arma sejam depositados nas mãos de atiradores como citado na introdução deste trabalho. E não menos importante, isso ainda é detectável pela técnica experimental LIBS mesmo se tratando de apenas um disparo (veja Figura 14).

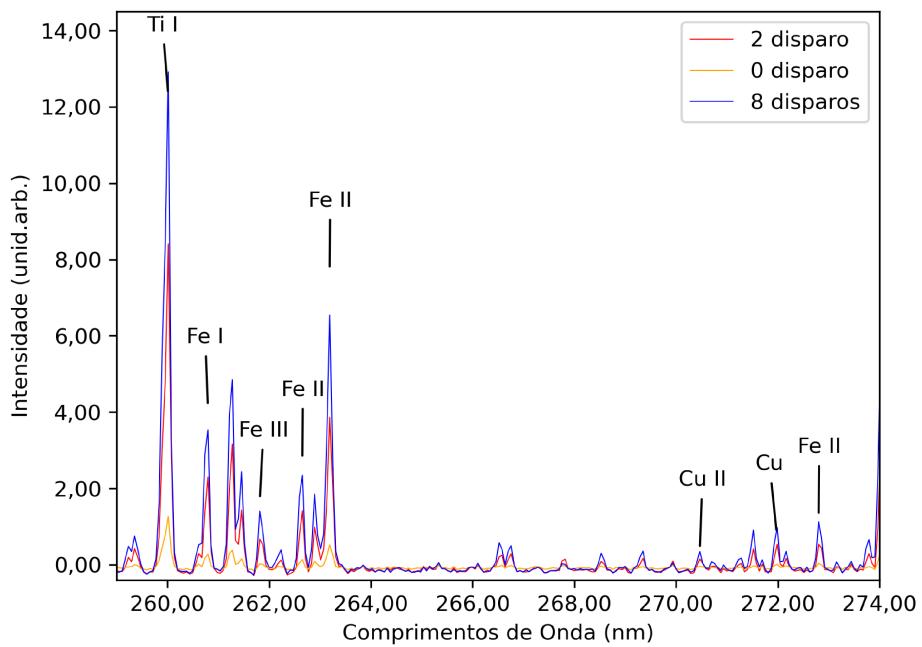


Figura 14 – Espectro médio de 0, 2 e 8 disparos no intervalo de 260 - 274 nm destacando marcadores químicos Cu II e Ti I. Fonte: Elaborada pelo autor

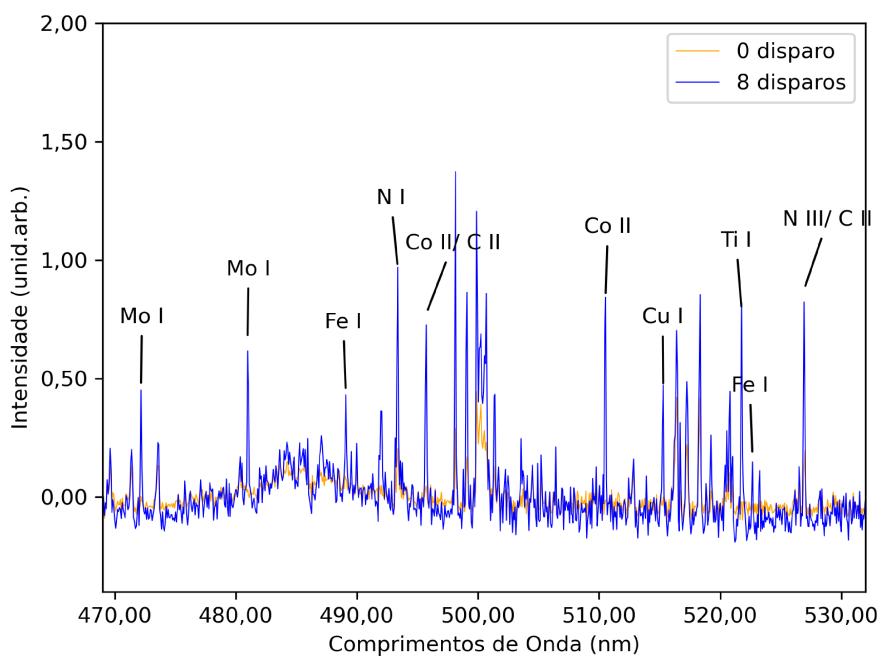


Figura 15 – Espectro médio de 8 e 0 disparos no intervalo de 470 - 530 nm destacando a presença de Compostos químicos de aço inox. Fonte: Elaborada pelo autor

Sendo mais preciso, os elementos químicos pertinentes a confecção de aço foram en-

contrados nos espectros em Ti I (260 nm , 521,7 nm), Mo I em (472,1 nm; 480,9 nm). Tanto em amostras de disparos com NTA e 0 disparo, identificamos linhas de K I (404,4 nm; 578,1 nm; 766,4 nm; 769,9 nm), N I (744,2 nm; 746,8 nm; 821,6 nm; 824,3 nm), que são provenientes provavelmente de sais minerais da mão e da atmosfera. A tabela 3 apresenta todas as linhas de emissão identificadas.

Tabela 3 – Linhas de emissão atômica medidas. Fonte: Elaborada pelo autor.

Linhas de Emissão	Comprimentos de Onda(nm)
Cu II	202,647; 206,253; 211,293; 213,667; 270,466;
Cu I	224,300; 249,365;
Co II	233,849; 238,277; 238,906; 242,480; 258,630; 510,537
Fe II	213,930; 233,34; 234,420; 239,598; 241,104; 251,699 253,844; 262,653; 263,190; 271,519; 274,025; 327,332
Ti I	260,016; 521,775;
Mn II	298,581;
Mn I	279,616; 472,174; 480,974
Mg II	280,36
Mg I	279,616
Mo I	472,174; 480,974
Sr I	489,086
N I	493,38; 744,281; 746,875; 821,66; 824,317
K I	578,158; 766,491; 769,917

A fim de garantir imparcialidade nos dados de espectroscopia ao diferenciar um atirador de um não atirador, os resultados a seguir partiram da análise em um modelo de classificação, aplicando métricas estatísticas para quantificar informações presentes nos espectros medidos.

4.2 Resultados Random Forest

Usando uma base de dados homogênea com 190 espectros de atiradores (2,6 e 8 disparos) e 190 espectros de não atiradores (0 disparo), e concentrando a análise no intervalo de 184 nm a 380 nm onde está localizado uma maior densidade de picos de transição eletrônica dos marcadores químicos Cu I, Cu II e Ti I. Temos os seguintes resultados; Usando 30 % dos dados para teste, o modelo de classificação selecionou de forma aleatória 58 espectros de 0 disparo, e 56 espectros de disparo para treino.

Nessas condições a acurácia a partir de 184 estimadores se encontra em 98 %. Analisando cada uma das classes, a precisão em não atiradores (Especificidade) se estabelece em 98 %, se mantendo igual as amostras de disparo de arma (Sensibilidade).

Dos erros de classificação apresentados na matriz de confusão (veja a Figura 16), em 58 espectros de não atiradores (classe - 0), 57 deles foram classificados corretamente e apenas 1 de forma incorreta. De forma equivalente, aos 56 espectros de atiradores (classe – 1),

temos que o modelo foi capaz de classificar os 55 espectros corretamente apresentando 1 falso positivo. Sendo assim, diante do número de erros, a de se constatar que o modelo não favorece uma classe em específico, visto que o número de erros é equivalente em ambas as classes.

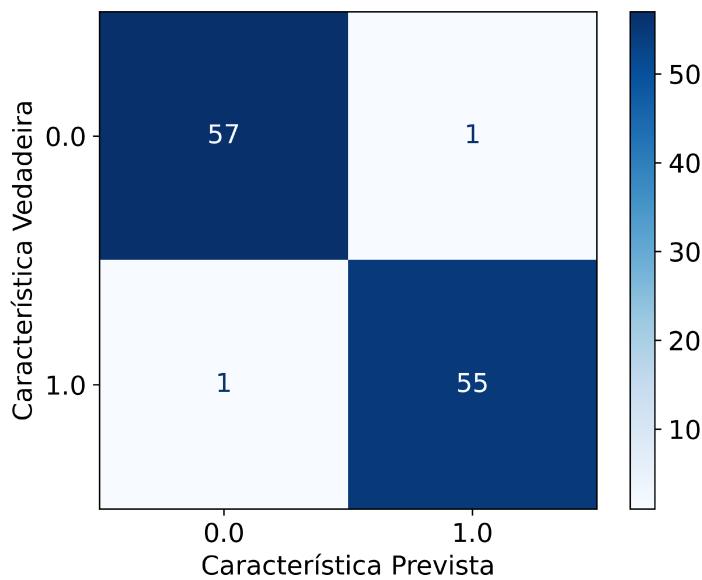


Figura 16 – Matriz de confusão das classes, 0 - não atirador, 1 - atirador. Fonte: Elaborada pelo autor

Tal distinção apresentada no modelo treinado é refletida nos picos de transição eletrônica dos espectros. O modelo julga as transições com uma importância de aproximadamente 82%, ou seja, durante o treino do modelo os elementos químicos são as características mais relevantes no momento de distinguir os dados, o demais dados numéricos são fisicamente interpretados como ruído.

Tal relevância é expressa no gráficos em barras (veja a Figura 17). Quanto maior a barra mais importante é tal característica, sendo assim, fica claro a irrelevância dos ruídos para o modelo perante as transições eletrônicas, como apresentado no intervalo de 230 – 245 nm. Olhando para os comprimentos de onda, sendo elas interpretadas como características pelo classificador, o marcador químico Ti I (260,076 nm) e Fe II (288,262 nm; 288,206 nm) se estabelecem sendo as duas primeiras variáveis mais relevantes durante a discriminação das classes (veja a Figura 18). Isso implica que o modelo foi bem treinado, se baseando em que há uma concordância entre o classificador e a identificação química.

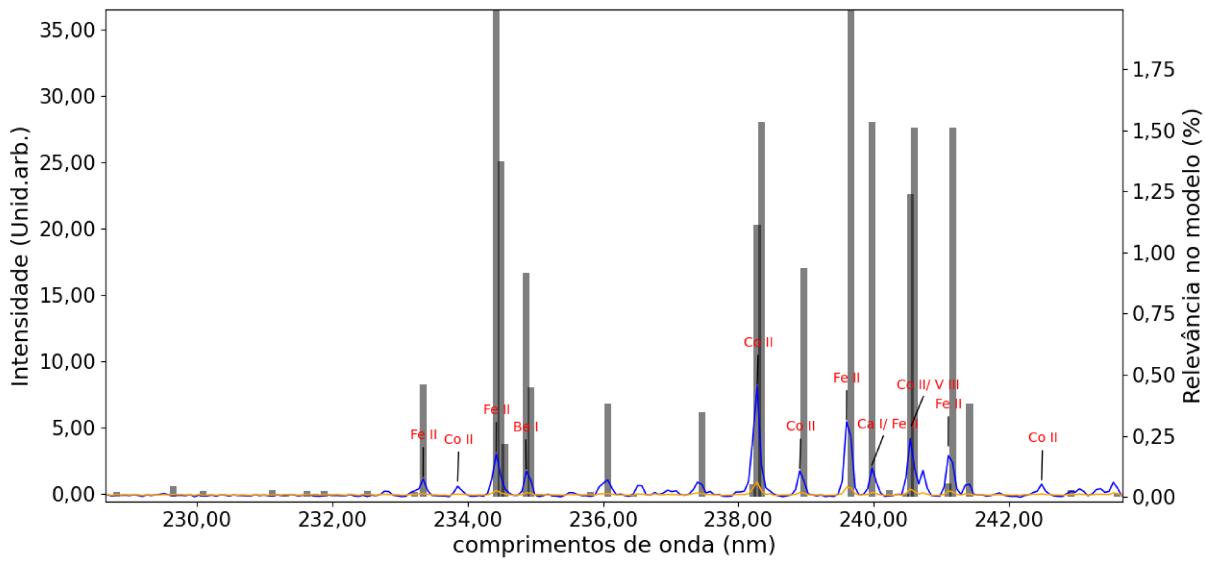


Figura 17 – O modelo de classificação estabelece uma medida estatística(%) de relevância para cada uma das características dos dados, sendo essas descritas em barras no gráfico. Fonte: Elaborada pelo autor.

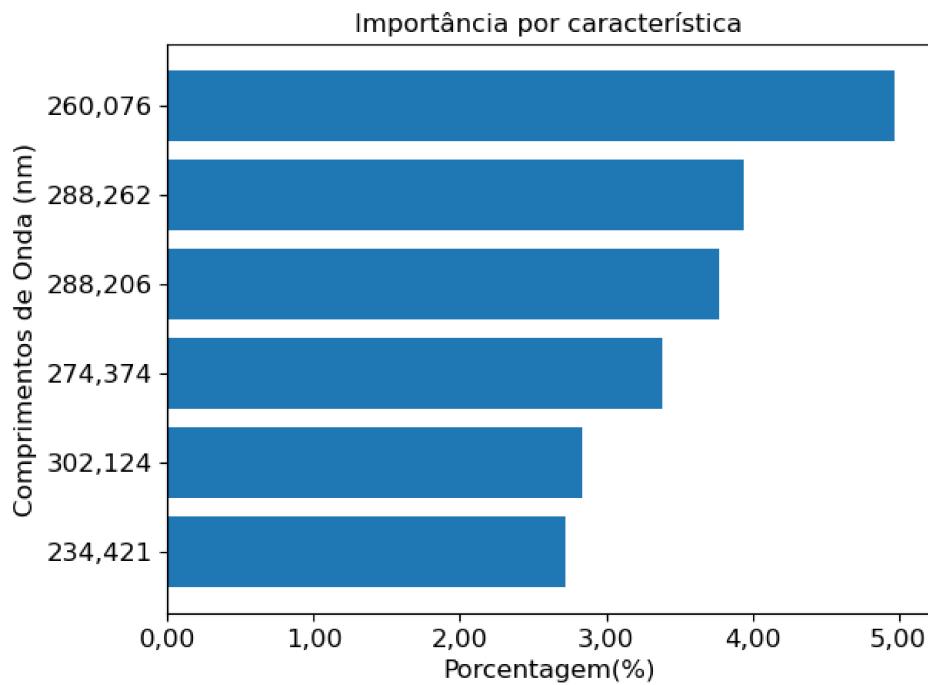


Figura 18 – As 5 características mais relevantes na discriminação das classes (0 e 1) são os elementos químicos: Ti I(260,076 nm), Fe II(288,262 nm, 288,206 nm), Fe I(274,374 nm), Fe I(302,124 nm) e Fe II(234,421 nm). Fonte: Elaborada pelo autor

Tal distinção entre as amostras, além de se manifestarem nos percentuais de relevância

entre características, também é possível visualizar a formação de grupos através do PCA dos espectros. A redução de dimensão foi realizada nos 30 % dos dados usados como teste do modelo após sua classificação, destacando em vermelho espectros classificados como 0 disparo, e em verde espectros classificados como atirador (veja a Figura 19). Como resultado, as componentes principais 1 e 2 de tal operação correspondem a 83 % dos dados, definindo dois grupos um a esquerda em vermelho e outro a direita em verde.

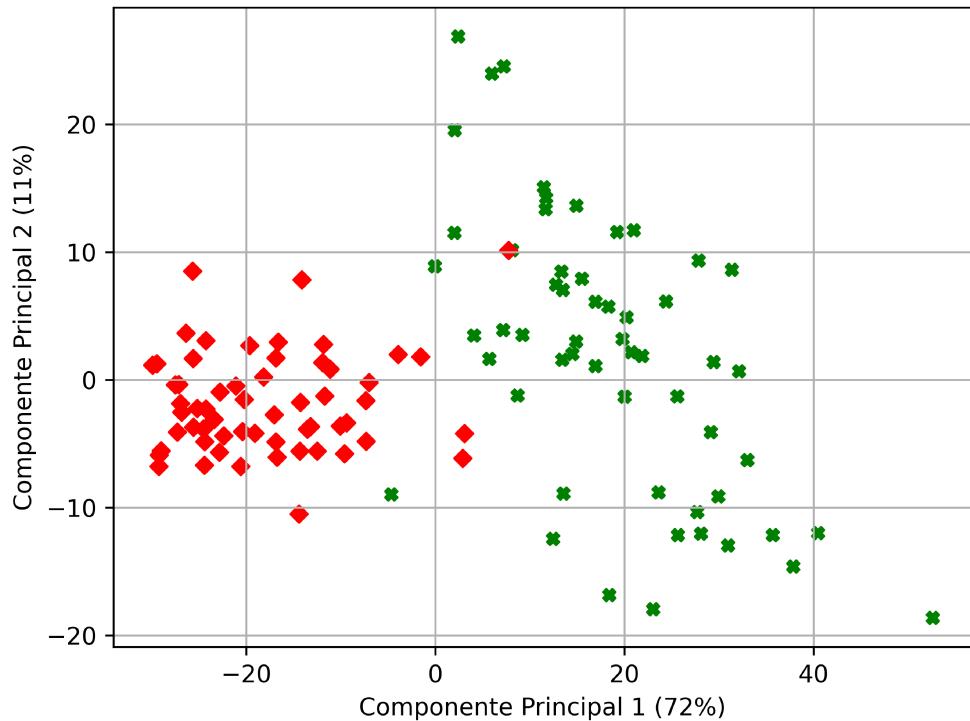


Figura 19 – Cada ponto é um espectro representado em 2d, onde em vermelho está a classe - 0 e em verde a classe - 1. Comparando esta figura com a matriz de confusão (Figura 17), o falso - positivo e o falso - negativo se destacam como um ponto verde em meio ao grupo da esquerda e um ponto vermelho em meio ao grupo da direita. Fonte: Elaborada pelo autor.

Portanto, afim de validar o grau de acurácia que garanta a fidelidade nos resultados fornecido nos parágrafos anteriores, a validação cruzada com 10 grupos de amostras a partir do método *K-fold* no modelo *Random Forest*, apresenta uma acurácia média de 99 % com desvio padrão de 0,01. Além disso, a título de comparação, usando também o classificador KNN(*k-Nearest Neighbors*) e aplicando o processo de validação cruzada com os mesmos parâmetros descritos aqui, é obtido o mesmo resultado de acurácia e desvio padrão encontrados no modelo *Random Forest*.

Na próxima seção, os mesmos parâmetros utilizados aqui nos modelos de classificação serão aplicados as amostras de 0 e 1 disparo NTA afim de validar os argumentos discutidos

a respeito da capacidade de detecção de marcadores químicos com apenas 1 disparo com arma de fogo.

4.2.1 *Random Forest* aplicado a espectros de 0 e 2 disparos NTA

No intuito de justificar os argumentos apresentados na sessão 4.1 (identificação de marcadores), mais precisamente nos dados abordados na Figura 13, é apresentado os seguintes resultados de classificação: Usando os mesmo parâmetros do modelo apresentado na sessão anterior, mas, em apenas 138 espectros de 0 e 2 disparos, é encontrado uma acurácia de 98%. Como consequência disso, olhando para sua matriz de confusão o modelo apresenta 2 falsos positivos e nenhum falso negativo (Figura 20), tendo assim, uma Especificidade de 100 % das amostras de classe - 0 (nenhum disparo) e 95% de Sensibilidade em amostras de classe - 1 (apenas dois disparos).

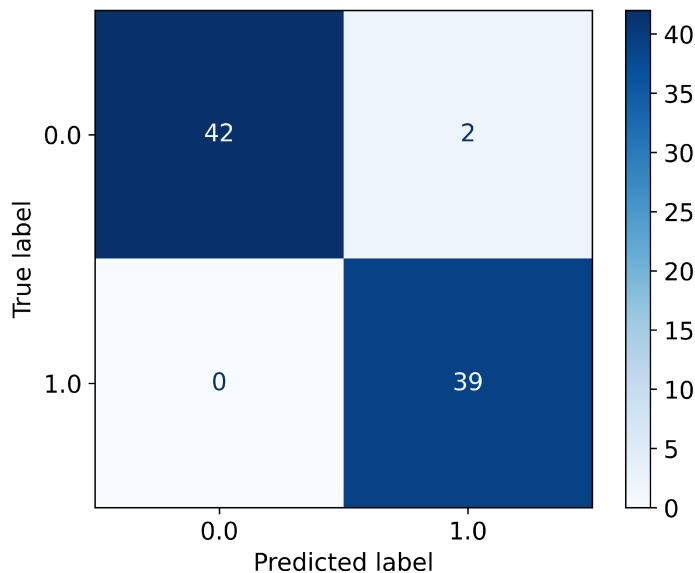


Figura 20 – Matriz de confusão do modelo *Random Forest* treinado a partir de 138 espectros de 0 e 2 disparos. Onde, as classes são definidas aqui como, 0 - nenhum disparo e 1 - apenas dois disparos. Fonte: Elaborada pelo autor.

E não menos importante, visualizando tais espectros classificados, através de componentes principais, é notável a formação de dois grupos assim como na Figura 19 . Onde a esquerda temos espectros classificados como nenhum disparo (classe - 0) e a direita 2 disparos (classe - 1) (Figura 21). Sendo assim, o número de disparos na arma de fogo não influencia significativamente nem mesmo em sua classificação.

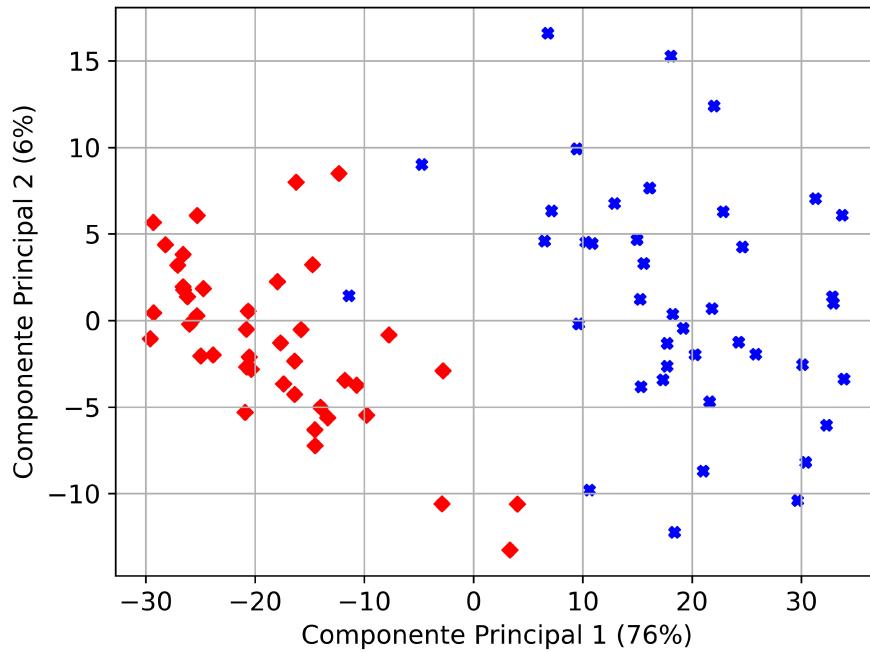


Figura 21 – A esquerda em vermelho temos os espectros classificados como 0 disparo, e a direita em azul os espectros classificados como 2 disparos. Fonte: Elaborada pelo autor.

Tal resultado corrobora a identificação de marcadores químicos em amostras NTA, sendo estes: Ti I (260 nm , 521,7 nm), Mo I em (472,1 nm; 480,9 nm), Cu I(217,9 nm; 224,30 nm; 249,36 nm) e Cu II(202,64 nm; 206,25 nm; 211,29 nm; 213,66 nm; 270,46 nm). Avaliando a distinção entre munições convencionais e NTA, a próxima sessão destacará o poder de generalização do modelo de classificação *Random Forest*.

4.3 Distinção Entre Munições Convencionais e NTA

Plotando e comparando o espectro médio de ambas as amostras convencionais e NTA, a região de 320 nm – 410 nm (Figura 22) onde se encontra os marcadores da munição convencional Pb I, Sb I, se presenta completamente ausente na amostra de NTA. O equivalente se aplica às munições convencionais. Olhando para a região onde há uma maior concentração dos marcadores químicos de NTA 204 – 225 nm (Figura 23), não é constatado nenhum sinal de transição eletrônica nas amostras de munições convencionais.

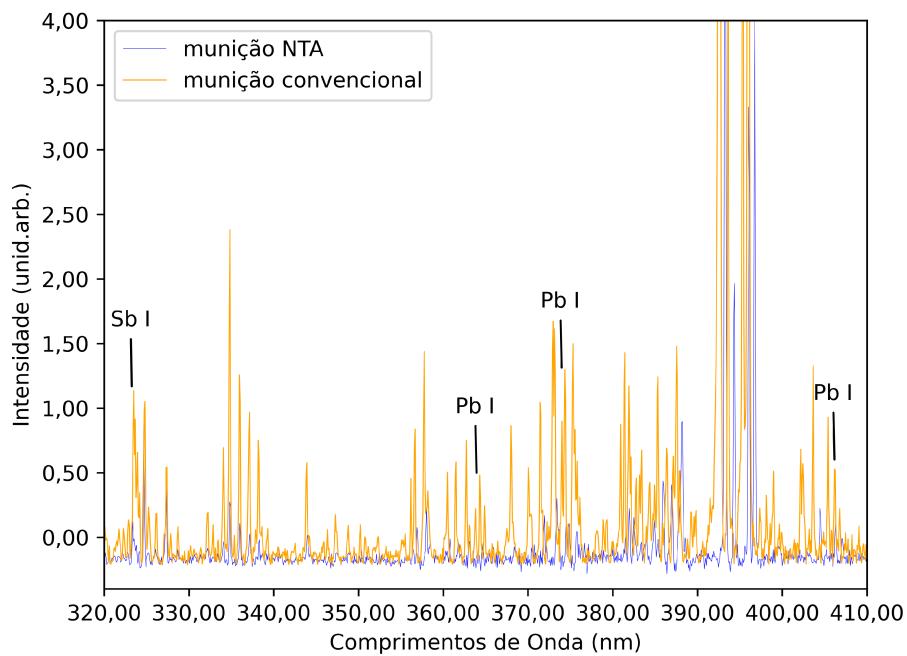


Figura 22 – Espectro médio das amostras de 8 disparos - NTA e 6 disparos - convencional, em uma das regiões onde se encontram os marcadores químicos da munição convencional, Sb I (antimônio) e Pb I (chumbo). Fonte: Elaborada pelo autor.

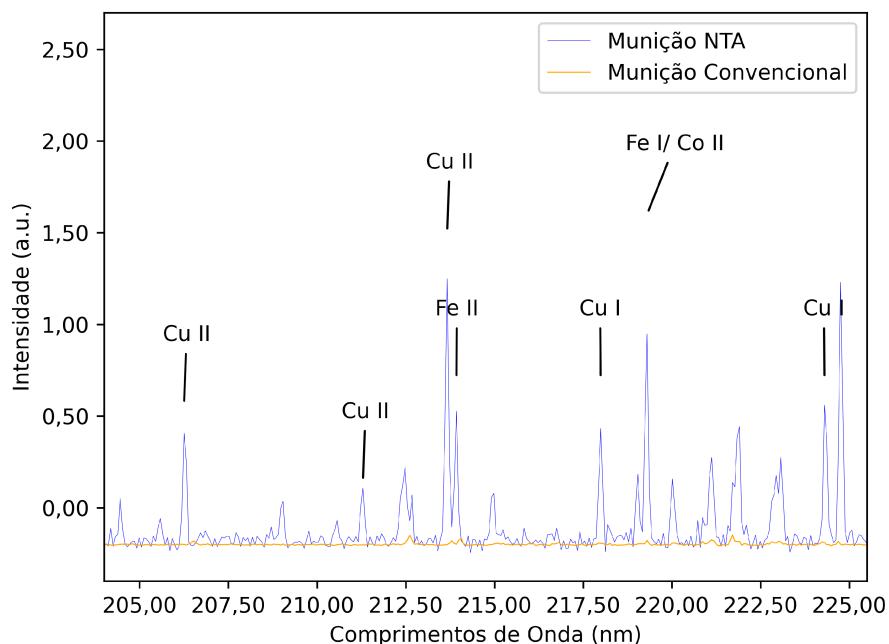


Figura 23 – Espectro médio das amostras de 8 disparos - NTA e 6 disparos - convencional, em uma das regiões onde se encontram os marcadores químicos de NTA. Fonte: Elaborado pelo autor.

Além disso, usando o classificador como uma ferramenta que é capaz de definir o quanto distinto são as duas amostras, o modelo *Random Forest* apresenta uma acurácia de 100%. Tamanha é a distinção entre as amostras, que visualizando ambas a partir do PCA (Figura 24), cada uma se encontra em dois grupos bem definidos, onde a munição convencional se estabelece a esquerda em azul, enquanto que a direita em amarelo se encontram as amostras de NTA.

Sendo assim, mesmo ambas as amostras pertencendo a munições de armas de fogo, as duas são absolutamente distintas em seu padrão espectroscópico e nos marcadores químicos que as definem, justificando a necessidade de um protocolo experimental que atenda incidentes com NTA.

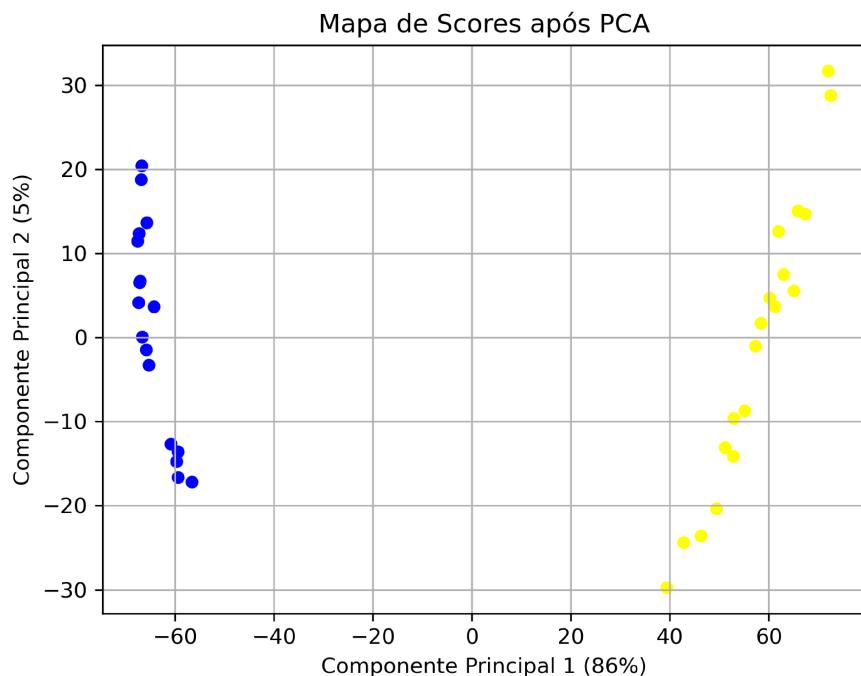


Figura 24 – A esquerda em azul, amostras classificadas como munição convencional, enquanto que a direita em amarelo, espectros classificados como NTA. Fonte: Elaborada pelo autor.

Partindo dessa distinção clara apresentada pelo modelo de classificação, foi testado seu poder de generalização treinando com 218 espectros de 0 disparo, e 218 espectros de todas as amostras de disparo de arma de fogo, ou seja, tanto munição NTA (2, 6 e 8 disparos) quanto convencionais (1, 6). Selecionando 30% dos espectros para teste e 70% para treino, e rotulando como 0 espectros de 0 disparo e 1 para espectros de disparo, independente de sua natureza, temos os seguintes resultados: O modelo de classificação *Random Forest* apresenta uma acurácia de 99%, no entanto, a classe de atirados apresenta o mesmo comportamento de super ajuste quando comparadas as amostras de NTA e munição convencional.

A métrica Sensibilidade do modelo resulta em 100% enquanto que a Especificidade tem um resultado de 98%. Nesta configuração, o modelo foi treinado com todo o espectro registrado, ou seja, usando o intervalo de 186 nm - 1042 nm, e mesmo com a presença de espectros de 0 disparo, o modelo ainda super ajusta a classe de atiradores perdendo generalidade. No PCA das amostras classificadas, é encontrado uma variância explicada de 86% para as PCs 1, 2 e 3. No gráfico de *score* (veja Figura 25), o padrão encontrado entre as amostras de NTA e convencionais ainda se mantém ao longo da componente principal 1, sugerindo uma distinção estatística maior das amostras de 0 disparo em relação as munições convencionais.

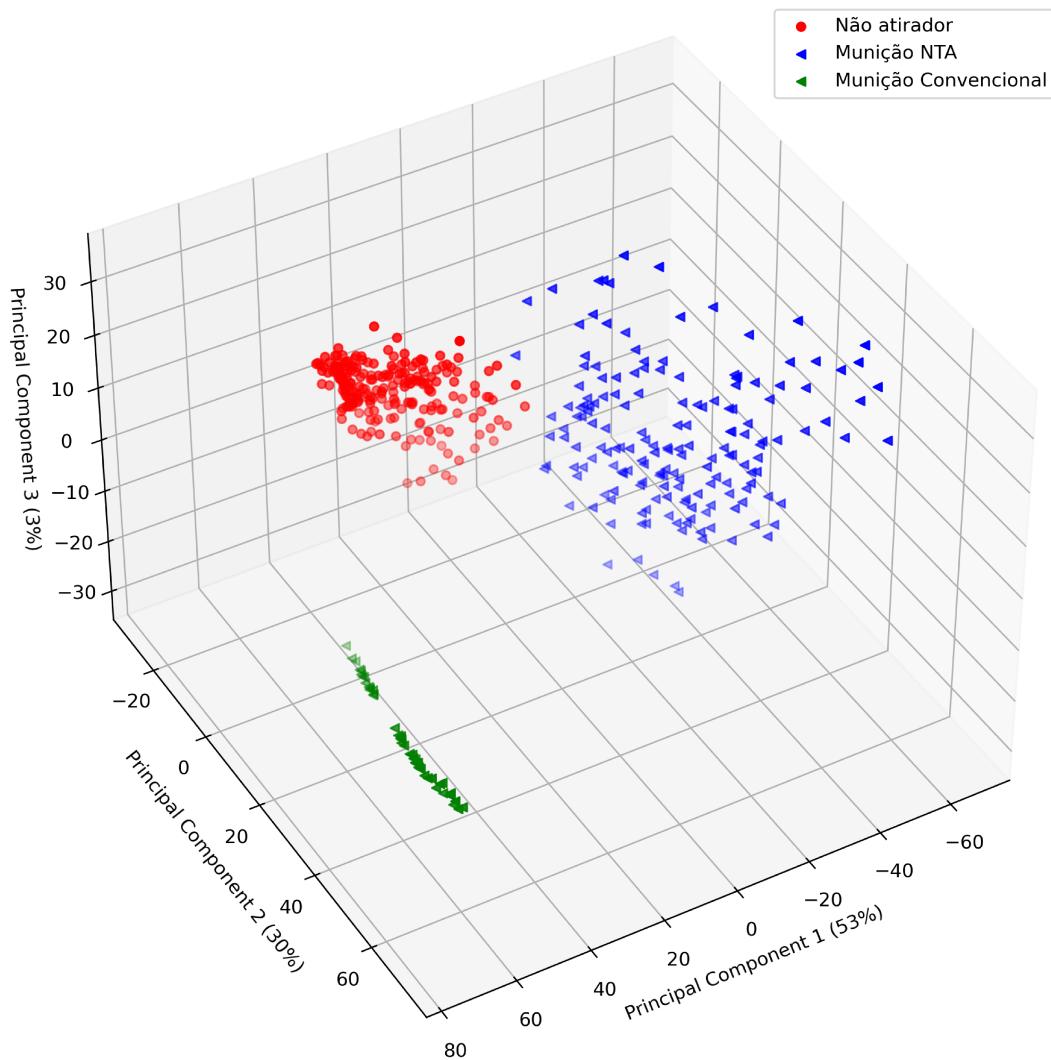


Figura 25 – PCA dos espectros classificados. Ao fundo em vermelho se encontra o grupo de amostras 0 disparo, e em azul e verde o grupo de atirados que se destacam em duas regiões ao longo da componente principal 1. Fonte: Elaborada pelo autor.

No próximo capítulo serão tratadas as conclusões acerca dos resultados apresentados nesta dissertação, tornando mais claro a necessidade de alguns métodos e suas implicações.

CAPÍTULO 5

Conclusão e Perspectivas

Como destacado na introdução deste trabalho, as munições aqui denominadas como convencionais possui marcadores claros por apresentar três elementos químicos juntos na mesma amostra, Ba, Pb e Sb. Os resultados deste trabalho mostram que a técnica experimental LIBS, possui um poder de detecção clara até mesmo de marcadores de munições NTA. E não menos importante, a técnica se mostra viável mesmo diante de apenas 2 disparos com a arma de fogo, registrando espectros característicos com uma resolução satisfatória.

Quanto aos marcadores químicos de NTA, a presença detectada de linhas de emissão atômica de Cu I(217,9 nm, 224,30 nm, 249,36 nm), Ti I (260 nm , 521,7 nm), e Mo I (472,1 nm, 480,9 nm), apenas nas amostras de NTA, não só revelam a sua ausência nas mãos de um indivíduo não portador de arma de fogo, mas também, as condições para que isso aconteça. Ou seja, mesmo havendo utensílios domésticos feitos de aço inox que venham a ter tais elementos químicos, os resultados apontam que provavelmente as condições abrasivas capazes de depositar tais resquícios na mão humana está relacionada ao disparo de uma arma, e não de atividades cotidianas. Sendo assim, os espectros constatam a existência de um padrão químico, que também, se manifestam de forma estatística no modelo de classificação.

As amostras de 0 disparo e NTA possuem tamanha distinção, que com apenas um disparo é possível alcançar uma porcentagem próxima de 98% ao identificar os portadores de arma de fogo. O mais relevante se enquadra nos atributos selecionados como nós de decisão do modelo. Ao executar o treinamento de aprendizagem de máquina, tanto ruído quanto linhas de emissão atômica dos espectros são comparados entre si, e mesmo sendo minoria, as linhas de emissão são os atributos mais relevantes nas decisões de classificação do modelo. Isso significa que os cálculos estatísticos, ou seja, a métrica de informação de Shannon da destaque a atributos que possuem uma interpretação física, neste caso, os elementos químicos.

Portanto, com um resultado em torno de 99% nas classificações a partir do *Random Forest*, onde também é encontrado pelo método de validação cruzada e através do método

KNN (apenas citado nos resultados), é constatado que não há enviesamento por parte dos modelos. Isso também se deve ao fato do número de amostras utilizadas para teste e treino serem balanceadas, seguindo uma uniformidade na quantidade de amostras de 0 disparos e NTA. Graças a isso, o protocolo experimental aqui proposto cumpriu seu objetivo e se mostrou como mais um método a contribuir para a ciência forense.

Como perspectiva futura, podemos propor testes mais robustos do modelo de classificação via espectros LIBS com dados de situações reais. Em se tratando de modelos de aprendizagem de máquina, é relevante a aplicação em um número estatisticamente representativo de amostras, testando as mais diversas situações possíveis.

Esta situação pode ser simulada aplicando métodos de produção de dados artificiais a partir dos espetros registrados em laboratório. Como um trabalho futuro, poderia ser aplicadas técnicas de engenharia de atributos, que consistem na criação de atributos e transformação dos já existentes nos espectros. A partir disso, é possível criar um modelo ainda mais robusto extrapolando o seu poder de generalização.

Referências

- AN, F. M. R. **An Introduction to Information Theory**. [S.l.]: Dover Publications, Inc., New York, 1961.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001. Disponível em: <<https://doi.org/10.1023/A:1010933404324>>.
- BREIMAN, L. et al. **Classification and Regression Trees**. [S.l.]: Routledge, 2017. Disponível em: <<https://doi.org/10.1201/9781315139470>>.
- BRO, R.; SMILDE, A. K. Principal component analysis. **Analytical methods**, Royal Society of Chemistry, v. 6, n. 9, p. 2812–2831, 2014. Disponível em: <<https://doi.org/10.1039/C3AY41907J>>.
- CBC. **Folheto de Munições NTA**. 2022. Disponível em: <<https://www.cbc.com.br/arquivos/Folheto-Municoes-NTA.pdf>>.
- CIOCCIA, G. et al. Probabilistic-based identification of gunshot residues (gsr) using laser-induced breakdown spectroscopy (libs) and support vector machine (svm) algorithm. **Microchemical Journal**, v. 207, p. 112142, 2024. ISSN 0026-265X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0026265X24022549>>.
- DALBY, O.; BUTLER, D.; BIRKETT, J. W. Analysis of gunshot residue and associated materials—a review. **Journal of forensic sciences**, Wiley Online Library, v. 55, n. 4, p. 924–943, 2010.
- DHANOA, M. et al. The link between multiplicative scatter correction (msc) and standard normal variate (snv) transformations of nir spectra. **Journal of Near Infrared Spectroscopy**, v. 2, n. 1, p. 43–47, 1994. Disponível em: <<https://doi.org/10.1255/jnirs.30>>.
- DOÑA-FERNÁNDEZ, A. et al. Real-time detection of gsr particles from crime scene: a comparative study of sem/edx and portable libs system. **Forensic science international**, Elsevier, v. 292, p. 167–175, 2018. Disponível em: <<https://doi.org/10.1016/j.forsciint.2018.09.021>>.
- DUARTE, A. et al. Characterization of brazilian ammunitions and their respective gunshot residues with ion beam techniques. **Forensic Chemistry**, Elsevier, v. 7, p. 94–102, 2018.

- FAMBRO, L. A. et al. Laser-induced breakdown spectroscopy for the rapid characterization of lead-free gunshot residues. **Applied Spectroscopy**, SAGE Publications Sage UK: London, England, v. 71, n. 4, p. 699–708, 2017.
- FORTES, F. J. et al. Laser-induced breakdown spectroscopy. **Analytical chemistry**, ACS Publications, v. 85, n. 2, p. 640–669, 2013. Disponível em: <<https://doi.org/10.1021/ac303220r>>.
- GOLDBERG, J. Neighbourhood component analysis. **Advances in Neural Information Processing Systems**, 2005.
- GOLDSTON, R. **Introduction to Plasma Physics**. CRC Press, 2020. Disponível em: <<https://doi.org/10.1201/9780367806958>>. ISBN 9781439822074. Disponível em: <<https://books.google.com.br/books?id=7kM7yEFUGnAC>>.
- GREENACRE, M. et al. Principal component analysis. **Nature Reviews Methods Primers**, Nature Publishing Group UK London, v. 2, n. 1, p. 100, 2022. Disponível em: <<https://doi.org/10.1038/s43586-022-00184-w>>.
- HANDELMAN, G. S. et al. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. **American Journal of Roentgenology**, Am Roentgen Ray Soc, v. 212, n. 1, p. 38–43, 2019. Disponível em: <<https://doi.org/10.2214/AJR.18.20224>>.
- HASTIE, T. et al. **The elements of statistical learning: data mining, inference, and prediction**. [S.l.]: Springer, 2009. v. 2.
- HEALD, M. A.; MARION, J. B. **Classical electromagnetic radiation**. [S.l.]: Courier Corporation, 2012.
- HEARD, B. J. **Handbook of firearms and ballistics: examining and interpreting forensic evidence**. [S.l.]: John Wiley & Sons, 2011.
- LÓPEZ-LÓPEZ, M. et al. An exploratory study of the potential of libs for visualizing gunshot residue patterns. **Forensic science international**, Elsevier, v. 273, p. 124–131, 2017. Disponível em: <<https://doi.org/10.1016/j.forsciint.2017.02.012>>.
- MICHIE, D. et al. **Machine learning, neural and statistical classification**. [S.l.]: Ellis Horwood, 1995.
- Ministério da Justiça e Segurança Pública (Brasil). **Procedimento Operacional Padrão - Perícia Criminal**. 2013. Disponível em: <https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/analise-e-pesquisa/download/pop/procedimento_operacional_padrao_pericia_criminal.pdf>.
- MINZIÈRE, V. R. et al. The relevance of gunshot residues in forensic science. **Wiley Interdisciplinary Reviews: Forensic Science**, Wiley Online Library, v. 5, n. 1, p. e1472, 2023. Disponível em: <<https://doi.org/10.1002/wfs2.1472>>.
- NICOLODELLI, G. et al. Influence of excitation wavelength on libs (1064 nm vs 266 nm) for multi-element mortar analysis. **Microchemical Journal**, v. 207, p. 111894, 2024. ISSN 0026-265X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0026265X2402006X>>.

- PÚBLICA, F. B. D. S. **Anuário Brasileiro de Segurança Pública.** [S.l.]: FBSP, 2024.
- RAI, V.; THAKUR, S. Physics of plasma in laser-induced breakdown spectroscopy. **Laser-induced breakdown spectroscopy**, Elsevier Amsterdam, The Netherlands, p. 83–111, 2007. Disponível em: <<https://doi.org/10.1016/B978-044451734-0.50007-7>>.
- RAINIO, O.; TEUHO, J.; KLÉN, R. Evaluation metrics and statistical tests for machine learning. **Scientific Reports**, Nature Publishing Group UK London, v. 14, n. 1, p. 6086, 2024. Disponível em: <<https://doi.org/10.1038/s41598-024-56706-x>>.
- RASHED, M. Rolling ball algorithm as a multitask filter for terrain conductivity measurements. **Journal of Applied Geophysics**, Elsevier, v. 132, p. 17–24, 2016. Disponível em: <<https://doi.org/10.1016/j.jappgeo.2016.06.016>>.
- RIBEIRO, C. M. et al. Discrimination of maize transgenic and non-transgenic varieties by laser induced spectroscopy (libs) and machine learning algorithms. **Microchemical Journal**, v. 203, p. 110898, 2024. ISSN 0026-265X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0026265X24010105>>.
- ROMOLO, F. S.; MARGOT, P. Identification of gunshot residue: a critical review. **Forensic science international**, Elsevier, v. 119, n. 2, p. 195–211, 2001. Disponível em: <[https://doi.org/10.1016/S0379-0738\(00\)00428-X](https://doi.org/10.1016/S0379-0738(00)00428-X)>.
- RUSSO, R. E. et al. Chapter 3 - laser ablation. In: SINGH, J. P.; THAKUR, S. N. (Ed.). **Laser-Induced Breakdown Spectroscopy (Second Edition)**. Second edition. Amsterdam: Elsevier, 2007. p. 41–70. ISBN 978-0-12-818829-3. Disponível em: <<https://doi.org/10.1016/B978-0-12-818829-3.00003-4>>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780128188293000034>>.
- RYBICKI, G. B.; LIGHTMAN, A. P. **Radiative processes in astrophysics.** [S.l.]: John Wiley & Sons, 1991.
- SEROL, M. et al. Chemical analysis of gunpowder and gunshot residues. **Molecules**, MDPI, v. 28, n. 14, p. 5550, 2023. Disponível em: <<https://doi.org/10.3390/molecules2814550>>.
- SHANNON, C. E. A mathematical theory of communication. **The Bell system technical journal**, Nokia Bell Labs, v. 27, n. 3, p. 379–423, 1948. Disponível em: <<https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>>.
- SHRIVASTAVA, P.; JAIN, V.; NAGPAL, S. Gunshot residue detection technologies—a review. **Egyptian Journal of Forensic Sciences**, Springer, v. 11, p. 1–21, 2021. Disponível em: <<https://doi.org/10.1186/s41935-021-00223-9>>.
- SHVEDCHENKO, D.; SUVOROVA, E. New method of automated statistical analysis of polymer-stabilized metal nanoparticles in electron microscopy images. **Crystallography Reports**, Springer, v. 62, p. 802–808, 2017. Disponível em: <<https://doi.org/10.1134/S1063774517050200>>.
- SINGH, J. P.; THAKUR, S. N. **Laser-induced breakdown spectroscopy.** [S.l.]: Elsevier, 2020.

TAURUS. **Manual 700 - Rev. 01.** 2016. Disponível em: <<https://pt.scribd.com/document/330322416/30004407-manual-700-rev-01>>.

TAYLOR; FRANCIS. **Lasers-Induced Plasmas and Applications.** [s.n.], 1989. (Optical Science and Engineering). ISBN 9780824780784. Disponível em: <<https://books.google.com.br/books?id=Yk8-g6-j5rYC>>.

WARDHANI, N. W. S. et al. Cross-validation metrics for evaluating classification performance on imbalanced data. In: **2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA).** [S.l.: s.n.], 2019. p. 14–18.

ZHANG, X.; LI, P. Lithological mapping from hyperspectral data by improved use of spectral angle mapper. **International Journal of Applied Earth Observation and Geoinformation**, v. 31, p. 95–109, 2014. ISSN 1569-8432. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0303243414000579>>.

ZHU, W. et al. Sensitivity, specificity, accuracy, associated confidence interval and roc analysis with practical sas implementations. **NESUG proceedings: health care and life sciences, Baltimore, Maryland**, v. 19, p. 67, 2010.

Apêndices

Bibliotecas Python

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from skimage import restoration
4 from scipy.signal import savgol_filter
5 from scipy.stats import zscore
6 import leitura_dados as ld
7 from sklearn.decomposition import PCA
8 from sklearn.neighbors import KNeighborsClassifier
9 from sklearn.model_selection import train_test_split
10 from sklearn.metrics import ConfusionMatrixDisplay, confusion_matrix
11 from sklearn.ensemble import RandomForestClassifier
12 from sklearn.metrics import roc_curve, auc, accuracy_score,
13     classification_report
14 from sklearn.preprocessing import label_binarize
15 from sklearn.inspection import permutation_importance
16 import shap
```

Processamento dos Espectros

```
2 def ajustespec(espectro ,ajustebase ,int_analise):
3
4     dadosM = []
5     dadosC = []
6
7
8     for i in range(0,int(espectro.shape[1]/int_analise)):
9         k = i*int_analise + int_analise
10        x = espectro[:, int_analise*i: k ]
11
12        media = np.mean(x, axis = 1, dtype = float)
13
14        background1 = restoration.rolling_ball(media, radius=ajustebase)
15
16        ajuste1 = media - background1
17
18        SNV1 = zscore(ajuste1 , axis = 0)
19
20        media1 = SNV1.reshape(-1, 1)
21
22        dadosM.append(media1)
23
24        Mg = np.concatenate(dadosM, axis=1)
25
```

```

#|||||||||||||||||_
|||||||||||||||||_
28
#OUTLIERS – SMA
30
mediag = np.mean(Mg, axis = 1, dtype = float)
32
for i in range(0,int(Mg.shape[1])):
34
    f = Mg[:, i]
36
    SAM = 0.95
38
    C = np.dot(f, mediag.reshape(-1, 1))/(np.linalg.norm(f)*np.linalg.norm
        (mediag.reshape(-1, 1)))
40
    if C >= SAM:
41
        dadosC.append(f.reshape(-1, 1))
42
44
S_outliers = np.concatenate(dadosC, axis=1)#matriz Mg sem outliers
46
media_outliers = np.mean(S_outliers, axis = 1,dtype = float)
48
return mediag, Mg, S_outliers, media_outliers

```

Rotina de Treino do Modelo *Random Forest*

```

1
X_train, X_test, y_train, y_test = train_test_split( treino3 , classe ,
    test_size=0.30, random_state=92)
3
clf = RandomForestClassifier(n_estimators= 184, criterion = 'entropy',
    max_depth = None,
5
    min_samples_split = 3 , min_samples_leaf = 3, max_features= 'log2',
    max_leaf_nodes= None, bootstrap = False)
7
model_rf = clf.fit(X_train, y_train)# Treinando o modelo
9
y_pred = model_rf.predict(X_test)# Teste do modelo
11
accuracy = accuracy_score(y_pred,y_test)
13
classification_rep = classification_report(y_test, y_pred)
15
print("\nClassification Report:\n", classification_rep)

```

Rotina da redução de dimensão e plot

```
def componentesp(x):
    2    pca = PCA()
    3    scores = pca.fit_transform(x)
    4
    5
    6    proporcao_variancia_explicada = pca.explained_variance_ratio_
    7    autovalores = pca.explained_variance_
    8    componentes_principais = pca.components_
    9
    10   return scores, proporcao_variancia_explicada, autovalores,
    11       componentes_principais
    12
    13 #plot
    14
    15 cp = componentesp(X_test)
    16
    17 for i in range(len(cp[0])):
    18     if y_pred[i] == 0:
    19         plt.scatter(cp[0][i, 0], cp[0][i, 1], c = 'red', s=25)
    20     else:
    21         plt.scatter(cp[0][i, 0], cp[0][i, 1], c = 'green', s=25)
    22
    23
    24 plt.legend()
    25
    26 plt.xlabel(f'Componente Principal 1 ({((cp[1][0])*100).astype(int)}%)')
    27 plt.ylabel(f'Componente Principal 2 ({((cp[1][1])*100).astype(int)}%)')
    28 plt.title('Mapa de Scores PCA')
    29 plt.grid(True)
    30 plt.show()
```