UNIVERSIDADE DE BRASÍLIA

Faculdade de Ciências de Saúde

Programa de Pós-Graduação em Odontologia

Tese de Doutorado

Classificação de ferimentos por armas de fogo com redes neurais convolucionais

Renato Queiroz Nogueira Lira

Brasília, 19 de dezembro de 2024

Renato Queiroz Nogueira Lira

Classificação de ferimentos por armas de fogo com redes neurais convolucionais

Tese apresentada ao Programa de Pós-Graduação em Odontologia da Faculdade de Ciências da Saúde da Universidade de Brasília, como requisito parcial à obtenção do título de Doutor em Odontologia.

Orientador: Prof. Dr. André Ferreira Leite

Coorientadora: Profa. Dra. Andreia Cristina Breda de Souza

Brasília, 2024

Renato Queiroz Nogueira Lira

Classificação de ferimentos por armas de fogo com redes neurais convolucionais

Tese aprovada, como requisito parcial para obtenção do grau de Doutor em Odontologia, Programa de Pós-Graduação em Odontologia da Faculdade de Ciências da Saúde da Universidade de Brasília.

Data da defesa: 13/12/2024

Banca examinadora:

Prof. Dr. André Ferreira Leite (Orientador)

Prof. Dr. Malthus Fonseca Galvão

Prof. Dr. Casimiro Abreu Possante de Almeida

Prof. Dr. Rhonan Ferreira da Silva

Prof. Dr. Fábio Wildson Gurgel Costa (Suplente)

Dedico este trabalho a minha esposa Lorena. Sem ela, minhas conquistas seriam escassas e limitadas. E aos meus pais, Helena e José Roberto (in memorian), por toda paciência em me educar e incentivo para buscar sempre destinos cada vez mais altos.

Agradecimentos

A Deus, por toda proteção e por me guiar todos os dias, me dando forças e inspirações para continuar com os meus planos.

À minha esposa Lorena, sempre presente em todas as etapas deste processo, incentivando-me a crescer continuamente e a focar em nossos projetos.

Aos meus pais, Helena Márcia e José Roberto (*in memorian*), por tudo que representam para mim e por todos os exemplos perfeitos de vida. Se estou aqui, é por causa deles.

Às minhas irmãs Ariane e Isabela. Mesmo morando em locais diferentes, estão no meu coração. Dizem que o irmão do meio é sempre o esquecido pelos demais...

À minha tia Tereza que sempre foi uma segunda mãe.

Ao meu avô José Alves que é um exemplo de pessoa que sempre buscou o conhecimento.

Ao meu orientador André, que bravamente assumiu a tarefa de me aconselhar, orientar e guiar durante a síntese deste trabalho. Obrigado por confiar no projeto e, acima de tudo, em mim.

À minha coorientadora Andreia, que desde 2018, juntamente com o ilustre professor Casimiro, me ajuda na criação e andamento desse trabalho e na admiração pela Odontologia Legal.

Ao prof. Pedro Garcia pela valiosíssima contribuição na condução e realização do trabalho.

Aos meus estagiários, Luana, Maísa e Renan. Sei que foram muitos dias em frente ao computador vendo imagens nada agradáveis, mas, sem vocês, esse trabalho seria impossível de ser feito.

Ao Bruno Scholles pela contribuição e perseverança em fazer esse trabalho dar certo.

Aos meus amigos, por serem verdadeiros pilares para mim. Irmandade sem laços sanguíneos que são formados durante a vida.

À Polícia Civil do Distrito Federal, em especial, ao Instituto de Criminalística e Instituto Médico Legal, unidades do Departamento de Polícia Técnica, pelos dados fornecidos e apoio para que esse trabalho fosse realizado.

Aos demais membros da banca que se comprometerem em avaliar esse trabalho e que devem ter contribuído com valiosos comentários.

A todos que de forma direta ou indireta contribuíram para o estudo e para a minha vida profissional e pessoal.

“Ao curvar-te com a lâmina rija de teu bisturi sobre o cadáver desconhecido, lembra-te que este corpo nasceu do amor de duas almas; cresceu embalado pela fé e esperança daquela que em seu seio o agasalhou, sorriu e sonhou os mesmos sonhos das crianças e dos jovens; por certo amou e foi amado e sentiu saudades dos outros que partiram, acalentou um amanhã feliz e agora jaz na fria lousa, sem que por ele tivesse derramado uma lágrima sequer, sem que tivesse uma só prece."

Karl Rokitansky

Resumo

No Brasil e no mundo, o uso de armas de fogo representa a maior causa de mortes violentas. Assim, visando a apuração dessas mortes, é imprescindível a compreensão de toda dinâmica relacionada e sobre como o crime ocorreu. Porém, em muitos casos, as equipes periciais, tanto do local de encontro cadavérico quanto do exame necroscópico, não conseguem recuperar elementos que sejam suficientes para conclusões mais robustas, restando somente os vestígios na vítima para serem observados e analisados. Além disso, para que ocorra a correta tipificação da infração penal, é necessário que as análises realizadas pelos peritos indiquem as circunstâncias em que os tiros foram realizados, como quais feridas correspondem a entradas e saídas e as suas distâncias em relação às vítimas. Com o objetivo de colaborar com essas análises foi desenvolvido um estudo que resultou em um artigo. Nesse artigo, com o objetivo de automatizar a classificação de ferimentos por armas de fogo, foram treinadas 59 redes neurais convolucionais para diferenciar ferimentos de entrada e saída e determinar a distâncias de tiro por meio de fotografias e documentos de casos de vítima fatais, examinadas pelas equipes de peritos da Polícia Civil do Distrito Federal entre os anos de 2012 e 2022. Uma base de dados abrangente foi construída com 2.551 imagens, incluindo 1.883 feridas de entrada e 668 de saída. A arquitetura ResNet152 demonstrou desempenho superior tanto na classificação de feridas de entrada e saída quanto na categorização médico-legal da distância de tiro. Para a primeira, alcançou precisão, recall, F1-score e especificidade de até 86,90% e uma AUC de 82,09%. Para a classificação médico-legal da distância de tiro, a ResNet152 mostrou uma precisão de até 92,48%, embora o desequilíbrio de amostras tenha afetado outras métricas como recall e F1-score. Os achados do estudo ressaltam os desafios de padronizar as imagens das feridas devido às diferentes condições de captura, mas refletem as realidades práticas do trabalho forense. Esta pesquisa destaca o potencial significativo do aprendizado profundo de máquinas em aprimorar as práticas de medicina forense, corroborando que a inteligência artificial é uma ferramenta de suporte para complementar a expertise humana nas investigações forenses.

Palavras-chave: Medicina Legal, Aprendizado de Máquina, Ferimentos por Arma de Fogo.

Abstract

In Brazil and around the world, the use of firearms represents the leading cause of violent deaths. Therefore, to investigate these deaths, it is crucial to understand the entire dynamic related to how the crime occurred. However, in many cases, forensic teams, both at the crime scene and during the autopsy, are unable to recover elements sufficient for more robust conclusions, leaving only the traces on the victim to be observed and analyzed. Furthermore, to correctly classify the criminal offense, it is necessary for the analyses performed by the experts to indicate the circumstances under which the shots were fired, such as which wounds correspond to entries and exits and their distances from the victims. With the aim of contributing to these analyses, a study was developed, resulting in an article. In this article, with the goal of automating the classification of gunshot wounds, 59 convolutional neural networks were trained to differentiate entry and exit wounds and to determine shooting distances through photographs and case documents of fatal victims examined by forensic teams from the Civil Police of the Federal District between 2012 and 2022. A comprehensive database was constructed with 2,551 images, including 1,883 entry wounds and 668 exit wounds. The ResNet152 architecture demonstrated superior performance in both entry and exit wound classification and medico-legal shooting distance categorization. For the first, it achieved accuracy, recall, F1-score, and specificity of up to 86.90% and an AUC of 82.09%. For the medico-legal shooting distance classification, the ResNet152 showed an accuracy of up to 92.48%, although sample imbalance affected other metrics such as recall and F1-score. Our findings highlight the challenges of standardizing wound images due to varying capture conditions but reflect the practical realities of forensic work. This research underscores the significant potential of deep learning in enhancing forensic medicine practices, advocating for artificial intelligence as a supportive tool to complement human expertise in forensic investigations.

Keywords: Forensic Medicine, Machine Learning, Gunshot Wounds.

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO, REVISÃO DA LITERATURA E OBJETIVOS 11
1.1 INTRODUÇÃO 11
1.2 REVISÃO DA LITERATURA 13
1.2.1 A Análise dos Ferimentos 13
1.2.2 Diferenciação Entre Ferimento de Entrada e de Saída 13
1.2.3 Categorização de Médico-Legal Distância do Tiro 13
1.2.4 Os Tipos de Modelos de Estudo de Lesões Produzidas por Projéteis 14
1.2.4.1. Cadáveres Humanos 14
1.2.4.2. Cadáveres de Outros Animais 14
1.2.4.3. Crânios Humanos 15
1.2.4.4. Crânios Humanos Artificiais 16
1.2.4.5. Gelatina Balística 16
1.2.4.6. Modelos Computacionais 16
1.2.5 Considerações Sobre os Estudos de Lesões Produzidas por Projéteis 16
1.2.6 Aprendizado de Máquina 17
1.3 OBJETIVOS 19
1.4 REFERÊNCIAS 20
CAPÍTULO 2 - DEEP LEARNING-BASED HUMAN GUNSHOT WOUNDS CLASSIFICATION 27
2.1 INTRODUCTION 28
2.2 MATERIAL AND METHODS 29
2.2.1 Data Acquisition 30
2.2.2 Data Preprocessing 31
2.2.3 Learning-Based Classification 32
2.2.4 Experimental Setup 36
2.2.5 Cross-validation 39
2.3 RESULTS 39
2.3.1 Wound Type Classification 40
2.3.2 Medico-legal Shooting Distance Classification 40
2.4 DISCUSSION 47
2.5 LIMITATIONS 50
2.6 CONCLUSIONS 51
AUTHOR’S CONTRIBUTIONS 51

DECLARATION OF COMPETING INTEREST 52
ACKNOWLEDGEMENTS 52
FUNDING SOURCES 52
DATA AVAILABILITY STATEMENTS 52
2.7 REFERENCES 53
CAPÍTULO 3 – DISCUSSÃO GERAL E CONCLUSÕES DA TESE 63
3.1 DISCUSSÃO GERAL 63
3.2 CONCLUSÕES 67
3.3 REFERÊNCIAS 68
CAPÍTULO 4 - PRESS RELEASE 70
APÊNDICE – ARTIGO PUBLICADO 71
ANEXO A - DOCUMENTO DE APROVAÇÃO PELO COMITÊ DE ÉTICA 87
ANEXO B – CARTA DE ACEITE 88

CAPÍTULO 1 - INTRODUÇÃO, REVISÃO DA LITERATURA E OBJETIVOS

1.1 INTRODUÇÃO

Mundialmente, as armas de fogo correspondem aos instrumentos mais utilizados para a prática de homicídios [1]. O Brasil tem sido anualmente o país com os maiores números absolutos, alcançando mais de 47 mil assassinatos com o emprego de armas de fogo no ano de 2017 [2,3].

Na elucidação desses crimes, as equipes de investigação policiais necessitam levantar o máximo de informações para agilizar e orientar a materialização e a busca dos autores [4]. Em casos que envolvem o uso de armas de fogo, a diferenciação entre ferimentos de entrada e saída do projétil, bem como a distância em que o tiro foi efetuado, são informações que podem ajudar a direcionar as investigações e elucidar como o evento ocorreu [5,6].

Como exemplo, existem os casos de vítimas de bala perdida, em que, na maioria das vezes, trata-se de pessoas inocentes que não estavam ligadas ao evento originário do tiro, fazendo com que as informações que direcionam a investigação sejam escassas [7]. Assim, a não definição do calibre pode acarretar a não identificação da arma utilizada e, em consequência, do autor do tiro, que pode sair impune [8].

Por sua vez, as equipes periciais de local de crime correspondem aos primeiros especialistas e são responsáveis pelas análises preliminares, tanto do local onde ocorreu o crime quanto de todos os vestígios relacionados, como as feridas nos corpos das vítimas fatais [4,9]. Assim, podendo promover a aceleração das diligências investigativas, a estimativa de calibre, sendo realizada durante o exame pericial do local de encontro cadavérico, a partir das feridas observadas, permitiria que a equipe de investigação se focasse na busca de um tipo específico de arma e em provas relacionadas a este instrumento.

Apesar da evidente vantagem, geralmente, no local do crime, não ocorre a troca de informações entre a equipe pericial e a de investigação [4,9]. Infelizmente, o mesmo ocorre entre os peritos de local e os médico-legistas. Uma pesquisa realizada no Estado do Pará mostrou que apenas 56,28% das informações dispostas nos laudos das duas categorias são concordantes [10].

A diferenciação entre ferimentos de entrada e saída do projétil é possível a partir das características apresentadas por essas lesões [11,12]. A distância médico-legal do tiro pode ser indicada a partir da análise dos elementos secundários observados próximos ao ferimento de entrada do projétil na vítima e categorizada em encostado, à curta distância e à distância [11,13,14]. Essa categorização fornece elementos importantes para que as autoridades consigam tipificar de maneira mais assertiva a conduta criminosa [15].

Diante do exposto, observa-se a importância da troca de informações específicas entre os envolvidos na apuração dos fatos, como: o calibre da arma utilizada, a localização do armamento e a possível intenção do autor. Esses dados são cruciais para uma investigação mais precisa e eficaz, permitindo uma compreensão mais clara dos acontecimentos e auxiliando na reconstituição detalhada do crime. O conhecimento gerado a partir desse cruzamento de informações pode ser decisivo para esclarecer a dinâmica do ocorrido e apoiar as conclusões periciais.

Além disso, a inteligência artificial, com sua alta capacidade de padronizar procedimentos e otimizar o tempo de análise, processando grandes volumes de dados de forma ágil e consistente, tem impulsionado consideravelmente o desenvolvimento e a pesquisa em diversos campos do conhecimento, inclusive na área forense [16–21]. Parte dessas iniciativas já começou a ser aplicada na balística forense [22,23].

Assim, utilizando-se das imagens dos ferimentos, por meio de aprendizagem de máquina, buscou-se treinar modelos computacionais que colaborem com a diferenciação entre ferimentos de entrada e saída do projétil e com a estimativa da distância médico-legal do tiro.

1.2 REVISÃO DA LITERATURA

1.2.1 A Análise dos Ferimentos

No campo da Medicina Legal, várias informações podem ser obtidas ao se analisar as lesões resultantes dos eventos traumáticos. Tratando-se de lesões produzidas por projéteis expelidos por arma de fogo, é possível, entre outras análises, diferenciar feridas de entrada e saída e estimar as distâncias em que os tiros foram realizados [24].

1.2.2 Diferenciação Entre Ferimento de Entrada e de Saída

O ato de determinar se um ferimento produzido por um projétil expelido por arma de fogo consiste em uma entrada ou saída no corpo da vítima é realizado pela análise das características apresentadas pelos próprios ferimentos. Os ferimentos de entrada costumam ter formato regular, variando do elíptico ao arredondado, bordas invertidas, podendo apresentar orlas de enxugo, e são acompanhados, em parte, pelos efeitos produzidos pelos elementos secundários dos tiros (zonas de tatuagem, de esfumaçamento, de chamuscamento) [13,14,25].

Em contrapartida, as lesões de saída geralmente apresentam formato irregular, dimensões maiores que as de entrada, bordas evertidas e ausência dos efeitos secundários dos tiros [13,14,25].

1.2.3 Categorização de Médico-Legal Distância do Tiros

A categorização de distância médico-legal do tiro se dá de maneira visual, por meio da análise dos ferimentos de entrada e, mais especificamente, dos elementos secundários presentes em suas adjacências [13,14,25].

Esses elementos constituem basicamente as zonas de chamuscamento, de esfumaçamento e de tatuagem. A partir da presença ou ausência desses elementos, os peritos indicam a distância médico-legal do tiro [11,26].

Quando nenhum elemento secundário é visualizado ao redor de um ferimento de entrada, o tiro é considerado como sendo realizado à distância. Ao serem visualizados quaisquer um dos efeitos secundários, isoladamente ou em conjunto, o tiro é considerado a curta distância. Por fim, o tiro encostado provoca uma lesão que leva os elementos primários e secundários de uma vez para o interior do corpo, provocando uma explosão de dentro para fora, deixando características únicas à lesão [11,27].

1.2.4 Os Tipos de Modelos de Estudo de Lesões Produzidas por Projéteis

Os trabalhos sobre lesões produzidas por projéteis podem ser divididos de acordo com os modelos de estudo utilizados: aqueles que utilizam cadáveres ou ossadas humanas e de animais; os que empregam modelos artificiais, tecidos sintéticos, gelatina balística e modelos computacionais [28,29].

1.2.4.1. Cadáveres Humanos

Phelps [30] fez um estudo com centenas de cadáveres, utilizando os calibres .22, .32, .38 e .44 em diversas distâncias de tiro. A partir dos resultados obtidos concluíram que, para todos os calibres abordados, na pele, o ferimento de entrada possuía dimensões menores do que as dos projéteis que os originaram, exceto para os casos de tiro encostado e os que impactaram em alguma curvatura craniana.

1.2.4.2. Cadáveres de Outros Animais

Rainio *et al*. [31], em uma amostra constituída por porcos anestesiados, efetuaram tiros com fuzis AK-47, variando a angulação e distância, para observar as diferenças morfológicas na pele durante o período do processo de putrefação. As conclusões do experimento foram que: as deduções de uso de diferentes armas e projéteis com base apenas nos ferimentos são complexas e as alterações post-mortem dificultaram o reconhecimento entre lesões de entrada e de saída.

Ao analisar ferimentos de entrada e saída em carcaças de porcos produzidos em diferentes distâncias de tiro, Wong *et al*. [32] concluíram que as características de ambas eram diferentes de acordo com a distância de produção. O mesmo estudo ainda sugeriu uma fórmula para calcular a distância do tiro a partir das dimensões do ferimento de entrada. Porém, dada a pequena amostra utilizada no trabalho, tal fórmula apresentou um coeficiente de determinação (*R square*) de apenas 17,1%.

O estudo de PIRCHER *et al*. [33] realizou tiros de projéteis de calibre .38, com pontas de diferentes formatos, contra segmentos da região abdominal de porcos com o intuito de avaliar sua influência no tamanho da orla de escoriação e dos ferimentos de entrada. Concluíram que há uma significativa correlação entre o formato do projétil e o tamanho da orla de escoriação. Porém, deve-se atentar para o fato de que modelos de estudo que utilizam cadáveres humanos e de animais possuem limitação devido à baixa reprodutibilidade, além dos problemas éticos envolvidos [34,35]. Além disso, Maiden & Byard [36] referiram que, em cadáveres, a força de tensão da pele pode variar de acordo com o intervalo entre a morte e o experimento, à medida que os efeitos da putrefação e do processo autolítico começam a se manifestar.

Já no caso de animais, a pele de cada espécie possui um direcionamento definido em relação à elasticidade e extensão, esticamento, como documentado pelas linhas de tensão mínima, conhecidas como Linhas de Langer. As Linhas de Langer representam as direções das fibras de colágeno na pele, influenciando a forma como esta se estica e responde a traumas ou incisões. As propriedades biomecânicas da pele, determinadas por essas linhas, influenciam diretamente sua resposta à expansão. Essas propriedades são os determinantes mais importantes de um modelo de estudo animal para a expansão de tecido humano [37].

1.2.4.3. Crânios Humanos

No trabalho de Phelps [30], foram realizados experimentos em 308 crânios atingidos por projéteis de calibres .22, .32, .38 e .44. Foi observado que mais de um terço das perfurações apresentavam formato circular e constatou-se que as demais morfologias não afetavam a média dos diâmetros. Por fim, o autor estabeleceu que as perfurações são ligeiramente maiores que os calibres dos projéteis que as produziram e que a distância não é um fator que afeta as dimensões e forma das lesões de entrada.

Os estudos mais recentes que avaliaram as lesões de entrada em crânios secos utilizaram amostras relativamente pequenas, variando, respectivamente, em 21, 73 e 35 espécimes [38–40] Suas conclusões foram que as perfurações de entrada possuíam diâmetros maiores do que os projéteis que as produziram [38,39] e que as perfurações de saída em crânios secos podiam não ser similares às de corpos íntegros [40]. Além disso, nas pesquisas não foi evidenciada diferença estatística entre as lesões produzidas por .22 e .25; entretanto, entre esses e o calibre .38 foi encontrada diferença [38,39]. Berryman, Smith & Symes [38] destacaram que, quando não há outros fatores envolvidos, uma lesão produzida por um calibre .22 talvez possa ser diferenciada de um .32 e definitivamente de um .45.

1.2.4.4. Crânios Humanos Artificiais

Mahoney *et al*. [41] utilizaram seis crânios fabricados artificialmente através de impressão 3D, obtidos a partir de um mapeamento interno e externo de um crânio humano e os cobriram com uma camada de pele sintética polimérica. O modelo artificial apresentou algumas características semelhantes às esperadas de ferimentos reais, como os aspectos dos ferimentos de entrada e saída e dos padrões de fratura macroscópicos. Porém, outros elementos, como as dimensões dos ferimentos artificiais e da perfuração no modelo ósseo, precisam de aprimoramento.

1.2.4.5. Gelatina Balística

Park *et al*. [42] testaram projéteis com pontas de angulações distintas, mesma massa e diâmetro, e com diferentes diâmetros em gelatina balística. Foi detectado que a diminuição do diâmetro do projétil aumenta sua penetração, devido à menor resistência. Constatou-se que o formato e o diâmetro dos projéteis são fatores importantes na produção de ferimentos.

1.2.4.6. Modelos Computacionais

Os estudos que avaliaram o diâmetro das perfurações com o uso de modelos computacionais do tipo elementos finitos resultaram em perfurações bem maiores do que as dos trabalhos em crânios secos [29,43].

1.2.5 Considerações Sobre os Estudos de Lesões Produzidas por Projéteis

Aplicando os diversos tipos de modelo de estudo, vários autores acreditam que o calibre do projétil não pode ser determinado pelo diâmetro da ferida de entrada na pele [25,27,44]. Enquanto em osso essa relação é mais tangível, apesar de necessitar bastante cuidado [30,38,39].

Feridas em áreas flácidas diferem daquelas que impactam em áreas mais rígidas, como quando o osso está próximo à superfície, como na cabeça [25,41]. Os mecanismos de produção dos ferimentos são determinados não só pelas características do projétil, como formato, construção, massa e velocidade, mas também pelas propriedades biomecânicas dos tecidos, como elasticidade e densidade [28].

A explicação para que o ferimento seja menor do que o calibre é devido à contração elástica sofrida pela pele imediatamente após a entrada, resultando em uma perfuração menor do que o calibre do projétil [25,44–46]. As propriedades físicas da pele não são unidimensionais e, especificamente, as elásticas mudam quanto maior for o estresse aplicado [37]. Além disso, destaca-se que a idade também afeta as propriedades físicas dos tecidos, como a redução da elasticidade com o aumento da idade [36].

Outro fator importante é a velocidade com que o projétil atinge. Para perfurar a pele necessita-se de projéteis com velocidades de 58 m/s e 75 m/s, respectivamente, para os calibres .38 e .22, enquanto para transpassar material ósseo, são necessários projéteis com velocidade entre 61 e 84 m/s [27]. Contudo, há necessidade de atenção ao fato de projéteis com velocidades menores resultarem em entradas maiores [45].

Além dos fatores descritos, ainda há os fenômenos de *yawing*[1] de *tumbling*[2] e de deformação durante o impacto que podem influenciar na morfologia das feridas [40].

1.2.6 Aprendizado de Máquina

O uso de inteligência artificial visa simular as funções cognitivas humanas. O *machine learning* (aprendizado de máquina) é uma forma de treinar a máquina a atingir essas simulações, por sua vez, *deep learning* é uma subcategoria bastante voltada para reconhecimento de padrões, textuais e de imagens, como exemplos [47,48].

Posto isso, o uso de inteligência artificial com aprendizado de máquina já é apontado como uma ferramenta que pode contribuir substancialmente em questões forenses [21]. Nesse cenário, para a estimativa de distância dos tiros, o aprendizado de máquina já foi testado em fotografias de lesões produzidas em carcaças de porcos [23], obtendo acurácia de até 98% a partir dos modelos computacionais obtidos. Entretanto, trabalhos em imagens de ferimentos em humanos ainda são um campo a ser explorado.

---

[1] É o desvio no eixo longitudinal do projétil em relação a um eixo de rotação vertical estabelecido pelo centro de gravidade do referido projétil. Quanto maior a velocidade do projétil, menor o grau de "guinada" ao atingir o objetivo (FIGUEROA, J. R. M. T.; MOLINA, M. G. G.; VELAZCO, F. A. Balística: Balística de efectos o balística de las heridas. Cirujano General, v. 23, n. 4, p. 266-72, 2001). [54]

[2] É o efeito de rotação impresso no projétil, que gira a ponta para trás e a extremidade traseira para a frente, para girar em torno do centro da massa do projétil (SAN ROMÁN, E., NEIRA, J., TISMINETZKY, G. Trauma - Prioridades. 2002).[55]

O estudo de Cheng *et al*. [22] testou diversas arquiteturas de *deep learning* com o intuito de diferenciar ferimentos de entrada dos de saída, obtendo 87,99% de acurácia.

1.3 OBJETIVOS

O objetivo geral deste trabalho foi avaliar modelos de aprendizado de máquina para analisar fotografias de ferimentos produzidos por projéteis expelidos por arma de fogo para extrair informações importantes à investigação forense.

O trabalho teve como objetivos específicos:

- Analisar modelos de aprendizado de máquina para diferenciar ferimentos de entrada e de saída produzidos por projéteis a partir de fotografias.

- Qualificar modelos de aprendizado de máquina para estimar a distância médico-legal de realização do tiro a partir da análise de fotografias de ferimentos de entrada do projétil.

As metodologias utilizadas para responder a esses objetivos específicos estão detalhadas no próximo capítulo desta tese, que consiste no artigo submetido e aprovado para publicação.

1.4 REFERÊNCIAS

1. UNODC. Global study on homicide 2023. 2023; Available from: https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html

2. Wisevoter. Gun Deaths by Country. 2024;1000. Available from: https://wisevoter.com/country-rankings/gun-deaths-by-country/

3. Cerqueira D, Bueno S, Lima RS de, Alves PP, Marques D, Lins G de OA. Atlas da Violencia 2023. 2023;1–121. Available from: https://www.ipea.gov.br/atlasviolencia/arquivos/artigos/9350-223443riatlasdaviolencia2023-final.pdf

4. Mingardi G. A Investigação de Homicídios - Construção de Um Modelo. 1º Concurs Nac Pesqui Apl Em Segurança Pública E Justiça Crim [Internet]. 2005;104. Available from: http://www.periciaoficial.al.gov.br/legislacao/artigos-pesquisas-cientificas-e-livros/invest_homic_guaracy.pdf

5. Kleypas DA, Badiye A. Evidence Collection. StatPearls. 2024.

6. Shrestha R, Kanchan T, Krishan K. Gunshot Wounds Forensic Pathology. StatPearls. 2024.

7. UNLIREC. Balas Perdidas II: Análisis de Casos de Balas Perdidas Reportados en Medios de Comunicación en América Latina y el Caribe (2014-2015). 2016.

8. Debeti LFF. A importância do exame pericial de comparação balística nas decisões dos julgadores em infrações penais com emprego de arma de fogo [Internet]. 2015. Available from: https://repositorio.ufsc.br/bitstream/handle/123456789/133842/TCC

Final.pdf?sequence=1

9. Oliveira JLM de. Perícia e Investigação Criminal: uma proposta de melhoria do modelo organizacional visando a otimização de resultados [Internet]. Fundação Getúlio Vargas; 2013. Available from: http://bibliotecadigital.fgv.br/dspace/handle/10438/11868

10. Rebello FBB. Perinecroscopia: conflitos entre peritos criminais e médicos-legistas [Internet]. 2002. p. 6. Available from: http://geocities.yahoo.com.br/filomenarebello/perinecrosco.htm

11. França. Medicina Legal. 11th ed. Med. Leg. Guanarabara Koogan; 2017.

12. Denton JS, Segovia A, Filkins JA. Practical pathology of gunshot wounds. Arch Pathol Lab Med. 2006;130:1283–9.

13. Zumwalt RE, Hirsch CS. Medicolegal Interpretation. of Gunshot Wounds. 1986.

14. Adelson L. A microscopic study of dermal gunshot wounds. Am J Clin Pathol. 1961;35:393–402.

15. Luz JWP. Da perícia em crimes contra a vida: parâmetros de atuação pelo Ministério Público. 2016;1–20.

16. Vodanović M, Subašić M, Milošević D, Galić I, Brkić H. Artificial intelligence in forensic medicine and forensic dentistry. J Forensic Odontostomatol. 2023;41:30–41.

17. Tournois L, Trousset V, Hatsch D, Delabarde T, Ludes B, Lefèvre T. Artificial intelligence in the practice of forensic medicine: a scoping review. Int J Legal Med. 2023;

18. Lefèvre T, Tournois L. Artificial Intelligence and Diagnostics in Medicine and Forensic Science. Diagnostics. 2023;13:1–12.

19. Wankhade TD, Ingale SW, Mohite PM, Bankar NJ. Artificial Intelligence in Forensic Medicine and Toxicology: The Future of Forensic Medicine. Cureus. 2022;14.

20. Galante N, Cotroneo R, Furci D, Lodetti G, Casali MB. Applications of artificial intelligence in forensic sciences: Current potential benefits, limitations and perspectives. Int J Legal Med. 2022;445–58.

21. Margagliotti G, Bollé T. Machine learning & forensic science. Forensic Sci Int. 2019;298:138–9.

22. Cheng J, Schmidt C, Wilson A, Wang Z, Hao W, Pantanowitz J, et al. Artificial intelligence for human gunshot wound classification. J Pathol Inform. 2024;15:0–3.

23. Junno JA, Kotiaho A, Oura P. Post-mortem computed tomography in forensic shooting distance estimation: a porcine cadaver study. BMC Res Notes. 2022;15:1–6.

24. Di Maio VJM. Gunshot Wounds: Practical Aspects of Firearms, Ballistics, and Forensic Techniques. 3rd ed. CRC Press; 2015.

25. Saukko P, Knight B. Knight's Forensic Pathology. 4th ed. CRC Press; 2016.

26. Zumwalt RE, Hirsch CS. Medicolegal interpretation of gunshot wounds. Am J Emerg Med. 1987;5:133–9.

27. Di Maio VJM, Dana SE. Handbook of Forensic Pathology. 2nd ed. Boca Raton:

CRC Press; 2006.

28. Humphrey C, Kumaratilake J. Ballistics and anatomical modelling – A review. Leg Med [Internet]. 2016;23:21–9. Available from: http://dx.doi.org/10.1016/j.legalmed.2016.09.002

29. Rodrigues LL. Morphological and biomechanical differences caused by gunshot wounds on the parietal bone using a computational simulation Diferenças morfológicas e biomecânicas causadas pelo ferimento por arma de fogo em osso parietal utilizando simulação computacional [Internet]. Universidade Estadual de Campinas; 2017. Available from: http://repositorio.unicamp.br/bitstream/REPOSIP/330558/1/Rodrigues_LarissaLopes_M.pdf

30. Phelps C. Traumatic injuries of the brain and its membranes : with a special study of pistol-shot wounds of the head in their medico-legal and surgical relations. London: Henry Kimpton; 1898.

31. Rainio J, Lalu K, Ranta H, Penttilä A. Morphology of experimental assault rifle skin wounds. Int J Legal Med [Internet]. 2003;117:19–26. Available from: http://link.springer.com/10.1007/s00414-002-0308-9

32. Wong YS, Khairul O, Atiah AAG, Abdul Aziz I, Muhd. Hilmi B, Noor Hazfalinda H SPAA. Porcine model: Differences in entry and exit wounds by semi-automatic pistol at different shooting distances. Malaysian J Forensic Sci [Internet]. 2013;4:7–14. Available from: https://www.researchgate.net/publication/260060250

33. Pircher R, Preiß D, Pollak S, Thierauf-Emberger A, Perdekamp MG, Geisenberger D. The influence of the bullet shape on the width of abrasion collars

and the size of gunshot entrance holes. Int J Legal Med. 2017;131:441–5.

34. Zhen T, Zhonghua Z, Gang Z, Yubin C, Tao L, Yinghui T. Establishment of a three-dimensional finite element model for gunshot wounds to the human mandible. J Med Coll PLA. 2012;27:87–100.

35. Tang Z, Tu W, Zhang G, Chen Y, Lei T, Tan Y. Dynamic simulation and preliminary finite element analysis of gunshot wounds to the human mandible. Injury [Internet]. 2012;43:660–5. Available from: http://dx.doi.org/10.1016/j.injury.2011.03.012

36. Maiden NR, Byard RW. Unpredictable tensile strength biomechanics may limit thawed cadaver use for simulant research. Aust J Forensic Sci. 2016;48:54–8.

37. Bartell TH, Mustoe TA. Animal models of human tissue expansion. Plast Reconstr Surg. 1989;83:681–6.

38. Berryman HE, Smith OC, Symes SA. Diameter of cranial gunshot wounds as a function of bullet caliber. J Forensic Sci. 1995;40:751–4.

39. Ross AH. Caliber estimation from cranial entrance defect measurements. J Forensic Sci. 1996;41:629–33.

40. Quatrehomme G, Işcan MY. Characteristics of gunshot wounds in the skull. J Forensic Sci. 1999;44:568–76.

41. Mahoney P, Carr D, Arm R, Gibb I, Hunt N, Delaney RJ. Ballistic impacts on an anatomically correct synthetic skull with a surrogate skin/soft tissue layer. Int J Legal Med. 2017;1–12.

42. Park NR, Kim KH, Mo JS, Yoon GH. An experimental study on the effects of the head angle and bullet diameter on the penetration of a gelatin block. Int J Impact Eng. 2017;106:73–85.

43. Lopez YAQ. Simulação dinâmica de projéteis disparados contra a região occipital do crânio humano – análise de elemento finitos Dynamic simulation of bullets fired at the occipital region of the human skull : finite element analysis [Internet]. Universidade Estadual de Campinas; 2016. Available from: http://repositorio.unicamp.br/bitstream/REPOSIP/304960/1/LopezQuintero_YuliAndrea_M.pdf

44. von See C, Stuehmer A, Gellrich N-C, Blum KS, Bormann K-H, Rücker M. Wound ballistics of injuries caused by handguns with different types of projectiles. Mil Med. 2009;174:757–61.

45. Fisher BAJ, Fisher DR. Techniques of Crime Scene Investigation. 8th ed. CRC Press; 2012.

46. Kneubühl BP, Coupland RM, Rothschild MA, Thali MJ. Wound Ballistics: Basics and applications. Int. J. Legal Med. Springer-Verlag; 2012.

47. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat. Rev. Cancer. Nature Publishing Group; 2018. p. 500–10.

48. Russell SJ (Stuart J, Norvig P. Artificial intelligence : a modern approach. Prentice Hall; 1995.

49. Oura P, Junno A, Junno JA. Deep learning in forensic gunshot wound

interpretation—a proof-of-concept study. Int J Legal Med. 2021;135.

50. Piraianu AI, Fulga A, Musat CL, Ciobotaru OR, Poalelungi DG, Stamate E, et al. Enhancing the Evidence with Algorithms: How Artificial Intelligence Is Transforming Forensic Medicine. Diagnostics. 2023;13.

51. Smith JC. Investigation of Fatal Gunshot Wounds. Cleveland-Marshall Law Rev [Internet]. 1964;13:296–308. Available from: https://engagedscholarship.csuohio.edu/clevstlrev

52. Wendel ASO. Techniques of Crime Scene Investigation. 2nd ed. Elsevier; 1965.

53. Ridge MD, Wright V. The directional effects of skin. A bio-engineering study of skin with particular reference to Langer’s lines. J Invest Dermatol [Internet]. 1966;46:341–6. Available from: http://dx.doi.org/10.1038/jid.1966.54

54. Figueroa JRM-T, Molina MGG, Velazco FA. Balística : Balística de efectos o balística de las heridas. Cir Gen. 2001;23:266–72.

55. San Román E, Neira JA, Tisminetzky G. Trauma Prioridades. 1st ed. Soc. argentina Med. y cirugía del trauma. Buenos Aires: Editorial Médica Panamericana; 2002.

CAPÍTULO 2 - DEEP LEARNING-BASED HUMAN GUNSHOT WOUNDS CLASSIFICATION

Abstract

In this paper, we present a forensic perspective on classifying gunshot wound patterns using Deep Learning (DL). Although DL has revolutionized various medical specialties, such as automating tasks like medical image classification, its applications in forensic contexts have been limited despite the inherently visual nature of the field. This study investigates the application of DL techniques (59 architectures) to classify gunshot wounds in a forensic context, focusing on distinguishing between entry and exit wounds and determining the Medical-Legal Shooting Distance (MLSD), which classifies wounds as contact, close range, or distant, based on digital images from real crime scene cases. A comprehensive database was constructed with 2,551 images, including 1,883 entries and 668 exit wounds. The ResNet152 architecture demonstrated superior performance in both entry and exit wound classification and MLSD categorization. For the first task, achieved accuracy of 86.90% and an AUC of 82.09%. For MLSD, the ResNet152 showed an accuracy of 92.48% and AUC up to 94.36%, though sample imbalance affected the metrics. Our findings underscore the challenges of standardizing wound images due to varying capture conditions but reflect the practical realities of forensic work. This research highlights the significant potential of DL in enhancing forensic pathology practices, advocating for Artificial Intelligence (AI) as a supportive tool to complement human expertise in forensic investigations.

Keywords: Forensic medicine, Gunshot wounds, Artificial intelligence, Deep learning

2.1 INTRODUCTION

Firearms correspond to one of the most used instruments worldwide for committing homicides, being related to 75% of cases in the Americas in 2021 [1]. In Brazil, the country with the highest number of firearm-related fatalities, there are more than 47,000 deaths attributed to firearms annually [2]. Given these high rates, it is common in forensic fieldwork to encounter gunshot wounds during external body examinations, whether at the crime scene or during an autopsy. When these are found, among various procedures, it is necessary to distinguish between entry and exit wounds and to classify the entry wounds according to the distance from which the shot was fired [3, 4].

Some characteristics, such as shape, size, and margins, allow for relatively easy differentiation between entry and exit wounds [5]. However, these characteristics are not always clearly visible, or they may be partially absent or even overlapping, complicating the accurate classification. Additionally, other factors can influence wound production, including the caliber and type of ammunition, the angle and speed of the shot, barriers encountered by the projectile, and the region of the body impacted [6, 7].

Similarly, the classification of the shot distance can be influenced by the same aforementioned factors [6]. This distance has a medico-legal classification, which is based on the characteristics presented by the injury itself and its naked eye appearances. These categories include: distant (or medium-distance) wound, close (or near) range wound and contact wound [7, 8]. Some authors also introduce additional categories between close-range and contact, such as near-contact and loose-contact [6]. Moreover, it is important to emphasize that this is a medico-legal classification based on the traces found on the victim, not necessarily reflecting the actual event, and serves as an indication or suggestion of what occurred at the crime scene [7].

Regardless, for forensic investigation purposes, any additional information can be crucial in aiding the elucidation of what happened, how it happened, and who was involved. However, to ensure impartiality and justice, it is important that the evidence produced is based on and grounded in scientific criteria [9].

In turn, in recent years, there has been an increase in the research and perspectives of use of AI with Machine Learning (ML) across various sectors, including forensic medicine [10–15]. In the pursuit of faster and more robust crime resolution, image recognition and classification have seen significant advancements, including initiating studies on the classification of firearm injuries [16, 17]. Despite these developments contributing to the enhancement of forensic methodologies and the efficiency of criminal investigations, and being the studies that initiated research in the area, they involved the use of only a few images and were based on injuries in pig carcasses [17] or, when trained with autopsy photographs of wounds on human bodies, did not analyze the MLSD aspect [16].

The current study aimed to develop a ML model that can analyze and classify photographs of real crime scene forensic cases of gunshot wounds in relation to determining whether they are entry or exit wounds and to classify MLSD of the gunshot wounds.

2.2 MATERIAL AND METHODS

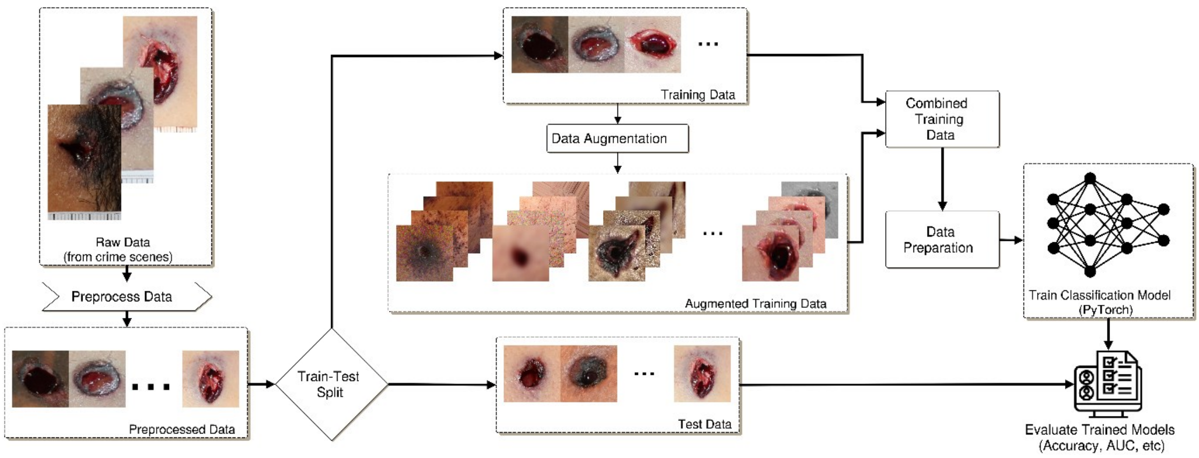


Figure 1: Developed framework for data augmentation, classification modeling, and used to evaluate the proposed methodology.

Figure 1 illustrated the framework employed during the development of this work. It includes three main stages responsible for preprocessing the captured photographies from the crime scene, classifying the preprocessed images using a deep neural network, and finally estimating the classification according to a specific forensic task. The following subsections details each stage.

2.2.1 Data Acquisition

In this study, photographic records of homicide and suicide cases examined by crime scene forensic experts and reports by medical examiners from incidents attended by the Federal District Civil Police (PCDF) (Brasilia, Federal District, Brazil) from the January of 2012 until March of 2022 were used. From the selected time period, a total of 2,012,584 photographs were stored. This dataset was obtained with the proper authorization of PCDF and the Committee for Ethics in Research of the University of Brasília approved this research (Protocol number 54418221.9.0000.0030).

The images and relevant data related to the wounds, such as age, sex, anatomical location, distance, and wound category, the following inclusion criteria were established:

1. the victim was examined and photographed in the crime scene by forensic experts;
2. the body did not show advanced signs of decomposition (e.g., greenish discoloration of the skin of the anterior abdominal wall in the right iliac fossa region);
3. the photos from the scene examination were in focus and with sufficient visualization of the wounds;
4. the victim was identified, due to the need for additional data - sex and age at the time of death;
5. there were bullets stored in the bodies;
6. the bullets had been recovered during the autopsy at the Legal Medical Institute (IML/PCDF);
7. the bullets recovered at the IML had been sent to the Institute of Criminalistics (IC/PCDF);
8. it was possible to determine the caliber(s) of the bullet(s) at the Forensic Ballistics Section (SBF/IC/PCDF);
9. only one caliber involved per victim was determined at SBF;
10. the respective autopsy reports, made by experienced forensics pathologists, indicates which anatomical location, shot distance (distant, close or contact) and

category (entry or exit) the photographed wound corresponds to. These reports correspond to the gold standard.

These criteria were designed based on the working protocol guidelines of the PCDF and, due to the use of this database, several wounds and cases were excluded from this study, reducing the available sample. However, these data may be used for further analyses, potentially involving caliber as one of the factors to be studied.

2.2.2 Data Preprocessing

The images retrieved from these cases were then manually cropped to isolate the wounds in a square format. This manual processing was handled using the GNU Image Manipulation Program (GIMP) image processing software, resulting in a preprocessed dataset of 2,551 images of gunshot wounds. These preprocessed images were then categorized according to wound type and MLSD. The wound types were categorized into entry and exit wounds, corresponding to 1,883 entry and 668 exit wounds. Table 1 details the wound types according to the number of cases, associated images, and sex. Moreover, considering only the subset of entry wound images, we obtained 1,609 from distance shots, 232 from close range, and 42 from contact shots as per the MLSD categorization, as depicted in Table 2.

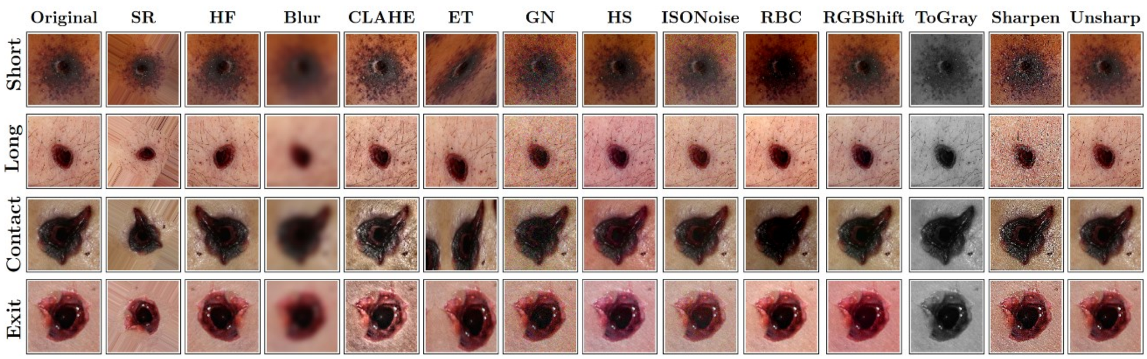


Figure 2: Examples of gunshot wounds images and some their augmented versions. First line corresponds to a close range entry, second line corresponds to distance shot entry, third line presents the contact shot entry image, and the last line shows an example of exit gunshot wound. The results obtained from the used augmentation are depicted from left to right: the original (preprocessed) image, SR, HF, blur, CLAHE, ET, GN, HS, ISO noise, RBC, gray representation of image (ToGray), sharpen, and unsharp mask.

Table 1 – Cases and images per sex and wound type (entry or exit).

Sex	# cases (%)	# images (%)	# entries (%)	# exits (%)
Female	24 (4.91)	114 (4.47)	85 (4.51)	29 (4.34)
Male	165 (95.09)	2437(95.53)	1798(95.49)	639 (95.66)
Total	489 (100)	2551 (100)	1883 (100)	668 (100)

Table 2 – Entry wounds images per medico-legal distance classification.

Distance	# images (%)
Distant	1609 (85.45)
Close range	232 (12.32)
Contact	42 (2.23)
Total	1883 (100)

Table 3 – Number of wound images per position on the body.

Anatomical location	# images (%)
Head and neck	548 (21.48)
Trunk	1310 (51.35)
Upper limbs	503 (19.72)
Lower limbs	190 (7.45)
Total	2551 (100)

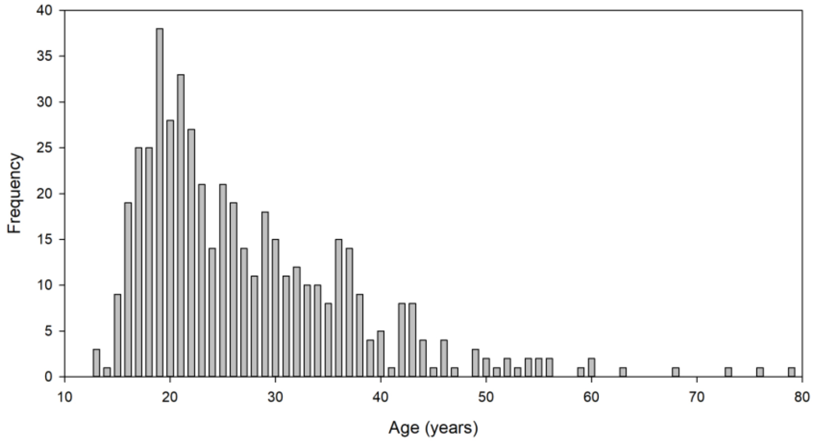
2.2.3 Learning-Based Classification

Considering the curated data as described in the previous section, we identified that the recognition of forensic characteristics of gunshot wound images can be expressed as an image classification problem. More specifically, the identification of the wound type is a binary classification problem (i.e., entry or exit), while the categorization of the MLSD is a multiclass classification problem, as can be inferred from Table 2. In other words, both are Computer Vision (CV) problems. Therefore, this study tested the potential of a series of DL models established in the literature for CV problems.

The neural networks used in this study were achieved from TorchVision [19], popular library that is part of the PyTorch [20] ecosystem. TorchVision provides tools and utilities specifically designed for computer vision tasks, such as state-of-the-art architectures and its pretrained models. From this library, fifty-nine state-of-the-art DL architectures were utilized to analyze forensic images for the differentiation of entry and exit wounds as well as the determination of MLSD. These models are listed in Tables 4 and 5. Additionally, we employed a supplementary abstraction software layer using Skorch [21], a library that provides a high-level interface for working with PyTorch models in a scikit-learn-like way. This additional layer was included to add Scikit-Learn [22] compatibility and make it easy to use PyTorch models with the scikit-learn ecosystem. This allows our software implementation to leverage tools and utilities from Scikit-Learn, such as grid search, cross-validation, hyperparameter tuning, and pipelines, with PyTorch models. Thus, with the use of Skorch, our implementation benefits from several desirable properties in terms of software engineering, such as decoupling, maintainability, efficiency, correctness, reusability, testability, etc. Our concern on these properties was to ensure reliability in conducting our experiments.

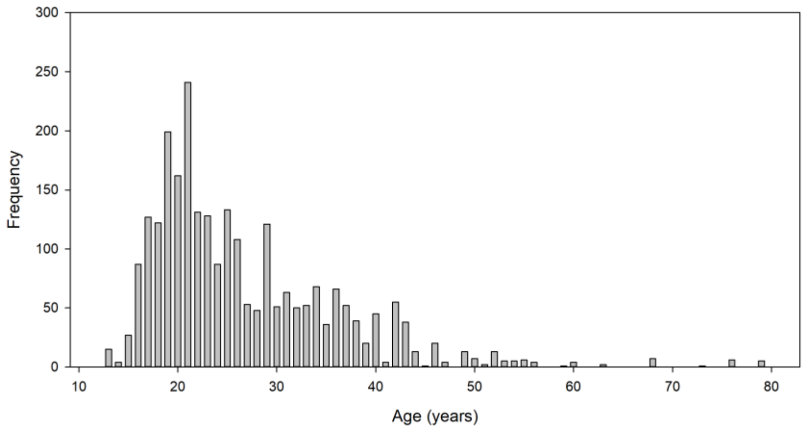
As can be noticed from Tables 1, 2 and 3, the studied dataset is highly imbalanced. This imbalance is usually common in forensic datasets, where certain types of wounds or shooting distances may be underrepresented compared to others. The data imbalance emerges from the real-world crime cases, where some type of wounds can often be limited in samples due to the sensitivity and rarity of such data. Taking that into account, we implemented some techniques in the data preprocessing when creating the training pipeline for the investigated models.

The main technique adopted in the scope of training data preparation was the data augmentation. The term “data augmentation” allude to procedures for building iterative sampling and optimization algorithms via the incorporation of unobserved data or latent variables. Its main goal is to increase the volume, quality and diversity of training data. In terms of the distribution of the target dataset, the greater the size, diversity, and representativeness of the training data, the more effectively the DL model performs on unseen data. It introduces variability into the training data, which helps prevent the model from memorizing specific features of the training set and overfitting. Augmentation techniques such as rotation, flipping, and changes in brightness or contrast help the model become more robust to these variations, ensuring accurate identification across different scenarios. On one hand, the variety of training samples should be ample enough for the model to manage various deviations in image appearance and even noisy target instances. On the other hand, this encourages the model to learn more generalizable features that are applicable to a wider range of forensic cases.



Mean	St. Dev	Median	Mode	Skewness	Kurtosis
27.319	10.395	25	19	1.52	3.20

(a) Selected cases.



Mean	St. Dev	Median	Mode	Skewness	Kurtosis
26.680	9.557	24	21	1.63	4.11

(b) Selected images.

Figure 3: Frequency of age at time of death.

In the context of this work, data augmentation helps in artificially increasing the size of the dataset, which is crucial for training deep learning models effectively. These techniques can help to balance the dataset by creating synthetic examples of the minority classes, improving the model's ability to recognize and classify all types of wounds and distances. However, there is a wide variety of augmentation techniques available in the literature [23], and choosing the appropriate set of augmentations for the problem can be quite challenging. Specifically, we use 13 augmentation techniques, including variants of rotation, flipping, cropping and resizing, brightness and contrast adjustments, Gaussian noise, color jitter (changes in hue and saturation), elastic transformations, and grayscale conversion. Examples results of these augmentation techniques are depicted in Figure 2.

The augmentation illustrated in Figure 2 were chosen to handle variations that may occur in real-world forensic cases. Rotation, flipping, and further geometric transformations such as SR can help the model learn to recognize wounds from different angles, making the model more adaptable to variations in the size and anatomical location of wounds. Elastic and affine transformations can mimic the natural variability in skin and tissue appearance, making the model more robust to such changes. Varying the brightness and contrast can help the model become robust to different lighting conditions that may be encountered in real-world forensic images taken from different crime scenes. Some noise and color augmentations such as ISOnoise, RBC, RGBShift, and Gaussian noise can make the model more resilient to image quality variations, such as those caused by different imaging equipment or environmental conditions. These augmentation techniques were discovered with AutoAulbument [24], an AutoML tool that automatically searches for the best augmentation policies based on training data. Using the its classification model, AutoAlbument provides a complete ready-to-use configuration for the augmentation pipeline. After discovered these augmentation techniques set, they are then incorporated in the training procedure with the Albumentations framework [25].

In practice, when using an augmentation library, it is essential that the quality of the synthetic samples does not deteriorate to a point of impairing performance on normal wound images. To avoid it, we combine synthetic examples with the original wound images rather than using only synthetic ones. The inclusion of the original preprocessed wound image sample was performed by adding the identity function as property in the Albumentations pipeline, which returns the image itself. This approach preserves the actual characteristics of the original data, ensuring that the model learns from authentic features and patterns present in the dataset. By combining original and augmented images, the model learns to recognize patterns from both the natural data distribution and the variations introduced by augmentation. By using this combination of training data, the model receives balanced exposure to both real and augmented data, enhancing its robustness to variations in real-world scenarios.

After combining the training data, a data resampling strategy is implemented in order to further reduce the remaining imbalance after data augmentation. Particularly, a combination of over- and under-sampling using Synthetic Minority Over-sampling Technique (SMOTE) and Edited Nearest Neighbours. (ENN) is applied for handling class imbalance with fewer noisy instances, leading to improved performance and robustness of ML models. It combines the oversampling of the minority class using SMOTE with the cleaning of the dataset using ENN [26]. This resampling strategy is implemented via Imbalanced-learn [27], an open-source Python library designed to handle imbalanced datasets in ML. Finally, the class weights are computed to train the TorchVision models.

2.2.4 Experimental Setup

The experiments were performed on a Linux machine (Ubuntu 22.04 LTS) using an AI software system developed by ourselves using the Python language (version 3.10). Computer hardware specifications included an Intel i7-8700 CPU, 32GB of RAM, and an RTX 3090 GPU. The TorchVision models are trained and tested separately. For each model and training round, 1000 epochs are used. The models trained with a batch size of 16 include RegnetY128GF and ViTH14. The models trained with a batch size of 32 include EfficientnetB6, EfficientnetB7, EfficientnetV2L, RegnetY128GF, Squeezenet10, SwinTransformerT, SwinTransformerS, SwinTransformerB, SwinTransformerV2T, SwinTransformerV2S, SwinTransformerV2B, ShuffleNetV2x20, ShuffleNetV2x15, ViTB16, ViTB32, ViTL16, ViTL32, ViTH14, and MaxVit. The remaining models not mentioned were trained with a batch size of 64. Early stopping with checkpointing is adopted to save the weights that perform best over the validation set during training.

Two cross-validation approaches were used: holdout and k-fold. For both approaches, the protocol consists of splitting the database into two content independent subsets —

one subset for training and another for testing. To avoid overfitting, the training and test data are split considering stratification based on the case number. This means that images associated with cases used in the training subset are not used for testing, and vice versa. With this constraint, 80% of cases are randomly selected for training, and the remaining 20% are used for testing in holdout cross-validation. Similarly, stratified grouped k-fold is adopted to ensure that each fold of the cross-validation procedure contains a balanced representation of these groups, providing a more realistic evaluation of the model's generalization capability. In this case, the groups used for stratification correspond to the crime cases.

We evaluate the performance of the tested models using regular classification metrics. More precisely, the models were assessed according to the following metrics:

• Accuracy: calculate the percentage of its correct predictions. It is defined as follows

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}.$$

• Precision: because the accuracy score may have as drawbacks the imbalance problem and being uninformative as a standalone classification metric, the precision measures the ability of a classifier not to label as Positive a Negative sample. It is defined as follows

$$\text{Precision} = \frac{TP}{TP + FP}.$$

• Recall: it is a ratio of predictions of the Positive class that are Positive by ground truth to the total number of Positive samples. In other words, recall measures the ability of a classifier to detect Positive samples. It is defined as follows

$$\text{Recall} = \frac{TP}{TP + FN}.$$

• F1-Score: it is the harmonic mean of precision and recall, representing both metrics in one score. The highest possible value of an F-score is 1, indicating perfect precision and recall, and the lowest possible value is 0, if either precision or recall is zero. It is defined as follows:

$$\text{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}.$$

• Specificity: it measures the ability of a classifier to correctly identify negative samples. It is calculated as the proportion of true negatives (TN) out of all actual negatives (TN + FP). Specificity is particularly useful when the cost of false positives is high, providing a comprehensive understanding of the performance of a classification model. It is defined as follows:

$$\text{Specificity} = \frac{TN}{TN + FP}.$$

• Area under the Curve (AUC): it evaluates the performance of a classification model by measuring the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (sensitivity) against the False Positive Rate (1 - specificity) at various threshold settings. The ROC curve is created by plotting the True Positive Rate (TPR)against the False Positive Rate (FPR)at various threshold settings. The AUC is then computed as the area under this ROC curve:

$$AUC \ = \int_0^1 TPRd(FPR)$$

$$= \sum_{i=1}^{n} \left(\frac{TPR_i + FPR_i}{2}\right).(FPR_i \ - \ FPR_{i-1})$$

where

$$\text{TPR} = \frac{TP}{TP + FN} \ \ and \ \text{FPR} = \frac{FP}{FP + TN} \ ,$$

are the ratio of true positives to all actual positives (also known as sensitivity) and the ratio of false positives to all actual negatives, respectively. Using the trapezoidal rule, n is the number of thresholds.

In the context of evaluating classification models, the above-described performance metrics can be aggregated with macro, micro, and weighted averages. Here is a brief overview of each:

• Macro average calculates the metric (e.g., precision, recall, or F1-Score) for each class individually and then takes the average of these metrics. In other words, it treats all classes equally, regardless of their frequency as follows:

$$Macro\ \Phi = \frac{1}{C}\sum_{i=1}^{C}\Phi_i,$$

• Micro average aggregates the contributions of all classes to compute the average metric. It treats each instance equally, regardless of class.

• Weighted average calculates the metric for each class and then takes the average weighted by the number of instances in each class. It can be expressed as:

$$Weighted\ \Phi = \frac{\sum_{i=1}^{C}(\Phi_1 \times Number\ of\ Instances\ in\ Class_i)}{\sum_{i=1}^{C} Number\ of\ Instances\ in\ Class_i}.$$

2.2.5 Cross-validation

In order to observe whether the ResNet152 results are stable, we conducted a K-fold cross-validation to ensure that this model is evaluated comprehensively and reliably. Using a boxplot to illustrate the k-fold cross-validation results is an effective way to visualize the distribution of the performance metrics across the different folds. The boxplot displays the median, quartiles, and potential outliers of the wound images, providing a clear summary of the results.

2.3 RESULTS

In this study, 2,551 wound images were analyzed, primarily from male subjects (95.09%) (Table 1), with most images showing wounds located on the trunk (51.35%) (Table 3). The majority of the wounds were classified as entry wounds (1,883), with 668 exit wounds. In terms of MLSD, most entry wounds were classified as distant (85.45%), while close-range and contact wounds were less frequent (Table 2). The age distribution of the cases indicated a predominance of younger victims, with a mean age of approximately 27 years and a significant skew towards individuals in their 20s (Figures 3a and 3b). These findings highlight key demographic and injury pattern trends in gunshot wound cases.

2.3.1 Wound Type Classification

Table 4 shows the accuracy, precision, recall and F1-score, AUC and specificity of would type classification using a houldout approach. This table discriminates the performance in terms of neural network architectures and its variants provided in the TorchVision library. In this table, the boldfaced values highlight the best model. Moreover, Table 4 shows the micro, macro, and weighted averages for each metric. From these results, we can notice that the results vary a lot for different models. However, it is noticeable that ResNet presents the best results. Among the variants of ResNet, the ResNet152 achieves the best performance for all metrics. Since The ResNet 152 achieved the highest values for 14 out of 16 evaluated metrics, we chose this model as the most suitable for wound type classification.

Figure 4a presents a boxplot with the results of the ResNet 152’s wound type classification metrics.

2.3.2 Medico-legal Shooting Distance Classification

Similarly to what was done for the classification of gunshot wound types, the results for the detection of MLSD are presented in Table 5. The ResNet 152 achieved the highest scores for the used metrics. The only exceptions were the macro precision values, which were higher for the ResNet 50 version, and the weighted specificity indices, for which the SqueezeNet 10 version reached the highest mark. Figure 4b presents the distribution of ResNet 152 MLSD classification metrics.

Table 4 – Performance metrics for different neural network architectures using Holdout validation to predict the wound type. Considering each metric separately, the best results in terms of maximum absolute value are boldfaced. Symbols M, m, and W illustrates macro, micro, and weighted, respectively

Architecture	Variant	ACC	Precision			Recall			F1-Score			AUC			Specificity		
			M	m	W	M	m	W	M	m	W	M	m	W	M	m	W
Alexnet [27]	-	0.7651	0.6948	0.7651	0.7799	0.7155	0.7651	0.7651	0.7029	0.7651	0.7710	0.7155	0.7155	0.7155	0.7155	0.7651	0.7651
Densenet [28]	121	0.7360	0.6401	0.7360	0.7266	0.6294	0.7360	0.7360	0.6339	0.7360	0.7307	0.6294	0.6294	0.6294	0.6294	0.7360	0.7360
	161	0.7526	0.6594	0.7526	0.7366	0.6349	0.7526	0.7526	0.6436	0.7526	0.7424	0.6349	0.6349	0.6349	0.6349	0.7526	0.7526
	169	0.7775	0.6995	0.7775	0.7679	0.6793	0.7775	0.7775	0.6877	0.7775	0.7716	0.6793	0.6793	0.6793	0.6793	0.7775	0.7775
	201	0.7443	0.6632	0.7443	0.7508	0.6711	0.7443	0.7443	0.6668	0.7443	0.7473	0.6711	0.6711	0.6711	0.6711	0.7443	0.7443
Efficientnet [29, 30]	B0	0.7443	0.6463	0.7443	0.7274	0.6238	0.7443	0.7443	0.6316	0.7443	0.7337	0.6238	0.6238	0.6238	0.6238	0.7443	0.7443
	B1	0.7339	0.6268	0.7339	0.7125	0.6030	0.7339	0.7339	0.6101	0.7339	0.7202	0.6030	0.6030	0.6030	0.6030	0.7339	0.7339
	B2	0.7505	0.6557	0.7505	0.7337	0.6307	0.7505	0.7505	0.6394	0.7505	0.7397	0.6307	0.6307	0.6307	0.6307	0.7505	0.7505
	B3	0.7568	0.6589	0.7568	0.7277	0.6043	0.7568	0.7568	0.6148	0.7568	0.7320	0.6043	0.6043	0.6043	0.6043	0.7568	0.7568
	B4	0.7152	0.5875	0.7152	0.6840	0.5655	0.7152	0.7152	0.5692	0.7152	0.6948	0.5655	0.5655	0.5655	0.5655	0.7152	0.7152
	B5	0.7443	0.6080	0.7443	0.6864	0.5348	0.7443	0.7443	0.5169	0.7443	0.6830	0.5348	0.5348	0.5348	0.5348	0.7443	0.7443
	B6	0.7588	0.6672	0.7588	0.7396	0.6335	0.7588	0.7588	0.6442	0.7588	0.7454	0.6335	0.6335	0.6335	0.6335	0.7588	0.7588
	B7	0.7505	0.3753	0.7505	0.5633	0.5000	0.7505	0.7505	0.4287	0.7505	0.6436	0.5000	0.5000	0.5000	0.5000	0.7505	0.7505
	V2L	0.7817	0.7097	0.7817	0.7603	0.6376	0.7817	0.7817	0.6543	0.7817	0.7595	0.6376	0.6376	0.6376	0.6376	0.7817	0.7817
	V2M	0.7256	0.6019	0.7256	0.6925	0.5724	0.7256	0.7256	0.5771	0.7256	0.7027	0.5724	0.5724	0.5724	0.5724	0.7256	0.7256
	V2S	0.7214	0.6211	0.7214	0.7137	0.6141	0.7214	0.7214	0.6172	0.7214	0.7173	0.6141	0.6141	0.6141	0.6141	0.7214	0.7214
MaxVit [31]	-	0.7796	0.7025	0.7796	0.7611	0.6529	0.7796	0.7796	0.6679	0.7796	0.7644	0.6529	0.6529	0.6529	0.6529	0.7796	0.7796
MNASNet [32]	5	0.7048	0.5887	0.7048	0.6883	0.5780	0.7048	0.7048	0.5816	0.7048	0.6953	0.5780	0.5780	0.5780	0.5780	0.7048	0.7048
	75	0.7173	0.5903	0.7173	0.6856	0.5669	0.7173	0.7173	0.5708	0.7173	0.6964	0.5669	0.5669	0.5669	0.5669	0.7173	0.7173
	10	0.6819	0.5703	0.6819	0.6775	0.5683	0.6819	0.6819	0.5692	0.6819	0.6796	0.5683	0.5683	0.5683	0.5683	0.6819	0.6819
	13	0.7464	0.6594	0.7464	0.7436	0.6558	0.7464	0.7464	0.6575	0.7464	0.7449	0.6558	0.6558	0.6558	0.6558	0.7464	0.7464
MobileNet [33, 34]	V2	0.7193	0.6085	0.7193	0.7016	0.5933	0.7193	0.7193	0.5983	0.7193	0.7088	0.5933	0.5933	0.5933	0.5933	0.7193	0.7193
	V3Large	0.6902	0.5655	0.6902	0.6717	0.5572	0.6902	0.6902	0.5595	0.6902	0.6798	0.5572	0.5572	0.5572	0.5572	0.6902	0.6902
	V3Small	0.7027	0.5846	0.7027	0.6852	0.5738	0.7027	0.7027	0.5773	0.7027	0.6927	0.5738	0.5738	0.5738	0.5738	0.7027	0.7027
Regnet [35]	X16GF	0.6902	0.5339	0.6902	0.6477	0.5238	0.6902	0.6902	0.5208	0.6902	0.6636	0.5238	0.5238	0.5238	0.5238	0.6902	0.6902

	X32GF	0.7235	0.5800	0.7235	0.6753	0.5460	0.7235	0.7235	0.5436	0.7235	0.6871	0.5460	0.5460	0.5460	0.5460	0.7235	0.7235
	X400MF	0.6632	0.5283	0.6632	0.6457	0.5253	0.6632	0.6632	0.5257	0.6632	0.6536	0.5253	0.5253	0.5253	0.5253	0.6632	0.6632
	X800MF	0.7027	0.5628	0.7027	0.6672	0.5460	0.7027	0.7027	0.5470	0.7027	0.6801	0.5460	0.5460	0.5460	0.5460	0.7027	0.7027
	X8GF	0.6985	0.5578	0.6985	0.6641	0.5433	0.6985	0.6985	0.5440	0.6985	0.6770	0.5433	0.5433	0.5433	0.5433	0.6985	0.6985
	Y16GF	0.7110	0.5710	0.7110	0.6717	0.5488	0.7110	0.7110	0.5495	0.7110	0.6847	0.5488	0.5488	0.5488	0.5488	0.7110	0.7110
	Y32GF	0.6715	0.5329	0.6715	0.6485	0.5280	0.6715	0.6715	0.5284	0.6715	0.6586	0.5280	0.5280	0.5280	0.5280	0.6715	0.6715
	Y400MF	0.6736	0.5401	0.6736	0.6538	0.5350	0.6736	0.6736	0.5359	0.6736	0.6626	0.5350	0.5350	0.5350	0.5350	0.6736	0.6736
	Y800MF	0.7027	0.5653	0.7027	0.6691	0.5488	0.7027	0.7027	0.5503	0.7027	0.6815	0.5488	0.5488	0.5488	0.5488	0.7027	0.7027
	Y8GF	0.7027	0.5628	0.7027	0.6672	0.5460	0.7027	0.7027	0.5470	0.7027	0.6801	0.5460	0.5460	0.5460	0.5460	0.7027	0.7027
ResNet [36]	152	**0.8690**	**0.8265**	**0.8690**	**0.8680**	**0.8209**	**0.8690**	**0.8690**	**0.8236**	**0.8690**	**0.8685**	**0.8209**	**0.8209**	**0.8209**	**0.8209**	**0.8690**	**0.8690**
	101	0.8628	0.8215	0.8628	0.8596	0.8029	0.8628	0.8628	0.8114	0.8628	0.8607	0.8029	0.8029	0.8029	0.8029	0.8628	0.8628
	50	0.8503	0.8021	0.8503	0.8480	0.7918	0.8503	0.8503	0.7967	0.8503	0.8490	0.7918	0.7918	0.7918	0.7918	0.8503	0.8503
	34	0.8503	0.7992	0.8503	0.8521	0.8057	0.8503	0.8503	0.8023	0.8503	0.8511	0.8057	0.8057	0.8057	0.8057	0.8503	0.8503
	18	0.8170	0.7580	0.8170	0.8101	0.7335	0.8170	0.8170	0.7440	0.8170	0.8125	0.7335	0.7335	0.7335	0.7335	0.8170	0.8170
ShuffleNet [37]	V2x05	0.7297	0.6279	0.7297	0.7165	0.6141	0.7297	0.7297	0.6194	0.7297	0.7221	0.6141	0.6141	0.6141	0.6141	0.7297	0.7297
	V2x10	0.7006	0.5327	0.7006	0.6461	0.5196	0.7006	0.7006	0.5121	0.7006	0.6641	0.5196	0.5196	0.5196	0.5196	0.7006	0.7006
	V2x15	0.7110	0.5801	0.7110	0.6790	0.5599	0.7110	0.7110	0.5629	0.7110	0.6904	0.5599	0.5599	0.5599	0.5599	0.7110	0.7110
	V2x20	0.6923	0.5508	0.6923	0.6598	0.5391	0.6923	0.6923	0.5396	0.6923	0.6725	0.5391	0.5391	0.5391	0.5391	0.6923	0.6923
SqueezeNet [38]	10	0.7505	0.3753	0.7505	0.5633	0.5000	0.7505	0.7505	0.4287	0.7505	0.6436	0.5000	0.5000	0.5000	0.5000	0.7505	0.7505
	11	0.7505	0.3753	0.7505	0.5633	0.5000	0.7505	0.7505	0.4287	0.7505	0.6436	0.5000	0.5000	0.5000	0.5000	0.7505	0.7505
SwinT [39]	B	0.7277	0.6143	0.7277	0.7033	0.5905	0.7277	0.7277	0.5968	0.7277	0.7119	0.5905	0.5905	0.5905	0.5905	0.7277	0.7277
	S	0.6715	0.5356	0.6715	0.6505	0.5308	0.6715	0.6715	0.5314	0.6715	0.6598	0.5308	0.5308	0.5308	0.5308	0.6715	0.6715
	T	0.7131	0.5944	0.7131	0.6907	0.5780	0.7131	0.7131	0.5825	0.7131	0.6995	0.5780	0.5780	0.5780	0.5780	0.7131	0.7131
	V2B	0.7505	0.3753	0.7505	0.5633	0.5000	0.7505	0.7505	0.4287	0.7505	0.6436	0.5000	0.5000	0.5000	0.5000	0.7505	0.7505
	V2S	0.7131	0.5574	0.7131	0.6613	0.5335	0.7131	0.7131	0.5284	0.7131	0.6763	0.5335	0.5335	0.5335	0.5335	0.7131	0.7131
	V2T	0.7505	0.3753	0.7505	0.5633	0.5000	0.7505	0.7505	0.4287	0.7505	0.6436	0.5000	0.5000	0.5000	0.5000	0.7505	0.7505
VGG [40]	11	0.7630	0.6734	0.7630	0.7424	0.6335	0.7630	0.7630	0.6454	0.7630	0.7477	0.6335	0.6335	0.6335	0.6335	0.7630	0.7630
	13	0.7900	0.7200	0.7900	0.7743	0.6710	0.7900	0.7900	0.6869	0.7900	0.7769	0.6710	0.6710	0.6710	0.6710	0.7900	0.7900
	16	0.7817	0.7057	0.7817	0.7723	0.6849	0.7817	0.7817	0.6935	0.7817	0.7759	0.6849	0.6849	0.6849	0.6849	0.7817	0.7817

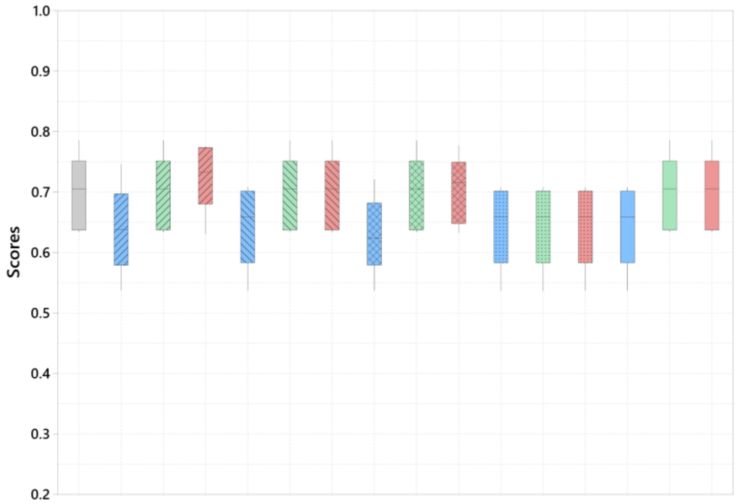
	19	0.7775	0.6995	0.7775	0.7679	0.6793	0.7775	0.7775	0.6877	0.7775	0.7716	0.6793	0.6793	0.6793	0.6793	0.7775	0.7775
ViT [41]	B16	0.7131	0.5869	0.7131	0.6841	0.5669	0.7131	0.7131	0.5707	0.7131	0.6946	0.5669	0.5669	0.5669	0.5669	0.7131	0.7131
	B32	0.6736	0.5589	0.6736	0.6690	0.5572	0.6736	0.6736	0.5580	0.6736	0.6712	0.5572	0.5572	0.5572	0.5572	0.6736	0.6736
	L16	0.6778	0.5439	0.6778	0.6563	0.5378	0.6778	0.6778	0.5388	0.6778	0.6657	0.5378	0.5378	0.5378	0.5378	0.6778	0.6778
	L32	0.6923	0.5631	0.6923	0.6694	0.5530	0.6923	0.6923	0.5553	0.6923	0.6790	0.5530	0.5530	0.5530	0.5530	0.6923	0.6923

Table 5 – Performance metrics for different neural network architectures using Holdout validation to predict the medico-legal shooting distance. Considering each metric separately, the best results in terms of maximum absolute value are boldfaced. Symbols M, m, and W illustrates macro, micro, and weighted, respectively.

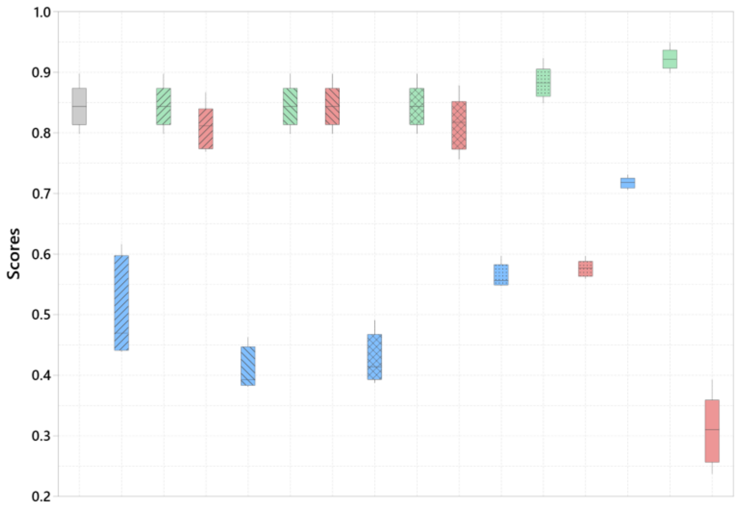
Architecture	Variant	ACC	Precision			Recall			F1-Score			AUC			Specificity		
			M	m	W	M	m	W	M	m	W	M	m	W	M	m	W
Alexnet [27]	-	0.8886	0.5371	0.8886	0.8623	0.3782	0.8886	0.8886	0.3947	0.8886	0.8559	0.5408	0.9164	0.5551	0.7034	0.9443	0.2215
Densenet [28]	121	0.8774	0.4221	0.8774	0.8251	0.3565	0.8774	0.8774	0.359	0.8774	0.841	0.5207	0.9081	0.5275	0.685	0.9387	0.1775
	161	0.8942	0.8238	0.8942	0.8708	0.4646	0.8942	0.8942	0.5276	0.8942	0.8625	0.5885	0.9206	0.5686	0.7124	0.9471	0.243
	169	0.883	0.4667	0.883	0.8195	0.3876	0.883	0.883	0.3977	0.883	0.844	0.548	0.9123	0.5626	0.7084	0.9415	0.2423
	201	0.9025	0.6336	0.9025	0.883	0.4825	0.9025	0.9025	0.527	0.9025	0.8861	0.6275	0.9269	0.6587	0.7724	0.9513	0.4148
Efficientnet [29, 30]	B0	0.8774	0.758	0.8774	0.8566	0.438	0.8774	0.8774	0.4829	0.8774	0.8577	0.5828	0.9081	0.5913	0.7275	0.9387	0.3052
	B1	0.8858	0.6701	0.8858	0.8515	0.4615	0.8858	0.8858	0.5081	0.8858	0.8566	0.5854	0.9143	0.5641	0.7094	0.9429	0.2423
	B2	0.8635	0.3926	0.8635	0.8173	0.36	0.8635	0.8635	0.3642	0.8635	0.8359	0.5235	0.8976	0.5306	0.687	0.9318	0.1976
	B3	0.8997	0.5602	0.8997	0.8734	0.3999	0.8997	0.8997	0.4239	0.8997	0.8689	0.5607	0.9248	0.5823	0.7215	0.9499	0.2649
	B4	0.883	0.4504	0.883	0.8372	0.3761	0.883	0.883	0.3864	0.883	0.851	0.5386	0.9123	0.5517	0.7012	0.9415	0.2205
	B5	0.8691	0.3575	0.8691	0.8063	0.3446	0.8691	0.8691	0.3399	0.8691	0.833	0.5132	0.9018	0.5227	0.6667	0.9429	0.1142
	B6	0.8858	0.2953	0.8858	0.7846	0.3333	0.8858	0.8858	0.3131	0.8858	0.8321	0.5	0.9143	0.5	0.6667	0.9429	0.1142
	B7	0.8858	0.4627	0.8858	0.8339	0.3421	0.8858	0.8858	0.3316	0.8858	0.8373	0.5079	0.9143	0.5107	0.6738	0.9429	0.1355
	V2L	0.8802	0.6167	0.8802	0.854	0.4479	0.8802	0.8802	0.4882	0.8802	0.8613	0.5882	0.9102	0.593	0.7286	0.9401	0.3057
	V2M	0.883	0.6137	0.883	0.846	0.4139	0.883	0.883	0.4484	0.883	0.8519	0.5576	0.9123	0.5519	0.7012	0.9415	0.2207
	V2S	0.8747	0.5918	0.8747	0.8427	0.4282	0.8747	0.8747	0.4646	0.8747	0.8524	0.5703	0.906	0.5686	0.7124	0.9373	0.2625
MaxVit [31]	-	0.8802	0.7648	0.8802	0.8585	0.4857	0.8802	0.8802	0.5494	0.8802	0.862	0.6071	0.9102	0.5928	0.7286	0.9401	0.3054
MNASNet [32]	5	0.844	0.344	0.844	0.8062	0.3439	0.844	0.844	0.342	0.844	0.824	0.519	0.883	0.5411	0.6941	0.922	0.2382
	75	0.8719	0.5214	0.8719	0.8432	0.436	0.8719	0.8719	0.4621	0.8719	0.8532	0.5773	0.9039	0.5779	0.7186	0.9359	0.284
	10	0.8719	0.5657	0.8719	0.8374	0.4184	0.8719	0.8719	0.4505	0.8719	0.8495	0.5684	0.9039	0.5777	0.7185	0.9359	0.2835
	13	0.8691	0.4095	0.8691	0.8244	0.3708	0.8691	0.8691	0.3777	0.8691	0.8421	0.5335	0.9018	0.5442	0.6961	0.9345	0.2193
MobileNet [33, 34]	V2	0.8858	0.4241	0.8858	0.8305	0.3596	0.8858	0.8858	0.3615	0.8858	0.8477	0.5309	0.9143	0.5533	0.7022	0.9429	0.2208
	V3Large	0.8719	0.4112	0.8719	0.829	0.3806	0.8719	0.8719	0.3875	0.8719	0.8472	0.5459	0.9039	0.5669	0.7113	0.9359	0.2619
	V3Small	0.8914	0.5477	0.8914	0.862	0.3617	0.8914	0.8914	0.3666	0.8914	0.8494	0.5259	0.9185	0.535	0.69	0.9457	0.1787

Regnet [35]	X16GF	0.8552	0.2941	0.8552	0.7814	0.3218	0.8552	0.8552	0.3073	0.8552	0.8166	0.4886	0.8914	0.4831	0.6554	0.9276	0.111
	X32GF	0.8719	0.3703	0.8719	0.8081	0.3456	0.8719	0.8719	0.3414	0.8719	0.8336	0.5107	0.9039	0.5136	0.6757	0.9359	0.1554
	X400MF	0.844	0.3581	0.844	0.8078	0.3526	0.844	0.844	0.3529	0.844	0.8247	0.5198	0.883	0.5304	0.687	0.922	0.2169
	X800MF	0.8552	0.4869	0.8552	0.8042	0.3771	0.8552	0.8552	0.3956	0.8552	0.8257	0.527	0.8914	0.5152	0.6768	0.9276	0.1752
	X8GF	0.8496	0.3487	0.8496	0.8017	0.346	0.8496	0.8496	0.3437	0.8496	0.8238	0.5103	0.8872	0.512	0.6746	0.9248	0.1744
	Y16GF	0.8747	0.3979	0.8747	0.82	0.3554	0.8747	0.8747	0.3568	0.8747	0.8402	0.5232	0.906	0.5366	0.6911	0.9373	0.1985
	Y32GF	0.8635	0.6595	0.8635	0.8143	0.3803	0.8635	0.8635	0.4069	0.8635	0.8296	0.5265	0.8976	0.509	0.6727	0.9318	0.1545
	Y400MF	0.8691	0.3332	0.8691	0.7976	0.3358	0.8691	0.8691	0.3259	0.8691	0.8292	0.5053	0.9018	0.5121	0.6747	0.9345	0.1551
	Y800MF	0.8802	0.5307	0.8802	0.8189	0.3866	0.8802	0.8802	0.4036	0.8802	0.8395	0.5363	0.9102	0.529	0.686	0.9401	0.1778
	Y8GF	0.8607	0.3236	0.8607	0.7941	0.3327	0.8607	0.8607	0.3233	0.8607	0.8246	0.5021	0.8955	0.5075	0.6716	0.9304	0.1542
ResNet [36]	152	**0.9248**	0.8096	**0.9248**	**0.9171**	**0.6366**	**0.9248**	**0.9248**	**0.7006**	**0.9248**	**0.9175**	**0.7334**	**0.9436**	**0.7454**	**0.8303**	**0.9624**	**0.5661**
	101	0.9136	0.8134	0.9136	0.9033	0.5711	0.9136	0.9136	0.6418	0.9136	0.8965	0.6738	0.9352	0.6647	0.7765	0.9568	0.4157
	50	0.9109	**0.8624**	0.9109	0.8981	0.6166	0.9109	0.9109	0.6956	0.9109	0.8956	0.696	0.9331	0.663	0.7754	0.9554	0.4152
	34	0.8969	0.7324	0.8969	0.8852	0.5911	0.8969	0.8969	0.6435	0.8969	0.889	0.6913	0.9227	0.6874	0.7916	0.9485	0.4779
	18	0.9136	0.5571	0.9136	0.8871	0.4489	0.9136	0.9136	0.4787	0.9136	0.8924	0.6126	0.9352	0.6646	0.7764	0.9568	0.4155
ShuffleNet [37]	V2x05	0.8552	0.3551	0.8552	0.8041	0.3481	0.8552	0.8552	0.3463	0.8552	0.8271	0.5124	0.8914	0.515	0.6767	0.9276	0.1749
	V2x10	0.8774	0.4115	0.8774	0.8298	0.374	0.8774	0.8774	0.3802	0.8774	0.8489	0.5436	0.9081	0.57	0.7133	0.9387	0.2625
	V2x15	0.8691	0.337	0.8691	0.7967	0.3358	0.8691	0.8691	0.3258	0.8691	0.8281	0.5017	0.9018	0.5014	0.6676	0.9345	0.1338
	V2x20	0.8802	0.3791	0.8802	0.8097	0.34	0.8802	0.8802	0.3296	0.8802	0.8342	0.5059	0.9102	0.5076	0.6717	0.9401	0.1349
SqueezeNet [38]	10	0.0947	0.0316	0.0947	0.009	0.3333	0.0947	0.0947	0.0577	0.0947	0.0164	0.5	0.321	0.5	0.6667	0.5474	0.9053
	11	0.8858	0.2953	0.8858	0.7846	0.3333	0.8858	0.8858	0.3131	0.8858	0.8321	0.5	0.9143	0.5	0.6667	0.9429	0.1142
SwinT [39]	B	0.8607	0.3539	0.8607	0.8086	0.3414	0.8607	0.8607	0.3386	0.8607	0.831	0.5173	0.8955	0.5399	0.6933	0.9304	0.2191
	S	0.8552	0.5225	0.8552	0.8159	0.3946	0.8552	0.8552	0.4203	0.8552	0.8313	0.5393	0.8914	0.5258	0.6839	0.9276	0.1965
	T	0.8719	0.727	0.8719	0.8397	0.4097	0.8719	0.8719	0.4491	0.8719	0.8449	0.5534	0.9039	0.5456	0.6971	0.9359	0.2193
	V2B	0.8858	0.2953	0.8858	0.7846	0.3333	0.8858	0.8858	0.3131	0.8858	0.8321	0.5	0.9143	0.5	0.6667	0.9429	0.1142
	V2S	0.8607	0.405	0.8607	0.8263	0.3764	0.8607	0.8607	0.3839	0.8607	0.841	0.5419	0.8955	0.5611	0.7074	0.9304	0.2615
	V2T	0.8719	0.4189	0.8719	0.8255	0.3631	0.8719	0.8719	0.3695	0.8719	0.8413	0.5266	0.9039	0.5352	0.6901	0.9359	0.1985
VGG [40]	11	0.9053	0.6663	0.9053	0.8886	0.466	0.9053	0.9053	0.5143	0.9053	0.8839	0.6126	0.929	0.6389	0.7593	0.9526	0.3725
	13	0.8747	0.5425	0.8747	0.8453	0.437	0.8747	0.8747	0.4674	0.8747	0.8558	0.5818	0.906	0.59	0.7267	0.9373	0.3053

	16	0.8635	0.3916	0.8635	0.8249	0.3775	0.8635	0.8635	0.3808	0.8635	0.8425	0.5464	0.8976	0.5729	0.7153	0.9318	0.2824
	19	0.883	0.5782	0.883	0.8521	0.4314	0.883	0.883	0.4658	0.883	0.8582	0.5735	0.9123	0.5733	0.7155	0.9415	0.2636
ViT [41]	B16	0.8384	0.342	0.8384	0.8009	0.3418	0.8384	0.8384	0.34	0.8384	0.8186	0.5098	0.8788	0.5167	0.6778	0.9192	0.195
	B32	0.8747	0.4247	0.8747	0.8371	0.3904	0.8747	0.8747	0.3996	0.8747	0.8528	0.5585	0.906	0.5899	0.7266	0.9373	0.3051
	L16	0.8691	0.7078	0.8691	0.8321	0.3999	0.8691	0.8691	0.4356	0.8691	0.84	0.5444	0.9018	0.5334	0.6889	0.9345	0.1977
	L32	0.8886	0.5371	0.8886	0.8623	0.3782	0.8886	0.8886	0.3947	0.8886	0.8559	0.5408	0.9164	0.5551	0.7034	0.9443	0.2215



(a) Wound type.



(b) Medico-legal shooting distance.

Figure 4: Classification performance for wound type (a) and shooting distance (b) using ResNet 152 over the k-folds.

2.4 DISCUSSION

This study marks a significant advancement in forensic science as it is among the first to employ DL tools to classify gunshot wound patterns with extensive labels such as entry and exit wounds and MLSD, utilizing a broad array of neural network architectures. Unique in its approach, this research utilizes a considerable dataset of forensic images captured during real crime scene investigations, a feat not matched by any previous work. The implementation of these methodologies holds the potential to notably accelerate the pace of real criminal investigations by offering rapid, reliable analyses that enhance traditional forensic methods. Furthermore, the technology serves as a complementary tool for forensic experts during the crucial phases of examination and report drafting, promoting more accurate and informed decision-making.

Clearly, for ethical reasons, conducting research with a completely controlled sample of gunshot wounds on human cadavers, like studies of Phelps in 1897 [43], is virtually impractical. Therefore, many studies still use animal carcasses [17, 44, 45]. However, as much as these may approximate what is seen in humans, there are significant differences [46] that could lead to the incorrect training of the ML model for future use in human cases.

In this study, for classifying entry and exit gunshot wounds images, the use of ResNet152 achieved percentages, with the exception of the AUC results, closely align with those by *Cheng et al.* [16], which reported an accuracy of 87.99%, precision of 83.99%, recall of 87.71%, specificity of 88.19%, F1-score of 85.81% and AUC of 94.6% with a larger sample, comprising 2,028 entry and 1,314 exit wound images. However, the images do not appear to have been obtained at crime scenes, but rather during necropsy examinations, which indicates a certain level of control over the captured images.

Additionally, in shooting distance classification, this study achieved a 92.48% accuracy rate using ResNet152. However, despite the high accuracy, the sample imbalance may have negatively impacted other metrics, including recall and the F1-Score. While newer models like ViT and EfficientNet offer advanced techniques and may perform exceptionally well in many scenarios, ResNet's combination of residual connections, depth, pretraining, and optimization make it a robust choice for a wide range of tasks, including the MLSD classification task. In contrast, Oura et al. [17] reached a 98% accuracy rate. However, this higher accuracy was probably facilitated by their use of a small, controlled sample comprising only 204 images in a laboratory setting. It is important to note that their study utilized pig carcasses instead of wound on humans corpses, which could affect the applicability of their findings to real-world forensic scenarios.

With this, it becomes evident that one of the greatest challenges in developing ML models for image categorization is acquiring a database with high-quality scientific data [47]. Due to the sensitive content nature of the area, this challenge may be even greater in the forensic field. Therefore, the photographs used in this study were taken by different individuals, under various environmental and exposure conditions, with a range of cameras, and in the routine circumstances of forensic work at crime scenes, where standardization becomes virtually unfeasible. Given this, it is understood that a standardized database would not only misalign with the realities of forensic work, where numerous factors can alter the characteristics of injuries, but also require an exhaustive effort.

The challenge in standardizing a database of gunshot wound images is also evident in the discrepancy between the number of entry (1,883) and exit wounds (668). Logically, for an exit wound to occur, there must inevitably be an entry wound. Additionally, in this study, inclusion criterion #5 required that bullets be lodged in the body. Therefore, it is expected that the number of entry wounds would surpass that of exit wounds. Furthermore, even with access to a database containing millions of images from crime scene investigations, the utility of this dataset was significantly limited. Many images were related to other types of police incidents or featured different instruments (e.g., knives), had poor image quality, and faced restrictive criteria, all of which significantly constrained the acquisition of a larger, more useful sample. Despite these limitations, the sample used in this study more accurately reflects the conditions forensic teams encounter in real-world scenarios and deviates from the laboratory standardization typical of other research [16, 17]. This approach enhances the practical relevance of the findings to real-world forensic investigations.

In turn, the classification of distance shooting is used in cases where the characteristics of close-range and contact shots are not present (examples: smoke pattern and tattooing) [48] and typically occur at a distance range starting from 18 inches-2 feet (0.45 to 0.60 meters) [5, 49]. Therefore, for MLSD classification purposes, there is a tendency for the majority of injuries to be of the distance type, as per the sample obtained (85.45%). Unfortunately, for these reasons, it was not possible to achieve a more uniform sample regarding the number of images available in each distance category.

Depending on the location of the wound on the body, the characteristics displayed also change. The flaccidity of the region, as well as its elasticity and density, are some of the factors that influence the appearance of the wounds [50–52]. In the sample used, most of the wounds were located on the trunk (51.35%). However, even within the torso, there are quite distinct areas, such as the abdomen compared to the sternum area. In turn, given that the wounds analyzed are situated on the skin, age greatly affects the activity of collagen and, consequently, elasticity [53]. Since 66.67% of the cases and 70.29% of the images are from individuals below thirty years of age, it is expected that the effect of age on elasticity did not have a significant influence. Finally, the differences in wounds due to sex, because of the arrangement of Langer's lines, skin thickness, and fat accumulation locations [52], were minimized by having a sample predominantly composed of male victims (95.09%).

In the field of scientific inquiry, it is crucial to acknowledge that AI serves not as a replacement for human expertise but as a formidable ally. The intrinsic value of human judgment and knowledge proves irreplaceable, especially in interpreting complex contexts and making well-informed decisions. Therefore, it is imperative to employ AI systems within forensic sciences as supportive instruments that enhance, rather than replace, the capabilities of forensic professionals [10, 47]. In real investigative scenarios, we can envision various applications, such as assisting pathologists in decision-making and accelerating criminal investigations. This approach ensures that the integration of AI into forensic practices augments the efficacy and accuracy of investigations, maintaining the indispensable balance between technological innovation and the nuanced understanding that only human experts can provide [47].

2.5 LIMITATIONS

This study presents significant strides in computer vision for the classification of gunshot wounds. However, while the dataset comprises images from actual forensic cases, there is a significant imbalance in the number of images available for each category of wounds, which could skew the accuracy of the model. Additionally, these images reflect a wide range of environmental and exposure conditions, contributing to variability in image quality that can affect the performance of the machine learning models. The imbalance in wound categories introduces potential selection bias, as certain types of wounds may be overrepresented, leading to models that perform well on the training data but struggle with generalization. Furthermore, the study lacks external validation, as the models were primarily validated using a dataset from a single institution located at Federal District, Brazil. Future work should focus on testing the models across diverse populations to improve robustness. To enhance model performance further, exploring resampling and training strategies, such as Data Center Interpolation with Improved Sparrow Search Algorithm [54], as well as incorporating classical feature-engineered models like Hu Moments, Haralick texture features, and local binary patterns, could offer complementary approaches for improving the robustness and accuracy of the models under real-world forensic conditions.

2.6 CONCLUSIONS

This study explored the use of fifty-nine deep neural network architectures for classifying gunshot wound images from real crime cases, focusing on distinguishing between entry and exit wounds, as well as determining the MLSD in a forensic context. The ResNet152 architecture demonstrated superior performance in wound type classification, achieving up to 86.90% accuracy, precision, recall, F1-score and specificity and an AUC of 82.09%. For MLSD, the same architecture reached an accuracy of 92.48%, precision, recall and F1-score up to 92.48%, AUC up to 94.36% and specifity up to 96.24%, although sample imbalance affected the metrics. Our findings underscore the challenges of standardizing wound images due to varying capture conditions, yet they reflect the practical realities of forensic work. This research highlights the significant potential of deep neural networks in forensic pathology, advocating for AI as a supportive tool to enhance, not replace, human expertise, ensuring the balance between technological innovation and expert judgment in forensic investigations.

AUTHOR'S CONTRIBUTIONS

Author 1 took a lead role, overseeing the conceptualization of the study, data curation, resource management, investigation, and the preparation of the original draft. Authors 2, 3, and 4 were pivotal in data curation, ensuring that the necessary data for the study's objectives were accurately collected and organized. Author 5 contributed significantly to the study's conceptualization, developed the methodology, managed the software aspects, and participated in the writing and editing processes. Author 6 enhanced the research with software development, validation, and additional investigation. Authors 7 and 8 played crucial roles in conceptualizing the project, supervising the research activities, managing the project, and refining the manuscript through their critical review and editing, ensuring both the rigor and integrity of the study.

DECLARATION OF COMPETING INTEREST

The authors have no competing interests to declare.

ACKNOWLEDGEMENTS

This work was supported by the Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF), by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq), and by the University of Brasília (UnB).

FUNDING SOURCES

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DATA AVAILABILITY STATEMENTS

The image dataset used in this paper is available at https://github.com/pedrogarciafreitas/FDCPUnBGunshotDB.git and the processed data used in the analysis were generated using the source code publicly available at https://gitlab.com/lisa-unb/leguwoi.git.

2.7 REFERENCES

[1] UNODC: Global Study on Homicide 2023. https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html Accessed 2024-03-08

[2] Wisevoter: Gun Deaths by Country 2023. https://wisevoter.com/ country-rankings/gun-deaths-by-country Accessed 2024-03-08

[3] Kleypas, D.A., Badiye, A.: Evidence Collection vol. 2, pp. 113–116 (2024). https://pubmed.ncbi.nlm.nih.gov/28722882/

[4] Shrestha, R., Kanchan, T., Krishan, K.: Gunshot Wounds Forensic Pathology, pp. 1–8 (2024). http://www.ncbi.nlm.nih.gov/pubmed/32310579

[5] Denton, J.S., Segovia, A., Filkins, J.A.: Practical pathology of gunshot wounds. Archives of pathology & laboratory medicine 130(9), 1283–9 (2006) https://doi.org/10.5858/2006-130-1283-PPOGW

[6] Di Maio, V.J.M.: Gunshot Wounds: Practical Aspects of Firearms, Ballistics, and Forensic Techniques, 3rd edn., p. 377. CRC Press(2015)

[7] Saukko, P.J., Knight, B.: Knight’s Forensic Pathology. CRC Press (2016)

[8] Adelson, L.: A microscopic study of dermal gunshot wounds. American

journal of clinical pathology 35(5), 393–402 (1961) https://doi.org/10.1093/ajcp/35.5.393

[9] Dror, I.E., Pierce, M.L.: ISO Standards Addressing Issues of Bias and Impartiality in Forensic Work. Journal of Forensic Sciences 65(3), 800–808 (2020) https://doi.org/10.1111/1556-4029.14265

[10] Vodanovi´c, M., Subaˇsi´c, M., Miloˇsevi´c, D., Gali´c, I., Brki´c, H.: Artificial intelligence in forensic medicine and forensic dentistry. Journal of Forensic Odonto-Stomatology 41(2), 30–41 (2023)

[11] Tournois, L., Trousset, V., Hatsch, D., Delabarde, T., Ludes, B., Lef`evre, T.: Artificial intelligence in the practice of forensic medicine: a scoping review. Inter- national Journal of Legal Medicine (0123456789) (2023) https://doi.org/10.1007/s00414-023-03140-9

[12] Lef`evre, T., Tournois, L.: Artificial Intelligence and Diagnostics in Medicine and Forensic Science. Diagnostics 13(23), 1–12 (2023) https://doi.org/10.3390/diagnostics13233554

[13] Wankhade, T.D., Ingale, S.W., Mohite, P.M., Bankar, N.J.: Artificial Intelligence in Forensic Medicine and Toxicology: The Future of Forensic Medicine. Cureus 14(8) (2022) https://doi.org/10.7759/cureus.28376

[14] Galante, N., Cotroneo, R., Furci, D., Lodetti, G., Casali, M.B.: Applications of artificial intelligence in forensic sciences: Current potential benefits, limitations and perspectives. International Journal of Legal Medicine, 445–458 (2022) https://doi.org/10.1007/s00414-022-02928-5

[15] Margagliotti, G., Boll´e, T.: Machine learning & forensic science. Forensic Sci- ence International 298, 138–139 (2019) https://doi.org/10.1016/j.forsciint.2019.02.045

[16] Cheng, J., Schmidt, C., Wilson, A., Wang, Z., Hao, W., Pantanowitz, J., Morris, C., Tashjian, R., Pantanowitz, L.: Artificial intelligence for human gunshot wound classification. Journal of Pathology Informatics 15, 100361 (2024) https://doi. org/10.1016/j.jpi.2023.100361

[17] Oura, P., Junno, A., Junno, J.-A.: Deep learning in forensic gunshot wound inter- pretation—a proof-of-concept study. International journal of legal medicine 135, 2101–2106 (2021) https://doi.org/10.1007/s00414-021-02566-3

[18] Molina, D.K., Dimaio, V., Cave, R.: Gunshot wounds: A review of firearm type, range, and location as pertaining to manner of death. American Journal of Forensic Medicine and Pathology 34, 366–371 (2013) https://doi.org/10.1097/PAF. 0000000000000065

[19] Marcel, S., Rodriguez, Y.: Torchvision the machine-vision package of

torch. In: Bimbo, A.D., Chang, S., Smeulders, A.W.M. (eds.) Proceedings of the 18th International Conference on Multimedia 2010, Firenze, Italy, October 25-29, 2010, pp. 1485–1488. ACM, (2010). https://doi.org/10.1145/1873951.1874254

[20] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., K¨opf, A., Yang, E.Z., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H.M., Larochelle, H., Beygelzimer, A., d'Alch´e-Buc, F., Fox, E.B., Garnett, R. (eds.) Advances in Neural Infor- mation Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pp. 8024–8035 (2019). https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html

[21] Tietz, M., Fan, T.J., Nouri, D., Bossan, B., skorch Developers: Skorch: A Scikit- learn Compatible Neural Network Library that Wraps PyTorch. (2017). https://skorch.readthedocs.io/en/stable/

[22] Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G.: Api design for machine learning software: experiences from the scikit-learn project., vol.

abs/1309.0238 (2013). http://dblp. uni-trier.de/db/journals/corr/corr1309.html#BuitinckLBPMGNPGGLVJHV13

[23] Mumuni, A., Mumuni, F.: Data augmentation: A comprehensive survey of modern approaches. Array 16, 100258 (2022) https://doi.org/10.1016/J.ARRAY.2022.100258

[24] Hataya, R., Zdenek, J., Yoshizoe, K., Nakayama, H.: Faster autoaugment: Learn- ing augmentation strategies using backpropagation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J. (eds.) Computer Vision - ECCV 2020 - 16th European Con- ference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXV. Lecture Notes in Computer Science, vol. 12370, pp. 1–16. Springer, (2020).https://doi.org/10.1007/978-3-030-58595-2 1 .

[25] Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. Informa- tion 11(2), 125 (2020) https://doi.org/10.3390/INFO11020125

[26] Batista, G.E.A.P.A., Prati, R.C., Monard, M.C.: A study of the behavior of sev- eral methods for balancing machine learning training data. SIGKDD Explor. 6(1), 20–29 (2004) https://doi.org/10.1145/1007730.1007735

[27] LemaA˜Zˇtre, G., Nogueira, F., Aridas, C.K.: Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. Journal of machine learning research 18(17), 1–5 (2017)

[28] Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Communications of the ACM 60(6), 84–90 (2017) https://doi.org/10.1145/3065386

[29] Huang, G., Liu, Z., Maaten, L., Weinberger, K.Q.: Densely connected convolu- tional networks, 2261–2269 (2017) https://doi.org/10.1109/CVPR.2017.243

[30] Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks 97, 6105–6114 (2019)

[31] Tan, M., Le, Q.V.: Efficientnetv2: Smaller models and faster training 139, 10096– 10106 (2021)

[32] Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A.C., Li, Y.: Maxvit: Multi-axis vision transformer 13684, 459–479 (2022) https://doi.org/10.1007/978-3-031-20053-3 27

[33] Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., Le, Q.V.: Mnasnet: Platform-aware neural architecture search for mobile, 2820–2828 (2019) https://doi.org/10.1109/CVPR.2019.00293

[34] Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., Chen, L.: Mobilenetv2: Inverted residuals and linear bottlenecks, 4510–4520 (2018) https://doi.org/10.

1109/CVPR.2018.00474

[35] Howard, A., Pang, R., Adam, H., Le, Q.V., Sandler, M., Chen, B., Wang, W., Chen, L., Tan, M., Chu, G., Vasudevan, V., Zhu, Y.: Searching for mobilenetv3, 1314–1324 (2019) https://doi.org/10.1109/ICCV.2019.00140

[36] Radosavovic, I., Kosaraju, R.P., Girshick, R.B., He, K., Doll´ar, P.: Designing network design spaces, 10425–10433 (2020) https://doi.org/10.1109/CVPR42600.2020.01044

[37] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition, 770–778 (2016) https://doi.org/10.1109/CVPR.2016.90

[38] Ma, N., Zhang, X., Zheng, H., Sun, J.: Shufflenet V2: practical guidelines for effi- cient CNN architecture design 11218, 122–138 (2018) https://doi.org/10.1007/978-3-030-01264-9 8

[39] Iandola, F.N., Moskewicz, M.W., Ashraf, K., Han, S., Dally, W.J., Keutzer, K.: Squeezenet: Alexnet-level accuracy with 50x fewer parameters and ¡1mb model size. CoRR abs/1602.07360 (2016)

[40] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows,

9992–10002 (2021) https://doi.org/10.1109/ICCV48922.2021.00986

[41] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2015)

[42] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021). https://openreview.net/forum?id=YicbFdNTTy

[43] Phelps, C.: Traumatic Injuries of the Brain and Its Membranes. Appleton (1897)

[44] Pircher, R., Preiß, D., Pollak, S., Thierauf-Emberger, A., Perdekamp, M.G., Geisenberger, D.: The influence of the bullet shape on the width of abrasion collars and the size of gunshot entrance holes. International Journal of Legal Medicine 131(2), 441–445 (2017) https://doi.org/10.1007/s00414-016-1501-6

[45] Rainio, J., Lalu, K., Ranta, H., Penttil¨a, A.: Morphology of experimental assault rifle skin wounds. International Journal of Legal Medicine 117(1), 19–26 (2003) https://doi.org/10.1007/s00414-002-0308-9

[46] Bartell, T.H., Mustoe, T.A.: Animal models of human tissue expansion. Plas- tic and reconstructive surgery 83(4), 681–6 (1989) https://doi.org/10.1097/00006534-198904000-00014

[47] Piraianu, A.I., Fulga, A., Musat, C.L., Ciobotaru, O.R., Poalelungi, D.G., Sta- mate, E., Ciobotaru, O., Fulga, I.: Enhancing the Evidence with Algorithms: How Artificial Intelligence Is Transforming Forensic Medicine. Diagnostics 13(18) (2023) https://doi.org/10.3390/diagnostics13182992

[48] Smith, J.C.: Investigation of Fatal Gunshot Wounds. Cleveland-Marshall Law Review 13(2), 296–308 (1964)

[49] Svensson, A., Wendel, O., Nicol, J.D.: Techniques of Crime Scene Investigation. American Elsevier Publishing Company (1965)

[50] Mahoney, P., Carr, D., Arm, R., Gibb, I., Hunt, N., Delaney, R.J.: Ballistic impacts on an anatomically correct synthetic skull with a surrogate skin/soft tis- sue layer. International Journal of Legal Medicine, 1–12 (2017) https://doi.org/ 10.1007/s00414-017-1737-9

[51] Humphrey, C., Kumaratilake, J.: Ballistics and anatomical modelling – A review.
Legal Medicine 23, 21–29 (2016) https://doi.org/10.1016/j.legalmed.2016.09.002

[52] Ridge, M.D., Wright, V.: The directional effects of skin. A bio-engineering study of skin with particular reference to Langer's lines. The Journal of investigative dermatology 46(4), 341–346 (1966) https://doi.org/10.1038/jid.1966.54

[53] Maiden, N.R., Byard, R.W.: Unpredictable tensile strength biomechanics may limit thawed cadaver use for simulant research. Australian Journal of Forensic Sciences 48(1), 54–58 (2016) https://doi.org/10.1080/00450618.2015.1025842

[54] Gu, Q., Tian, J., Li, X., Jiang, S.: A novel random forest integrated model for imbalanced data classification problem. Knowledge-Based Systems 250, 109050 (2022) https://doi.org/10.1016/j.knosys.2022.109050

CAPÍTULO 3 – DISCUSSÃO GERAL E CONCLUSÕES DA TESE

3.1 DISCUSSÃO GERAL

Este estudo marca um avanço significativo na ciência forense, pois é o primeiro conhecido a empregar ferramentas de aprendizado profundo para classificar padrões de ferimentos por arma de fogo com etiquetas extensivas, como ferimentos de entrada e saída e distâncias médico-legais, utilizando uma ampla variedade de arquiteturas de redes neurais. Único em sua abordagem, pois utiliza um conjunto considerável de imagens forenses capturadas durante investigações de cenas de crimes reais, uma conquista não alcançada por nenhum trabalho anterior. A implementação dessas metodologias possui o potencial de acelerar notavelmente o ritmo das investigações criminais reais ao oferecer análises rápidas e confiáveis que aprimoram os métodos forenses tradicionais. Além disso, a tecnologia fornece aos especialistas forenses uma segunda opinião crucial durante as fases de exame e elaboração de laudos, promovendo tomadas de decisão mais precisas e embasadas em ciência. Essas contribuições fortalecem significativamente o papel da inteligência artificial em aplicações forenses, melhorando a precisão e a eficiência das investigações criminais e auxiliando na resolução acelerada de casos.

Claramente, por razões éticas, realizar pesquisas com uma amostra completamente controlada de ferimentos por arma de fogo em cadáveres humanos, como os estudos de Phelps em 1897 [1], é impraticável. Portanto, muitos estudos ainda usam carcaças de animais [2,3,4]. No entanto, por mais que estas possam se aproximar do que é visto em humanos, existem diferenças significativas [5] que podem levar ao treinamento incorreto do modelo de aprendizado de máquina para uso futuro em casos humanos.

Neste estudo, para a classificação de imagens de ferimentos de entrada e saída por arma de fogo, o uso do ResNet152 alcançou percentuais que, com exceção dos resultados de AUC, alinham-se de perto com aqueles reportados por Cheng et al. [6], que indicou uma acurácia (*accuracy*) de 87,99%, precisão (*precision*) de 83,99%, taxa de detecção (*recall*) de 87,71%, especificidade (*specificity*) de 88,19%, pontuação F1 (*F1-score*) de 85,81% e AUC de 94,6% com uma amostra maior, composta por 2.028 imagens de ferimentos de entrada e 1.314 de saída. No entanto, as imagens utilizadas não parecem ter sido obtidas em cenas de crime, mas sim durante exames necroscópicos, o que indica um certo controle sobre as imagens capturadas. Em contrapartida, as imagens usadas em nosso estudo foram produzidas no próprio local em que o cadáver foi encontrado e sob condições de captura diversas e, por isso, correspondem melhor à realidade da rotina do trabalho pericial.

Adicionalmente, na classificação de distância de tiro, este estudo alcançou uma taxa de precisão de 92,48% usando o ResNet152. No entanto, apesar da alta precisão (*precision*), o desequilíbrio da amostra pode ter impactado negativamente outras métricas, incluindo taxa de detecção (*recall*) e pontuação F1 (*F1-score)*. Em contraste, Oura et al. [4] alcançou uma taxa de precisão de 98%. No entanto, essa maior precisão provavelmente foi facilitada pelo uso de uma amostra pequena e controlada, composta por apenas 204 imagens em um ambiente laboratorial. É importante notar que seu estudo utilizou carcaças de porcos em vez de feridas em cadáveres humanos, o que poderia afetar a aplicabilidade de seus achados em cenários forenses do mundo real.

Com isso, torna-se evidente que um dos maiores desafios no desenvolvimento de modelos de aprendizado de máquina para categorização de imagens é adquirir um banco de dados com dados científicos de alta qualidade [7]. Devido à natureza sensível do conteúdo da área, esse desafio pode ser ainda maior no campo forense. Portanto, as fotografias usadas neste estudo foram tiradas por diferentes indivíduos, sob várias condições ambientais e de exposição, com uma variedade de câmeras, nas circunstâncias rotineiras do trabalho forense em cenas de crime, onde a padronização se torna inviável. Dado isso, entende-se que um banco de dados padronizado não só estaria desalinhado com as realidades do trabalho forense, onde inúmeros fatores podem alterar as características das lesões.

O desafio em padronizar um banco de dados de imagens de ferimentos por arma de fogo também é evidente na discrepância entre o número de ferimentos de entrada (1.883) e de saída (668). Logicamente, para que ocorra um ferimento de saída, deve inevitavelmente haver um de entrada. Portanto, é esperado que o número de ferimentos de entrada supere o de saída. Além disso, mesmo com acesso a um banco de dados contendo milhões de imagens de investigações de cenas de crime, a utilidade deste conjunto de dados foi significativamente limitada. Muitas imagens estavam relacionadas a outros tipos de incidentes policiais ou apresentavam diferentes instrumentos (por exemplo, facas), tinham qualidade de imagem inadequada e enfrentavam critérios restritivos, todos os quais limitavam significativamente a obtenção de uma amostra maior e mais útil. Apesar dessas limitações, a amostra usada neste estudo reflete mais precisamente as condições que as equipes forenses encontram em cenários do mundo real e se afasta da padronização laboratorial típica de outras pesquisas [4,6,8]. Esta abordagem aumenta a relevância prática dos achados para investigações forenses do mundo real.

Por sua vez, a classificação de distância de tiro é usada em casos onde as características de tiros de curta distância e contato não estão presentes (exemplos: padrão de fumaça e tatuagem) [91] e ocorrem tipicamente em uma faixa de distância a partir de 18 polegadas-2 pés (0.45 a 0.60 metros) [10,11]. Portanto, para fins de classificação médico-legal, há uma tendência para que a maioria das lesões seja do tipo à distância, conforme a amostra obtida (85.45%). Infelizmente, por essas razões, não foi possível alcançar uma amostra mais uniforme quanto ao número de imagens disponíveis em cada categoria de distância.

Dependendo da localização da ferida no corpo, as características exibidas também mudam. A flacidez da região, bem como sua elasticidade e densidade, são alguns dos fatores que influenciam a aparência das feridas [12,13,14]. Na amostra utilizada, a maioria das feridas estava localizada no tronco (51.35%). No entanto, mesmo dentro do tronco, existem áreas bastante distintas, como o abdômen comparado à área do esterno. Além disso, considerando que as feridas analisadas estão situadas na pele, a idade afeta bastante a atividade do colágeno e, consequentemente, a elasticidade15. Como 66.67% dos casos e 70.29% das imagens são de indivíduos abaixo de trinta anos, espera-se que o efeito da idade sobre a elasticidade não tenha tido uma influência significativa. Finalmente, as diferenças nas feridas relacionadas ao sexo, por causa da disposição das linhas de Langer, espessura da pele e locais de acumulação de gordura [14], foram minimizadas por ter uma amostra predominantemente composta por vítimas do sexo masculino (95.09%).

No campo da investigação científica, é crucial reconhecer que a inteligência artificial não serve como substituto para a perícia humana, mas sim como um aliado formidável. No presente estudo, a aplicação de modelos de aprendizado de máquina para a análise de ferimentos por arma de fogo mostrou-se uma ferramenta promissora para auxiliar peritos forenses, contribuindo para a diferenciação entre ferimentos de entrada e saída e para a estimativa da distância médico-legal do tiro. No entanto, o valor intrínseco do julgamento humano e do conhecimento permanece insubstituível, especialmente na interpretação de contextos complexos e na tomada de decisões bem fundamentadas, como a análise de fatores anatômicos, o posicionamento do corpo e as circunstâncias do crime. Portanto, a utilização da inteligência artificial, como desenvolvida nesta tese, deve ser vista como um complemento, que aprimora a capacidade dos profissionais forenses, ao fornecer uma segunda opinião técnica e apoiar análises mais rápidas e precisas [7,16]. Essa integração, como demonstrado pelos resultados da pesquisa, reforça a eficácia e a precisão das investigações, sem abrir mão do entendimento matizado que apenas os especialistas humanos podem proporcionar [7].

3.2 CONCLUSÕES

Esta tese apresentou um estudo sobre a aplicação de técnicas de aprendizado de máquina na análise de ferimentos produzidos por projéteis expelidos por arma de fogo, visando melhorar a precisão e a objetividade nas investigações forenses. Foram testados 59 modelos de aprendizado de máquina, especificamente, para classificar ferimentos de entrada e saída de projéteis e estimar a distância médico-legal dos tiros. A arquitetura ResNet152 demonstrou alta precisão e confiabilidade, alcançando até 86,90% na classificação de ferimentos de entrada e saída e até 92,48% na estimativa da distância médico-legal do tiro, resultados esses significativamente superiores aos métodos tradicionais.

Este trabalho não apenas contribui para o avanço científico na aplicação de inteligência artificial na medicina forense, mas também tem o potencial de impactar positivamente a sociedade ao melhorar a eficácia das investigações criminais e, consequentemente, a justiça. A continuidade desta linha de pesquisa e o aperfeiçoamento dos modelos propostos são essenciais para consolidar a utilização de tecnologias de aprendizado de máquina na prática forense.

3.3 REFERÊNCIAS

1. Phelps C. Traumatic injuries of the brain and its membranes : with a special study of pistol-shot wounds of the head in their medico-legal and surgical relations. London: Henry Kimpton; 1898.

2. Rainio J, Lalu K, Ranta H, Penttilä A. Morphology of experimental assault rifle skin wounds. Int J Legal Med [Internet]. 2003;117:19–26. Available from: http://link.springer.com/10.1007/s00414-002-0308-9

3. Pircher R, Preiß D, Pollak S, Thierauf-Emberger A, Perdekamp MG, Geisenberger D. The influence of the bullet shape on the width of abrasion collars and the size of gunshot entrance holes. Int J Legal Med. 2017;131:441–5.

4. Oura P, Junno A, Junno JA. Deep learning in forensic gunshot wound interpretation—a proof-of-concept study. Int J Legal Med. 2021;135.

5. Bartell TH, Mustoe TA. Animal models of human tissue expansion. Plast Reconstr Surg. 1989;83:681–6.

6. Cheng J, Schmidt C, Wilson A, Wang Z, Hao W, Pantanowitz J, et al. Artificial intelligence for human gunshot wound classification. J Pathol Inform. 2024;15:0–3.

7. Junno JA, Kotiaho A, Oura P. Post-mortem computed tomography in forensic shooting distance estimation: a porcine cadaver study. BMC Res Notes. 2022;15:1–6.

8. Piraianu AI, Fulga A, Musat CL, Ciobotaru OR, Poalelungi DG, Stamate E, et al. Enhancing the Evidence with Algorithms: How Artificial Intelligence Is Transforming Forensic Medicine. Diagnostics. 2023;13.

9. Smith JC. Investigation of Fatal Gunshot Wounds. Cleveland-Marshall Law Rev [Internet]. 1964;13:296–308. Available from: https://engagedscholarship.csuohio.edu/clevstlrev

10. Denton JS, Segovia A, Filkins JA. Practical pathology of gunshot wounds. Arch Pathol Lab Med. 2006;130:1283–9.

11. Wendel ASO. Techniques of Crime Scene Investigation. 2nd ed. Elsevier; 1965.

12. Humphrey C, Kumaratilake J. Ballistics and anatomical modelling – A review. Leg Med [Internet]. 2016;23:21–9. Available from: http://dx.doi.org/10.1016/j.legalmed.2016.09.002

13. Mahoney P, Carr D, Arm R, Gibb I, Hunt N, Delaney RJ. Ballistic impacts on an anatomically correct synthetic skull with a surrogate skin/soft tissue layer. Int J Legal Med. 2017;1–12.

14. Ridge MD, Wright V. The directional effects of skin. A bio-engineering study of skin with particular reference to Langer's lines. J Invest Dermatol [Internet].

15. Maiden NR, Byard RW. Unpredictable tensile strength biomechanics may limit thawed cadaver use for simulant research. Aust J Forensic Sci. 2016;48:54–8.

16. Vodanović M, Subašić M, Milošević D, Galić I, Brkić H. Artificial intelligence in forensic medicine and forensic dentistry. J Forensic Odontostomatol. 2023;41:30–41.

CAPÍTULO 4 - PRESS RELEASE

As armas de fogo são a principal causa de mortes violentas no mundo, com o Brasil registrando mais de 47.000 fatalidades anuais. Na prática forense, distinguir entre ferimentos de entrada e saída e determinar a distância médico-legal do tiro são passos essenciais, mas frequentemente desafiadores, devido a fatores como tipo de munição, ângulo do tiro e condição do corpo. Métodos tradicionais dependem da interpretação de peritos, que pode ser subjetiva e inconsistente. A equipe construiu uma base de dados abrangente com 2.551 imagens de cenas de crimes reais, incluindo 1.883 ferimentos de entrada e 668 de saída, coletadas pela Polícia Civil do Distrito Federal entre 2012 e 2022. Utilizando a arquitetura ResNet152, os modelos de inteligência artificial alcançaram uma precisão de até 86,90% na classificação de tipos de ferimentos e 92,48% na determinação da distância médico-legal do tiro. Os modelos de inteligência artificial se mostraram ferramentas úteis que podem auxiliar os peritos em seus trabalhos de análise de ferimentos produzidos por arma de fogo. A pesquisa destaca um dos primeiros usos de aprendizado profundo na medicina forense, demonstrando seu potencial para melhorar a precisão e a rapidez das investigações criminais. A utilização de inteligência artificial pode acelerar a resolução de crimes, proporcionando análises rápidas e confiáveis que complementam os métodos forenses tradicionais. Fornece aos peritos forenses uma segunda opinião crucial durante as fases de exame e elaboração de laudos, promovendo decisões mais precisas e informadas. Contribui para a justiça, garantindo que as investigações sejam baseadas em critérios científicos rigorosos, reduzindo a subjetividade e aumentando a imparcialidade nas conclusões forenses. Essa pesquisa representa um passo significativo na integração da inteligência artificial nas práticas forenses, promovendo um equilíbrio entre inovação tecnológica e a expertise humana.

APÊNDICE – ARTIGO PUBLICADO

International Journal of Legal Medicine
https://doi.org/10.1007/s00414-024-03355-4

ORIGINAL ARTICLE

Deep learning-based human gunshot wounds classification

Renato Queiroz Nogueira Lira[1] · Luana Geovana Motta de Sousa[1] · Maisa Luana Memoria Pinho[1] · Renan Cesar Pinto da Silva Andrade de Lima[1] · Pedro Garcia Freitas[2] · Bruno Scholles Soares Dias[3] · Andreia Cristina Breda de Souza[4] · André Ferreira Leite[1]

Received: 10 June 2024 / Accepted: 18 October 2024

Abstract
In this paper, we present a forensic perspective on classifying gunshot wound patterns using Deep Learning (DL). Although DL has revolutionized various medical specialties, such as automating tasks like medical image classification, its applications in forensic contexts have been limited despite the inherently visual nature of the field. This study investigates the application of DL techniques (59 architectures) to classify gunshot wounds in a forensic context, focusing on distinguishing between entry and exit wounds and determining the Medical-Legal Shooting Distance (MLSD), which classifies wounds as contact, close range, or distant, based on digital images from real crime scene cases. A comprehensive database was constructed with 2,551 images, including 1,883 entries and 668 exit wounds. The ResNet152 architecture demonstrated superior performance in both entry and exit wound classification and MLSD categorization. For the first task, achieved accuracy of 86.90% and an AUC of 82.09%. For MLSD, the ResNet152 showed an accuracy of 92.48% and AUC up to 94.36%, though sample imbalance affected the metrics. Our findings underscore the challenges of standardizing wound images due to varying capture conditions but reflect the practical realities of forensic work. This research highlights the significant potential of DL in enhancing forensic pathology practices, advocating for Artificial Intelligence (AI) as a supportive tool to complement human expertise in forensic investigations.

Keywords Forensic medicine · Gunshot wounds · Artificial intelligence · Deep learning

Introduction

Firearms correspond to one of the most used instruments worldwide for committing homicides, being related to 75% of cases in the Americas in 2021 [1]. In Brazil, the country with the highest number of firearm-related fatalities, there are more than 47,000 deaths attributed to firearms annually [2]. Given these high rates, it is common in forensic fieldwork to encounter gunshot wounds during external body examinations, whether at the crime scene or during an autopsy. When these are found, among various procedures, it is necessary to distinguish between entry and exit wounds and to classify the entry wounds according to the distance from which the shot was fired [3, 4].

Luana Geovana Motta de Sousa, Maisa Luana Memoria Pinho, and Renan Cesar Pinto da Silva Andrade de Lima contributed equally to this work.

Extended author information available on the last page of the article

Some characteristics, such as shape, size, and margins, allow for relatively easy differentiation between entry and exit wounds [5]. However, these characteristics are not always clearly visible, or they may be partially absent or even overlapping, complicating the accurate classification. Additionally, other factors can influence wound production, including the caliber and type of ammunition, the angle and speed of the shot, barriers encountered by the projectile, and the region of the body impacted [6, 7].

Similarly, the classification of the shot distance can be influenced by the same aforementioned factors [6]. This distance has a medico-legal classification, which is based on the characteristics presented by the injury itself and its naked eye appearances. These categories include: distant (or medium-distance) wound, close (or near) range wound and contact wound [7, 8]. Some authors also introduce additional categories between close-range and contact, such as near-contact and loose-contact [6]. Moreover, it is important to emphasize that this is a medico-legal classification based on the traces found on the victim, not necessarily reflecting the actual event, and serves as an indication or suggestion of what occurred at the crime scene [7].

Regardless, for forensic investigation purposes, any additional information can be crucial in aiding the elucidation of what happened, how it happened, and who was involved. However, to ensure impartiality and justice, it is important that the evidence produced is based on and grounded in scientific criteria [9].

In turn, in recent years, there has been an increase in the research and perspectives of use of AI with Machine Learning (ML) across various sectors, including forensic medicine [10–15]. In the pursuit of faster and more robust crime resolution, image recognition and classification have seen significant advancements, including initiating studies on the classification of firearm injuries [16, 17]. Despite these developments contributing to the enhancement of forensic methodologies and the efficiency of criminal investigations, and being the studies that initiated research in the area, they involved the use of only a few images and were based on injuries in pig carcasses [17] or, when trained with autopsy photographs of wounds on human bodies, did not analyze the MLSD aspect [16].

The current study aimed to develop a ML model that can analyze and classify photographs of real crime scene forensic cases of gunshot wounds in relation to determining whether they are entry or exit wounds and to classify MLSD of the gunshot wounds.

Material and methods

Figure 1 illustrated the framework employed during the development of this work. It includes three main stages responsible for preprocessing the captured photographies from the crime scene, classifying the preprocessed images using a deep neural network, and finally estimating the classification according to a specific forensic task. The following subsections details each stage.

Data acquisition

In this study, photographic records of homicide and suicide cases examined by crime scene forensic experts and reports by medical examiners from incidents attended by the Federal District Civil Police (PCDF) (Brasília, Federal District, Brazil) from the January of 2012 until March of 2022 were used. From the selected time period, a total of 2,012,584 photographs were stored. This dataset was obtained with the proper authorization of PCDF and the Committee for Ethics in Research of the University of Brasília approved this research (Protocol number 54418221.9.0000.0030).

The images and relevant data related to the wounds, such as age, sex, anatomical location, distance, and wound category, the following inclusion criteria were established:

1. the victim was examined and photographed in the crime scene by forensic experts;
2. the body did not show advanced signs of decomposition (e.g., greenish discoloration of the skin of the anterior abdominal wall in the right iliac fossa region);
3. the photos from the scene examination were in focus and with sufficient visualization of the wounds;
4. the victim was identified, due to the need for additional data - sex and age at the time of death;
5. there were bullets stored in the bodies;
6. the bullets had been recovered during the autopsy at the Legal Medical Institute (IML/PCDF);
7. the bullets recovered at the IML had been sent to the Institute of Criminalistics (IC/PCDF);
8. it was possible to determine the caliber(s) of the bullet(s) at the Forensic Ballistics Section (SBF/IC/PCDF);
9. only one caliber involved per victim was determined at SBF;
10. the respective autopsy reports, made by experienced forensics pathologists, indicates which anatomical location, shot distance (distant, close or contact) and category (entry or exit) the photographed wound corresponds to. These reports correspond to the gold standard.

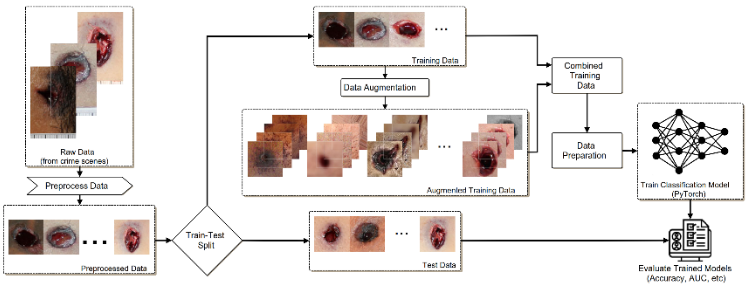


Fig. 1 Developed framework for data augmentation, classification modeling, and used to evaluate the proposed methodology

Table 1 Cases and images per sex and wound type (entry or exit)

Sex	# Cases (%)	# Images (%)	# Entries (%)	# Exits (%)
Female	24 (4.91)	114 (4.47)	85 (4.51)	29 (4.34)
Male	465 (95.09)	2437 (95.53)	1798 (95.49)	639 (95.66)
Total	489 (100)	2551 (100)	1883 (100)	668 (100)

These criteria were designed based on the working protocol guidelines of the PCDF and, due to the use of this database, several wounds and cases were excluded from this study, reducing the available sample. However, these data may be used for further analyses, potentially involving caliber as one of the factors to be studied.

Data preprocessing

The images retrieved from these cases were then manually cropped to isolate the wounds in a square format. This manual processing was handled using the GNU Image Manipulation Program (GIMP) image processing software, resulting in a preprocessed dataset of 2,551 images of gunshot wounds. These preprocessed images were then categorized according to wound type and MLSD. The wound types were categorized into entry and exit wounds, corresponding to 1,883 entry and 668 exit wounds. Table 1 details the wound types according to the number of cases, associated images, and sex. Moreover, considering only the subset of entry wound images, we obtained 1,609 from distance shots, 232 from close range, and 42 from contact shots as per the MLSD categorization, as depicted in Table 2.

As observed in Tables 1, 2 and 3, the dataset under study exhibits a significant imbalance. Such imbalance is typical in forensic datasets, where certain types of wounds or shooting distances are often underrepresented compared to others [18]. This imbalance stems from real-world crime cases, where certain types of wounds are rare or sensitive, leading to limited sample availability. In response to this, techniques were applied during data preprocessing when constructing the training pipeline for the models under investigation.

Learning-based classification

Considering the curated data as described in the previous section, we identified that the recognition of forensic characteristics of gunshot wound images can be expressed as an image classification problem. More specifically, the identification of the wound type is a binary classification problem (i.e., entry or exit), while the categorization of the MLSD is a multiclass classification problem, as can be inferred from Table 2. In other words, both are Computer Vision (CV) problems. Therefore, this study tested the potential of a series of DL models established in the literature for CV problems.

The neural networks used in this study were achieved from TorchVision [19], popular library that is part of the PyTorch [20] ecosystem. TorchVision provides tools and utilities specifically designed for computer vision tasks, such as state-of-the-art architectures and its pretrained models. From this library, fifty-nine state-of-the-art DL architectures were utilized to analyze forensic images for the differentiation of entry and exit wounds as well as the determination of MLSD. These models are listed in Tables 4 and 5. Additionally, we employed a supplementary abstraction software layer using Skorch [21], a library that provides a high-level interface for working with PyTorch models in a scikit-learn-like way. This additional layer was included to add Scikit-Learn [22] compatibility and make it easy to use PyTorch models with the scikit-learn ecosystem. This allows our software implementation to leverage tools and utilities from Scikit-Learn, such as grid search, cross-validation, hyperparameter tuning, and pipelines, with PyTorch models. Thus, with the use of Skorch, our implementation benefits from several desirable properties in terms of software engineering, such as decoupling, maintainability, efficiency, correctness, reusability, testability, etc. Our concern on these properties was to ensure reliability in conducting our experiments.

As can be noticed from Tables 1, 2 and 3, the studied dataset is highly imbalanced. This imbalance is usually common in forensic datasets, where certain types of wounds or shooting distances may be underrepresented compared to others. The data imbalance emerges from the real-world crime cases, where some type of wounds can often be lim-

Table 2 Entry wounds images per medico-legal distance classification

Distance	# Images (%)
Distant	1609 (85.45)
Close range	232 (12.32)
Contact	42 (2.23)
Total	1883 (100)

Table 3 Number of wound images per anatomical location on the body

Anatomical location	# Images (%)
Head and neck	548 (21.48)
Trunk	1310 (51.35)
Upper limbs	503 (19.72)
Lower limbs	190 (7.45)
Total	2551 (100)

Table 4 Performance metrics for different neural network architectures using hold-out validation to predict the wound type

Architecture	Variant	ACC	Precision			Recall			F1-Score			AUC			Specificity		
			$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$
Alexnet [28]		0.765	0.695	0.765	0.780	0.716	0.765	0.765	0.703	0.765	0.771	0.716	0.716	0.716	0.716	0.765	0.765
DenseNet	121	0.736	0.640	0.736	0.727	0.629	0.736	0.736	0.634	0.736	0.731	0.629	0.629	0.629	0.629	0.736	0.736
[29]	161	0.753	0.659	0.753	0.737	0.635	0.753	0.753	0.644	0.753	0.742	0.635	0.635	0.635	0.635	0.753	0.753
	169	0.777	0.700	0.777	0.768	0.679	0.777	0.777	0.688	0.777	0.772	0.679	0.679	0.679	0.679	0.777	0.777
	201	0.744	0.663	0.744	0.751	0.671	0.744	0.744	0.667	0.744	0.747	0.671	0.671	0.671	0.671	0.744	0.744
EfficientNet	B0	0.744	0.646	0.744	0.727	0.624	0.744	0.744	0.632	0.744	0.734	0.624	0.624	0.624	0.624	0.744	0.744
[30, 31]	B1	0.734	0.627	0.734	0.713	0.603	0.734	0.734	0.610	0.734	0.720	0.603	0.603	0.603	0.603	0.734	0.734
	B2	0.750	0.656	0.750	0.734	0.631	0.750	0.750	0.639	0.750	0.740	0.631	0.631	0.631	0.631	0.750	0.750
	B3	0.757	0.659	0.757	0.728	0.604	0.757	0.757	0.615	0.757	0.732	0.604	0.604	0.604	0.604	0.757	0.757
	B4	0.715	0.588	0.715	0.684	0.566	0.715	0.715	0.569	0.715	0.695	0.566	0.566	0.566	0.566	0.715	0.715
	B5	0.744	0.608	0.744	0.686	0.535	0.744	0.744	0.517	0.744	0.683	0.535	0.535	0.535	0.535	0.744	0.744
	B6	0.759	0.667	0.759	0.740	0.633	0.759	0.759	0.644	0.759	0.745	0.633	0.633	0.633	0.633	0.759	0.759
	B7	0.750	0.375	0.750	0.563	0.500	0.750	0.750	0.429	0.750	0.644	0.500	0.500	0.500	0.500	0.750	0.750
	V2L	0.782	0.710	0.782	0.760	0.638	0.782	0.782	0.654	0.782	0.759	0.638	0.638	0.638	0.638	0.782	0.782
	V2M	0.726	0.602	0.726	0.693	0.572	0.726	0.726	0.577	0.726	0.703	0.572	0.572	0.572	0.572	0.726	0.726
	V2S	0.721	0.621	0.721	0.714	0.614	0.721	0.721	0.617	0.721	0.717	0.614	0.614	0.614	0.614	0.721	0.721
MaxVit [32]		0.780	0.703	0.780	0.761	0.653	0.780	0.780	0.668	0.780	0.764	0.653	0.653	0.653	0.653	0.780	0.780
MNASNet	5	0.705	0.589	0.705	0.688	0.578	0.705	0.705	0.582	0.705	0.695	0.578	0.578	0.578	0.578	0.705	0.705
[33]	75	0.717	0.590	0.717	0.686	0.567	0.717	0.717	0.571	0.717	0.696	0.567	0.567	0.567	0.567	0.717	0.717
	10	0.682	0.570	0.682	0.677	0.568	0.682	0.682	0.569	0.682	0.680	0.568	0.568	0.568	0.568	0.682	0.682
	13	0.746	0.659	0.746	0.744	0.656	0.746	0.746	0.657	0.746	0.745	0.656	0.656	0.656	0.656	0.746	0.746
MobileNet	V2	0.719	0.609	0.719	0.702	0.593	0.719	0.719	0.598	0.719	0.709	0.593	0.593	0.593	0.593	0.719	0.719
[34, 35]	V3Large	0.690	0.566	0.690	0.672	0.557	0.690	0.690	0.559	0.690	0.680	0.557	0.557	0.557	0.557	0.690	0.690
	V3Small	0.703	0.585	0.703	0.685	0.574	0.703	0.703	0.577	0.703	0.693	0.574	0.574	0.574	0.574	0.703	0.703
Regnet	X16GF	0.690	0.534	0.690	0.648	0.524	0.690	0.690	0.521	0.690	0.664	0.524	0.524	0.524	0.524	0.690	0.690
[36]	X32GF	0.724	0.580	0.724	0.675	0.546	0.724	0.724	0.544	0.724	0.687	0.546	0.546	0.546	0.546	0.724	0.724
	X400MF	0.663	0.528	0.663	0.646	0.525	0.663	0.663	0.526	0.663	0.654	0.525	0.525	0.525	0.525	0.663	0.663
	X800MF	0.703	0.563	0.703	0.667	0.546	0.703	0.703	0.547	0.703	0.680	0.546	0.546	0.546	0.546	0.703	0.703
	X8GF	0.699	0.558	0.699	0.664	0.543	0.699	0.699	0.544	0.699	0.677	0.543	0.543	0.543	0.543	0.699	0.699
	Y16GF	0.711	0.571	0.711	0.672	0.549	0.711	0.711	0.549	0.711	0.685	0.549	0.549	0.549	0.549	0.711	0.711

Table 4 continued

Architecture	Variant	ACC	Precision			Recall			F1-Score			AUC			Specificity		
			$\mathbb{M}$	$\mathbb{m}$	$\mathbb{W}$	$\mathbb{M}$	$\mathbb{m}$	$\mathbb{W}$	$\mathbb{M}$	$\mathbb{m}$	$\mathbb{W}$	$\mathbb{M}$	$\mathbb{m}$	$\mathbb{W}$	$\mathbb{M}$	$\mathbb{m}$	$\mathbb{W}$
	Y32GF	0.671	0.533	0.671	0.648	0.528	0.671	0.671	0.528	0.671	0.659	0.528	0.528	0.528	0.528	0.671	0.671
	Y400MF	0.674	0.540	0.674	0.654	0.535	0.674	0.674	0.536	0.674	0.663	0.535	0.535	0.535	0.535	0.674	0.674
	Y800MF	0.703	0.565	0.703	0.669	0.549	0.703	0.703	0.550	0.703	0.681	0.549	0.549	0.549	0.549	0.703	0.703
	Y8GF	0.703	0.563	0.703	0.667	0.546	0.703	0.703	0.547	0.703	0.680	0.546	0.546	0.546	0.546	0.703	0.703
ResNet	152	**0.869**	**0.827**	**0.869**	**0.868**	**0.821**	**0.869**	**0.869**	**0.824**	**0.869**	**0.869**	**0.821**	**0.821**	**0.821**	**0.821**	**0.869**	**0.869**
[37]	101	0.863	0.822	0.863	0.860	0.803	0.863	0.863	0.811	0.863	0.861	0.803	0.803	0.803	0.803	0.863	0.863
	50	0.850	0.802	0.850	0.848	0.792	0.850	0.850	0.797	0.850	0.849	0.792	0.792	0.792	0.792	0.850	0.850
	34	0.850	0.799	0.850	0.852	0.806	0.850	0.850	0.802	0.850	0.851	0.806	0.806	0.806	0.806	0.850	0.850
	18	0.817	0.758	0.817	0.810	0.734	0.817	0.817	0.744	0.817	0.812	0.734	0.734	0.734	0.734	0.817	0.817
ShuffleNet	V2x05	0.730	0.628	0.730	0.717	0.614	0.730	0.730	0.619	0.730	0.722	0.614	0.614	0.614	0.614	0.730	0.730
[38]	V2x10	0.701	0.533	0.701	0.646	0.520	0.701	0.701	0.512	0.701	0.664	0.520	0.520	0.520	0.520	0.701	0.701
	V2x15	0.711	0.580	0.711	0.679	0.560	0.711	0.711	0.563	0.711	0.690	0.560	0.560	0.560	0.560	0.711	0.711
	V2x20	0.692	0.551	0.692	0.660	0.539	0.692	0.692	0.540	0.692	0.672	0.539	0.539	0.539	0.539	0.692	0.692
SqueezeNet	10	0.750	0.375	0.750	0.563	0.500	0.750	0.750	0.429	0.750	0.644	0.500	0.500	0.500	0.500	0.750	0.750
[39]	11	0.750	0.375	0.750	0.563	0.500	0.750	0.750	0.429	0.750	0.644	0.500	0.500	0.500	0.500	0.750	0.750
SwinT	B	0.728	0.614	0.728	0.703	0.591	0.728	0.728	0.597	0.728	0.712	0.591	0.591	0.591	0.591	0.728	0.728
[40]	S	0.671	0.536	0.671	0.650	0.531	0.671	0.671	0.531	0.671	0.660	0.531	0.531	0.531	0.531	0.671	0.671
	T	0.713	0.594	0.713	0.691	0.578	0.713	0.713	0.583	0.713	0.700	0.578	0.578	0.578	0.578	0.713	0.713
	V2B	0.750	0.375	0.750	0.563	0.500	0.750	0.750	0.429	0.750	0.644	0.500	0.500	0.500	0.500	0.750	0.750
	V2S	0.713	0.557	0.713	0.661	0.533	0.713	0.713	0.528	0.713	0.676	0.533	0.533	0.533	0.533	0.713	0.713
	V2T	0.750	0.375	0.750	0.563	0.500	0.750	0.750	0.429	0.750	0.644	0.500	0.500	0.500	0.500	0.750	0.750
VGG	11	0.763	0.673	0.763	0.742	0.633	0.763	0.763	0.645	0.763	0.748	0.633	0.633	0.633	0.633	0.763	0.763
[41]	13	0.790	0.720	0.790	0.774	0.671	0.790	0.790	0.687	0.790	0.777	0.671	0.671	0.671	0.671	0.790	0.790
	16	0.782	0.706	0.782	0.772	0.685	0.782	0.782	0.694	0.782	0.776	0.685	0.685	0.685	0.685	0.782	0.782
	19	0.777	0.700	0.777	0.768	0.679	0.777	0.777	0.688	0.777	0.772	0.679	0.679	0.679	0.679	0.777	0.777
ViT	B16	0.713	0.587	0.713	0.684	0.567	0.713	0.713	0.571	0.713	0.695	0.567	0.567	0.567	0.567	0.713	0.713
[42]	B32	0.674	0.559	0.674	0.669	0.557	0.674	0.674	0.558	0.674	0.671	0.557	0.557	0.557	0.557	0.674	0.674
	L16	0.678	0.544	0.678	0.656	0.538	0.678	0.678	0.539	0.678	0.666	0.538	0.538	0.538	0.538	0.678	0.678
	L32	0.692	0.563	0.692	0.669	0.553	0.692	0.692	0.555	0.692	0.679	0.553	0.553	0.553	0.553	0.692	0.692

Considering each metric separately, the best results in terms of maximum absolute value are **boldfaced**. Symbols $\mathbb{M}$, $\mathbb{m}$, and $\mathbb{W}$ illustrates macro, micro, and weighted, respectively

Table 5 Performance metrics for different neural network architectures using hold-out validation to predict the medico-legal shooting distance

Architecture	Variant	ACC	Precision			Recall			F1-Score			AUC			Specificity		
			$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$
Alexnet [28]		0.889	0.537	0.889	0.862	0.378	0.889	0.889	0.395	0.889	0.856	0.541	0.916	0.555	0.703	0.944	0.222
DenseNet	121	0.877	0.422	0.877	0.825	0.356	0.877	0.877	0.359	0.877	0.841	0.521	0.908	0.527	0.685	0.939	0.177
[29]	161	0.894	0.824	0.894	0.871	0.465	0.894	0.894	0.528	0.894	0.863	0.589	0.921	0.569	0.712	0.947	0.243
	169	0.883	0.467	0.883	0.820	0.388	0.883	0.883	0.398	0.883	0.844	0.548	0.912	0.563	0.708	0.942	0.242
	201	0.902	0.634	0.902	0.883	0.482	0.902	0.902	0.527	0.902	0.886	0.627	0.927	0.659	0.772	0.951	0.415
EfficientNet	B0	0.877	0.758	0.877	0.857	0.438	0.877	0.877	0.483	0.877	0.858	0.583	0.908	0.591	0.728	0.939	0.305
[30, 31]	B1	0.886	0.670	0.886	0.852	0.462	0.886	0.886	0.508	0.886	0.857	0.585	0.914	0.564	0.709	0.943	0.242
	B2	0.864	0.393	0.864	0.817	0.360	0.864	0.864	0.364	0.864	0.836	0.523	0.898	0.531	0.687	0.932	0.198
	B3	0.900	0.560	0.900	0.873	0.400	0.900	0.900	0.424	0.900	0.869	0.561	0.925	0.582	0.722	0.950	0.265
	B4	0.883	0.450	0.883	0.837	0.376	0.883	0.883	0.386	0.883	0.851	0.539	0.912	0.552	0.701	0.942	0.221
	B5	0.869	0.357	0.869	0.806	0.345	0.869	0.869	0.340	0.869	0.833	0.513	0.902	0.523	0.667	0.943	0.114
	B6	0.886	0.295	0.886	0.785	0.333	0.886	0.886	0.313	0.886	0.832	0.500	0.914	0.500	0.667	0.943	0.114
	B7	0.886	0.463	0.886	0.834	0.342	0.886	0.886	0.332	0.886	0.837	0.508	0.914	0.511	0.674	0.943	0.136
	V2L	0.880	0.617	0.880	0.854	0.448	0.880	0.880	0.488	0.880	0.861	0.588	0.910	0.593	0.729	0.940	0.306
	V2M	0.883	0.614	0.883	0.846	0.414	0.883	0.883	0.448	0.883	0.852	0.558	0.912	0.552	0.701	0.942	0.221
	V2S	0.875	0.592	0.875	0.843	0.428	0.875	0.875	0.465	0.875	0.852	0.570	0.906	0.569	0.712	0.937	0.263
MaxVit [32]		0.880	0.765	0.880	0.859	0.486	0.880	0.880	0.549	0.880	0.862	0.607	0.910	0.593	0.729	0.940	0.305
MNASNet	5	0.844	0.344	0.844	0.806	0.344	0.844	0.844	0.342	0.844	0.824	0.519	0.883	0.541	0.694	0.922	0.238
[33]	75	0.872	0.521	0.872	0.843	0.436	0.872	0.872	0.462	0.872	0.853	0.577	0.904	0.578	0.719	0.936	0.284
	10	0.872	0.566	0.872	0.837	0.418	0.872	0.872	0.451	0.872	0.850	0.568	0.904	0.578	0.719	0.936	0.283
	13	0.869	0.409	0.869	0.824	0.371	0.869	0.869	0.378	0.869	0.842	0.533	0.902	0.544	0.696	0.934	0.219
MobileNet	V2	0.886	0.424	0.886	0.831	0.360	0.886	0.886	0.361	0.886	0.848	0.531	0.914	0.553	0.702	0.943	0.221
[34, 35]	V3Large	0.872	0.411	0.872	0.829	0.381	0.872	0.872	0.388	0.872	0.847	0.546	0.904	0.567	0.711	0.936	0.262
	V3Small	0.891	0.548	0.891	0.862	0.362	0.891	0.891	0.367	0.891	0.849	0.526	0.918	0.535	0.690	0.946	0.179
Regnet	X16GF	0.855	0.294	0.855	0.781	0.322	0.855	0.855	0.307	0.855	0.817	0.489	0.891	0.483	0.655	0.928	0.111
[36]	X32GF	0.872	0.370	0.872	0.808	0.346	0.872	0.872	0.341	0.872	0.834	0.511	0.904	0.514	0.676	0.936	0.155
	X400MF	0.844	0.358	0.844	0.808	0.353	0.844	0.844	0.353	0.844	0.825	0.520	0.883	0.530	0.687	0.922	0.217
	X800MF	0.855	0.487	0.855	0.804	0.377	0.855	0.855	0.396	0.855	0.826	0.527	0.891	0.515	0.677	0.928	0.175
	X8GF	0.850	0.349	0.850	0.802	0.346	0.850	0.850	0.344	0.850	0.824	0.510	0.887	0.512	0.675	0.925	0.174
	Y16GF	0.875	0.398	0.875	0.820	0.355	0.875	0.875	0.357	0.875	0.840	0.523	0.906	0.537	0.691	0.937	0.199

Table 5 continued

Architecture	Variant	ACC	Precision			Recall			F1-Score			AUC			Specificity		
			$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$	$\mathbb{M}$	m	$\mathbb{W}$
	Y32GF	0.864	0.659	0.864	0.814	0.380	0.864	0.864	0.407	0.864	0.830	0.526	0.898	0.509	0.673	0.932	0.154
	Y400MF	0.869	0.333	0.869	0.798	0.336	0.869	0.869	0.326	0.869	0.829	0.505	0.902	0.512	0.675	0.934	0.155
	Y800MF	0.880	0.531	0.880	0.819	0.387	0.880	0.880	0.404	0.880	0.840	0.536	0.910	0.529	0.686	0.940	0.178
	Y8GF	0.861	0.324	0.861	0.794	0.333	0.861	0.861	0.323	0.861	0.825	0.502	0.895	0.507	0.672	0.930	0.154
ResNet	152	**0.925**	0.810	**0.925**	**0.917**	**0.637**	**0.925**	**0.925**	**0.701**	**0.925**	**0.917**	**0.733**	**0.944**	**0.745**	**0.830**	**0.962**	0.566
[37]	101	0.914	0.813	0.914	0.903	0.571	0.914	0.914	0.642	0.914	0.896	0.674	0.935	0.665	0.776	0.957	0.416
	50	0.911	**0.862**	0.911	0.898	0.617	0.911	0.911	0.696	0.911	0.896	0.696	0.933	0.663	0.775	0.955	0.415
	34	0.897	0.732	0.897	0.885	0.591	0.897	0.897	0.643	0.897	0.889	0.691	0.923	0.687	0.792	0.949	0.478
	18	0.914	0.557	0.914	0.887	0.449	0.914	0.914	0.479	0.914	0.892	0.613	0.935	0.665	0.776	0.957	0.415
ShuffleNet	V2x05	0.855	0.355	0.855	0.804	0.348	0.855	0.855	0.346	0.855	0.827	0.512	0.891	0.515	0.677	0.928	0.175
[38]	V2x10	0.877	0.411	0.877	0.830	0.374	0.877	0.877	0.380	0.877	0.849	0.544	0.908	0.570	0.713	0.939	0.263
	V2x15	0.869	0.337	0.869	0.797	0.336	0.869	0.869	0.326	0.869	0.828	0.502	0.902	0.501	0.668	0.934	0.134
	V2x20	0.880	0.379	0.880	0.810	0.340	0.880	0.880	0.330	0.880	0.834	0.506	0.910	0.508	0.672	0.940	0.135
SqueezeNet	10	0.095	0.032	0.095	0.009	0.333	0.095	0.095	0.058	0.095	0.016	0.500	0.321	0.500	0.667	0.547	**0.905**
[39]	11	0.886	0.295	0.886	0.785	0.333	0.886	0.886	0.313	0.886	0.832	0.500	0.914	0.500	0.667	0.943	0.114
SwinT	B	0.861	0.354	0.861	0.809	0.341	0.861	0.861	0.339	0.861	0.831	0.517	0.895	0.540	0.693	0.930	0.219
[40]	S	0.855	0.522	0.855	0.816	0.395	0.855	0.855	0.420	0.855	0.831	0.539	0.891	0.526	0.684	0.928	0.197
	T	0.872	0.727	0.872	0.840	0.410	0.872	0.872	0.449	0.872	0.845	0.553	0.904	0.546	0.697	0.936	0.219
	V2B	0.886	0.295	0.886	0.785	0.333	0.886	0.886	0.313	0.886	0.832	0.500	0.914	0.500	0.667	0.943	0.114
	V2S	0.861	0.405	0.861	0.826	0.376	0.861	0.861	0.384	0.861	0.841	0.542	0.895	0.561	0.707	0.930	0.262
	V2T	0.872	0.419	0.872	0.826	0.363	0.872	0.872	0.369	0.872	0.841	0.527	0.904	0.535	0.690	0.936	0.199
VGG	11	0.905	0.666	0.905	0.889	0.466	0.905	0.905	0.514	0.905	0.884	0.613	0.929	0.639	0.759	0.953	0.372
[41]	13	0.875	0.542	0.875	0.845	0.437	0.875	0.875	0.467	0.875	0.856	0.582	0.906	0.590	0.727	0.937	0.305
	16	0.864	0.392	0.864	0.825	0.378	0.864	0.864	0.381	0.864	0.843	0.546	0.898	0.573	0.715	0.932	0.282
	19	0.883	0.578	0.883	0.852	0.431	0.883	0.883	0.466	0.883	0.858	0.574	0.912	0.573	0.716	0.942	0.264
ViT	B16	0.838	0.342	0.838	0.801	0.342	0.838	0.838	0.340	0.838	0.819	0.510	0.879	0.517	0.678	0.919	0.195
[42]	B32	0.875	0.425	0.875	0.837	0.390	0.875	0.875	0.400	0.875	0.853	0.558	0.906	0.590	0.727	0.937	0.305
	L16	0.869	0.708	0.869	0.832	0.400	0.869	0.869	0.436	0.869	0.840	0.544	0.902	0.533	0.689	0.934	0.198
	L32	0.847	0.376	0.847	0.817	0.371	0.847	0.847	0.372	0.847	0.831	0.537	0.885	0.553	0.702	0.923	0.260

Considering each metric separately, the best results in terms of maximum absolute value are **boldfaced**. Symbols $\mathbb{M}$, m, and $\mathbb{W}$ illustrates macro, micro, and weighted, respectively ited in samples due to the sensitivity and rarity of such data. Taking that into account, we implemented some techniques in the data preprocessing when creating the training pipeline for the investigated models.

The main technique adopted in the scope of training data preparation was the data augmentation. The term “data augmentation” allude to procedures for building iterative sampling and optimization algorithms via the incorporation of unobserved data or latent variables. Its main goal is to increase the volume, quality and diversity of training data. In terms of the distribution of the target dataset, the greater the size, diversity, and representativeness of the training data, the more effectively the DL model performs on unseen data. It introduces variability into the training data, which helps prevent the model from memorizing specific features of the training set and overfitting. Augmentation techniques such as rotation, flipping, and changes in brightness or contrast help the model become more robust to these variations, ensuring accurate identification across different scenarios. On one hand, the variety of training samples should be ample enough for the model to manage various deviations in image appearance and even noisy target instances. On the other hand, this encourages the model to learn more generalizable features that are applicable to a wider range of forensic cases.

In the context of this work, data augmentation helps in artificially increasing the size of the dataset, which is crucial for training deep learning models effectively. These techniques can help to balance the dataset by creating synthetic examples of the minority classes, improving the model’s ability to recognize and classify all types of wounds and distances. However, there is a wide variety of augmentation techniques available in the literature [23], and choosing the appropriate set of augmentations for the problem can be quite challenging. Specifically, we use 13 augmentation techniques, including variants of rotation, flipping, cropping and resizing, brightness and contrast adjustments, Gaussian noise, color jitter (changes in hue and saturation), elastic transformations, and grayscale conversion. Examples results of these augmentation techniques are depicted in Fig. 2.

The augmentation illustrated in Fig. 2 were chosen to handle variations that may occur in real-world forensic cases. Rotation, flipping, and further geometric transformations such as SR can help the model learn to recognize wounds from different angles, making the model more adaptable to variations in the size and anatomical location of wounds. Elastic and affine transformations can mimic the natural variability in skin and tissue appearance, making the model more robust to such changes. Varying the brightness and contrast can help the model become robust to different lighting conditions that may be encountered in real-world forensic images taken from different crime scenes. Some noise and color augmentations such as ISOnoise, RBC, RGBShift, and Gaussian noise can make the model more resilient to image quality variations, such as those caused by different imaging equipment or environmental conditions. These augmentation techniques were discovered with AutoAulbument [24], an AutoML tool that automatically searches for the best augmentation policies based on training data. Using the its classification model, AutoAlbument provides a complete ready-to-use configuration for the augmentation pipeline. After discovered these augmentation techniques set, they are then incorporated in the training procedure with the Albumentations framework [25].

In practice, when using an augmentation library, it is essential that the quality of the synthetic samples does not deteriorate to a point of impairing performance on normal wound images. To avoid it, we combine synthetic examples with the original wound images rather than using only synthetic ones. The inclusion of the original preprocessed wound image sample was performed by adding the identity function as property in the Albumentations pipeline, which returns the image itself. This approach preserves the actual characteristics of the original data, ensuring that the model learns from authentic features and patterns present in the dataset.

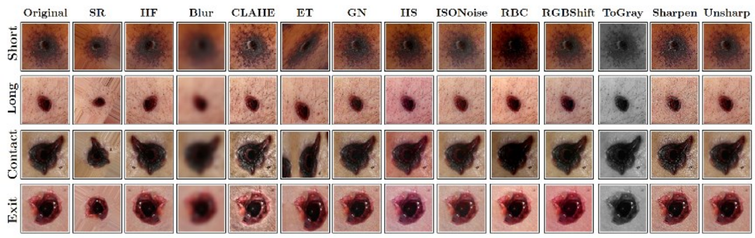


Fig. 2 Examples of gunshot wounds images and some their augmented versions. First line corresponds to a close range entry, second line corresponds to distance shot entry, third line presents the contact shot entry image, and the last line shows an example of exit gunshot wound. The results obtained from the used augmentation are depicted from left to right: the original (preprocessed) image, Safe Rotation (SR), Horizontal Flip (HF), blur, Contrast-Limited Adaptive Histogram Equalization (CLAHE), Elastic Transform (ET), Gaussian Noise (GN), Hue Saturation (HS), ISO noise, Random Brightness Constrast (RBC), gray representation of image (ToGray), sharpen, and unsharp mask

By combining original and augmented images, the model learns to recognize patterns from both the natural data distribution and the variations introduced by augmentation. By using this combination of training data, the model receives balanced exposure to both real and augmented data, enhancing its robustness to variations in real-world scenarios.

After combining the training data, a data resampling strategy is implemented in order to further reduce the remaining imbalance after data augmentation. Particularly, a combination of over- and under-sampling using Synthetic Minority Over-sampling Technique (SMOTE) and Edited Nearest Neighbours. (ENN) is applied for handling class imbalance with fewer noisy instances, leading to improved performance and robustness of ML models. It combines the oversampling of the minority class using SMOTE with the cleaning of the dataset using ENN [26]. This resampling strategy is implemented via Imbalanced-learn [27], an open-source Python library designed to handle imbalanced datasets in ML. Finally, the class weights are computed to train the TorchVision models.

Experimental setup

The experiments were performed on a Linux machine (Ubuntu 22.04 LTS) using an AI software system developed by ourselves using the Python language (version 3.10). Computer hardware specifications included an Intel i7-8700 CPU, 32GB of RAM, and an RTX 3090 GPU. The TorchVision models are trained and tested separately. For each model and training round, 1000 epochs are used. The models trained with a batch size of 16 include RegnetY128GF and ViTH14. The models trained with a batch size of 32 include EfficientnetB6, EfficientnetB7, EfficientnetV2L, RegnetY128GF, Squeezenet10, SwinTransformerT, SwinTransformerS, SwinTransformerB, SwinTransformerV2T, SwinTransformerV2S, SwinTransformerV2B, ShuffleNetV 2x20, ShuffleNetV2x15, ViTB16, ViTB32, ViTL16, ViTL32, ViTH14, and MaxVit. The remaining models not mentioned were trained with a batch size of 64. Early stopping with checkpointing is adopted to save the weights that perform best over the validation set during training.

Two cross-validation approaches were used: holdout and k-fold. For both approaches, the protocol consists of splitting the database into two content-independent subsets - one subset for training and another for testing. To avoid overfitting, the training and test data are split considering stratification based on the case number. This means that images associated with cases used in the training subset are not used for testing, and vice versa. With this constraint, 80% of cases are randomly selected for training, and the remaining 20% are used for testing in holdout cross-validation. Similarly, stratified grouped k-fold is adopted to ensure that each fold of the cross-validation procedure contains a balanced representation of these groups, providing a more realistic evaluation of the model's generalization capability. In this case, the groups used for stratification correspond to the crime cases.

We evaluate the performance of the tested models using regular classification metrics. More precisely, the models were assessed according to the following metrics:

- **Accuracy** (ACC): calculate the percentage of its correct predictions. It is defined as follows

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}.$$

- **Precision**: because the accuracy score may have as drawbacks the imbalance problem and being uninformative as a standalone classification metric, the precision measures the ability of a classifier not to label as Positive a Negative sample. It is defined as follows

$$\text{Precision} = \frac{TP}{TP + FP}.$$

- **Recall**: it is a ratio of predictions of the Positive class that are Positive by ground truth to the total number of Positive samples. In other words, recall measures the ability of a classifier to detect Positive samples. It is defined as follows

$$\text{Recall} = \frac{TP}{TP + FN}.$$

- **F1-Score**: it is the harmonic mean of precision and recall, representing both metrics in one score. The highest possible value of an F-score is 1, indicating perfect precision and recall, and the lowest possible value is 0, if either precision or recall is zero. It is defined as follows:

$$\text{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}.$$

- **Specificity**: it measures the ability of a classifier to correctly identify negative samples. It is calculated as the proportion of true negatives (TN) out of all actual negatives (TN + FP). Specificity is particularly useful when the cost of false positives is high, providing a comprehensive understanding of the performance of a classification model. It is defined as follows:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- **Area under the Curve (AUC)**: it evaluates the performance of a classification model by measuring the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (sensitivity) against the False Positive Rate (1 - specificity) at various threshold settings. The ROC curve is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The AUC is then computed as the area under this ROC curve:

$$\text{AUC} = \int_0^1 TPR\, d(FPR) = \sum_{i=1}^{n} \left(\frac{TPR_i + FPR_i}{2} \right) \cdot (FPR_i - FPR_{i-1}),$$

where

$$\text{TPR} = \frac{TP}{TP + FN} \quad \text{and} \quad \text{FPR} = \frac{FP}{FP + TN},$$

are the ratio of true positives to all actual positives (also known as sensitivity) and the ratio of false positives to all actual negatives, respectively. Using the trapezoidal rule, n is the number of thresholds.

In the context of evaluating classification models, the above-described performance metrics can be aggregated with macro, micro, and weighted averages. Here is a brief overview of each:

- **Macro average** calculates the metric (e.g., precision, recall, or F1-Score) for each class individually and then takes the average of these metrics. In other words, it treats all classes equally, regardless of their frequency as follows:

$$\text{Macro } \Phi = \frac{1}{C} \sum_{i=1}^{C} \Phi_i,$$

  where C is the number of classes and Φ_i is the performance metric for class i.
- **Micro average** aggregates the contributions of all classes to compute the average metric. It treats each instance equally, regardless of class.
- **Weighted average** calculates the metric for each class and then takes the average weighted by the number of instances in each class. It can be expressed as:

$$\text{Weighted } \Phi = \frac{\sum_{i=1}^{C} (\Phi_i \times \text{Number of Instances in Class}_i)}{\sum_{i=1}^{C} \text{Number of Instances in Class}_i}.$$

Cross-validation

In order to observe whether the ResNet152 results are stable, we conducted a K-fold cross-validation to ensure that this model is evaluated comprehensively and reliably. Using a boxplot to illustrate the k-fold cross-validation results is an effective way to visualize the distribution of the performance metrics across the different folds. The boxplot displays the median, quartiles, and potential outliers of the wound images, providing a clear summary of the results.

Results

In this study, 2,551 wound images were analyzed, primarily from male subjects (95.09%)(Table 1), with most images showing wounds located on the trunk (51.35%)(Table 3). The majority of the wounds were classified as entry wounds (1,883), with 668 exit wounds. In terms of MLSD, most entry wounds were classified as distant (85.45%), while close-range and contact wounds were less frequent (Table 2). The age distribution of the cases indicated a predominance of younger victims, with a mean age of approximately 27 years and a significant skew towards individuals in their 20s (Fig. 3a and b). These findings highlight key demographic and injury pattern trends in gunshot wound cases.

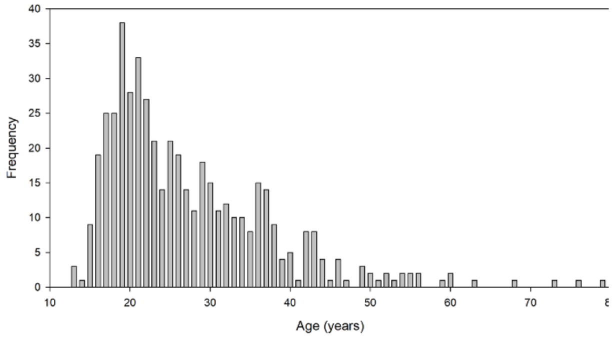
Wound type classification

Table 4 shows the accuracy, precision, recall and F1-score, AUC and specificity of would type classification using a houldout approach. This table discriminates the performance in terms of neural network architectures and its variants provided in the TorchVision library. In this table, the boldfaced values highlight the best model. Moreover, Table 4 shows the micro, macro, and weighted averages for each metric. From these results, we can notice that the results vary a lot for different models. However, it is noticeable that ResNet presents the best results. Among the variants of ResNet, the ResNet152 achieves the best performance for all metrics. Since The ResNet 152 achieved the highest values for 14 out of 16 evaluated metrics, we chose this model as the most suitable for wound type classification.

Figure 4a presents a boxplot with the results of the ResNet 152's wound type classification metrics.

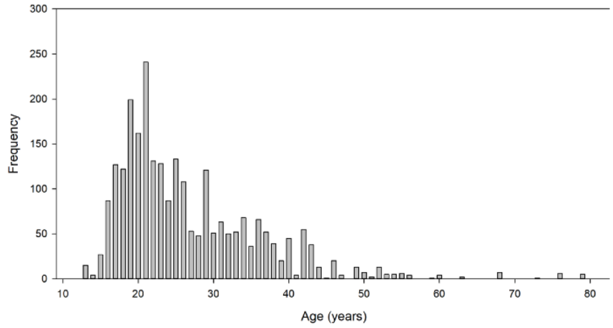
Medico-legal shooting distance classification

Similarly to what was done for the classification of gunshot wound types, the results for the detection of MLSD are presented in Table 5. The ResNet 152 achieved the highest scores for the used metrics. The only exceptions were the macro precision values, which were higher for the ResNet 50 version, and the weighted specificity indices, for which the SqueezeNet 10 version reached the highest mark. Figure 4b presents the distribution of ResNet 152 MLSD classification metrics.



Mean	St. Dev	Median	Mode	Skewness	Kurtosis
27.319	10.395	25	19	1.52	3.20

(a) Selected cases.



Mean	St. Dev	Median	Mode	Skewness	Kurtosis
26.680	9.557	24	21	1.63	4.11

(b) Selected images.

Fig. 3 Frequency of age at time of death

Discussion

This study marks a significant advancement in forensic science as it is among the first to employ DL tools to classify gunshot wound patterns with extensive labels such as entry and exit wounds and MLSD, utilizing a broad array of neural network architectures. Unique in its approach, this research utilizes a considerable dataset of forensic images captured during real crime scene investigations, a feat not matched by any previous work. The implementation of these methodologies holds the potential to notably accelerate the pace of real criminal investigations by offering rapid, reliable analyses that enhance traditional forensic methods. Furthermore, the technology serves as an complementary tool for forensic experts during the crucial phases of examination and report drafting, promoting more accurate and informed decision-making.

Clearly, for ethical reasons, conducting research with a completely controlled sample of gunshot wounds on human cadavers, like studies of Phelps in 1897 [43], is virtually impractical. Therefore, many studies still use animal carcasses [17, 44, 45]. However, as much as these may approximate what is seen in humans, there are significant differences [46] that could lead to the incorrect training of the ML model for future use in human cases.

In this study, for classifying entry and exit gunshot wounds images, the use of ResNet152 achieved percentages, with the exception of the AUC results, closely align with those by Cheng et al. [16], which reported an accuracy of 87.99%, precision of 83.99%, recall of 87.71%, specificity of 88.19%, F1-score of 85.81% and AUC of 94.6% with a larger sample, comprising 2,028 entry and 1,314 exit wound images. However, the images do not appear to have been obtained at crime scenes, but rather during necropsy examinations,

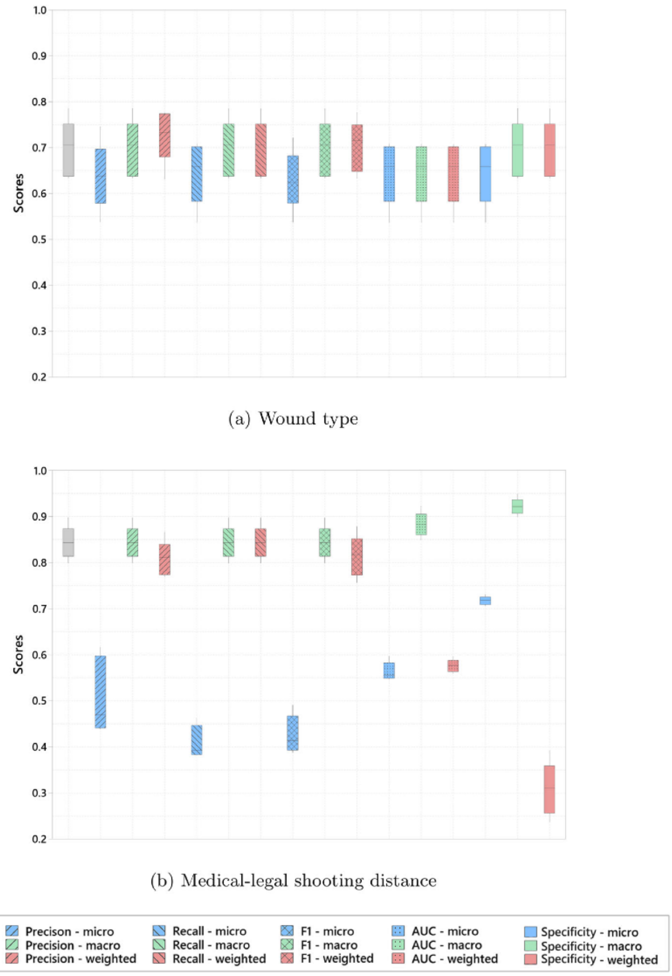


Fig. 4 Classification performance for wound type (a) and shooting distance (b) using ResNet 152 over the k-folds

which indicates a certain level of control over the captured images.

Additionally, in shooting distance classification, this study achieved a 92.48% accuracy rate using ResNet152. However, despite the high accuracy, the sample imbalance may have negatively impacted other metrics, including recall and the F1-Score. While newer models like ViT and EfficientNet offer advanced techniques and may perform exceptionally well in many scenarios, ResNet's combination of residual connections, depth, pretraining, and optimization make it a robust choice for a wide range of tasks, including the MLSD classification task. In contrast, Oura et al. [17] reached a 98% accuracy rate. However, this higher accuracy was probably facilitated by their use of a small, controlled sample comprising only 204 images in a laboratory setting. It is important to note that their study utilized pig carcasses instead of wound on humans corpses, which could affect the applicability of their findings to real-world forensic scenarios.

With this, it becomes evident that one of the greatest challenges in developing ML models for image categorization is acquiring a database with high-quality scientific data [47]. Due to the sensitive content nature of the area, this challenge may be even greater in the forensic field. Therefore, the photographs used in this study were taken by different individuals, under various environmental and exposure conditions, with a range of cameras, and in the routine circumstances of forensic work at crime scenes, where standardization becomes virtually unfeasible. Given this, it is understood that a standardized database would not only misalign with the realities of forensic work, where numerous factors can alter the characteristics of injuries, but also require an exhaustive effort.

The challenge in standardizing a database of gunshot wound images is also evident in the discrepancy between the number of entry (1,883) and exit wounds (668). Logically, for an exit wound to occur, there must inevitably be an entry wound. Additionally, in this study, inclusion criterion #5 required that bullets be lodged in the body. Therefore, it is expected that the number of entry wounds would surpass that of exit wounds. Furthermore, even with access to a database containing millions of images from crime scene investigations, the utility of this dataset was significantly limited. Many images were related to other types of police incidents or featured different instruments (e.g., knives), had poor image quality, and faced restrictive criteria, all of which significantly constrained the acquisition of a larger, more useful sample. Despite these limitations, the sample used in this study more accurately reflects the conditions forensic teams encounter in real-world scenarios and deviates from the laboratory standardization typical of other research [16, 17]. This approach enhances the practical relevance of the findings to real-world forensic investigations.

In turn, the classification of distance shooting is used in cases where the characteristics of close-range and contact shots are not present (examples: smoke pattern and tattooing) [48] and typically occur at a distance range starting from 18 inches-2 feet (0.45 to 0.60 meters) [5, 49]. Therefore, for MLSD classification purposes, there is a tendency for the majority of injuries to be of the distance type, as per the sample obtained (85.45%). Unfortunately, for these reasons, it was not possible to achieve a more uniform sample regarding the number of images available in each distance category.

Depending on the location of the wound on the body, the characteristics displayed also change. The flaccidity of the region, as well as its elasticity and density, are some of the factors that influence the appearance of the wounds [50–52]. In the sample used, most of the wounds were located on the trunk (51.35%). However, even within the torso, there are quite distinct areas, such as the abdomen compared to the sternum area. In turn, given that the wounds analyzed are situated on the skin, age greatly affects the activity of collagen and, consequently, elasticity [53]. Since 66.67% of the cases and 70.29% of the images are from individuals below thirty years of age, it is expected that the effect of age on elasticity did not have a significant influence. Finally, the differences in wounds due to sex, because of the arrangement of Langer's lines, skin thickness, and fat accumulation locations [52], were minimized by having a sample predominantly composed of male victims (95.09%).

In the field of scientific inquiry, it is crucial to acknowledge that AI serves not as a replacement for human expertise but as a formidable ally. The intrinsic value of human judgment and knowledge proves irreplaceable, especially in interpreting complex contexts and making well-informed decisions. Therefore, it is imperative to employ AI systems within forensic sciences as supportive instruments that enhance, rather than replace, the capabilities of forensic professionals [10, 47]. In real investigative scenarios, we can envision various applications, such as assisting pathologists in decision-making and accelerating criminal investigations. This approach ensures that the integration of AI into forensic practices augments the efficacy and accuracy of investigations, maintaining the indispensable balance between technological innovation and the nuanced understanding that only human experts can provide [47].

Limitations

This study presents significant strides in computer vision for the classification of gunshot wounds. However, while the dataset comprises images from actual forensic cases, there is a significant imbalance in the number of images available for each category of wounds, which could skew the accuracy of the model. Additionally, these images reflect a wide range of environmental and exposure conditions, contributing to variability in image quality that can affect the performance of the machine learning models. The imbalance in wound categories introduces potential selection bias, as certain types of wounds may be overrepresented, leading to models that perform well on the training data but struggle with generalization. Furthermore, the study lacks external validation, as the models were primarily validated using a dataset from a single institution located at Federal District, Brazil. Future work should focus on testing the models across diverse populations to improve robustness. To enhance model performance further, exploring resampling and training strategies, such as Data Center Interpolation with Improved Sparrow Search Algorithm [54], as well as incorporating classical feature-engineered models like Hu Moments, Haralick texture features, and local binary patterns, could offer complementary approaches for improving the robustness and accuracy of the models under real-world forensic conditions.

Conclusions

This study explored the use of fifty-nine deep neural network architectures for classifying gunshot wound images from real crime cases, focusing on distinguishing between entry and exit wounds, as well as determining the MLSD in a forensic context. The ResNet152 architecture demonstrated superior performance in wound type classification, achieving up to 86.90% accuracy, precision, recall, F1-score and specificity and an AUC of 82.09%. For MLSD, the same architecture reached an accuracy of 92.48%, precision, recall and F1-score up to 92.48%, AUC up to 94.36% and specifity up to 96.24% , although sample imbalance affected the metrics. Our findings underscore the challenges of standardizing wound images due to varying capture conditions, yet they reflect the practical realities of forensic work. This research highlights the significant potential of deep neural networks in forensic pathology, advocating for AI as a supportive tool to enhance, not replace, human expertise, ensuring the balance between technological innovation and expert judgment in forensic investigations.

Acknowledgements This work was supported by the Fundação de Apoio a Pesquisa do Distrito Federal (FAP-DF), by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and by the University of Brasília (UnB).

Author Contributions Author 1 took a lead role, overseeing the conceptualization of the study, data curation, resource management, investigation, and the preparation of the original draft. Authors 2, 3, and 4 were pivotal in data curation, ensuring that the necessary data for the study's objectives were accurately collected and organized. Author 5 contributed significantly to the study's conceptualization, developed the methodology, managed the software aspects, and participated in the writing and editing processes. Author 6 enhanced the research with software development, validation, and additional investigation. Authors 7 and 8 played crucial roles in conceptualizing the project, supervising the research activities, managing the project, and refining the manuscript through their critical review and editing, ensuring both the rigor and integrity of the study.

Funding Sources This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability The image dataset used in this paper is available at https://github.com/pedrogarciafreitas/FDCPUnBGunshotDB.git and the processed data used in the analysis were generated using the source code publicly available at https://gitlab.com/lisa-unb/leguwoi.git.

Declarations

The research involved the use of photographic records and reports from homicide and suicide cases handled by the PCDF in Brasília, Brazil. Proper authorization was obtained from the institution, and the study was approved by the Ethics Committee for Research at the University of Brasília under protocol number 54418221.9.0000.0030. As the study did not involve direct participation of individuals, consent to participate was not applicable.

Competing Interest The authors have no competing interests to declare.

References

1. UNODC: Global study on homicide 2023. https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html. Accessed 08 Mar 2024
2. Wisevoter: Gun deaths by country 2023. https://wisevoter.com/country-rankings/gun-deaths-by-country. Accessed 08 Mar 2024
3. Kleypas DA, Badiye A (2024) Evidence collection vol 2, pp 113–116. https://pubmed.ncbi.nlm.nih.gov/28722882/
4. Shrestha R, Kanchan T, Krishan K (2024) Gunshot wounds forensic pathology, pp 1–8. StatPearls Publishing, Treasure Island, Florida, United States. http://www.ncbi.nlm.nih.gov/pubmed/32310579
5. Denton JS, Segovia A, Filkins JA (2006) Practical pathology of gunshot wounds. Arch Pathol Lab Med 130(9):1283–9. https://doi.org/10.5858/2006-130-1283-PPOGW
6. Di Maio VJM (2015) Gunshot wounds: practical aspects of firearms, ballistics, and forensic techniques, 3rd edn. CRC Press, Boca Raton, Florida, United States, p 377
7. Saukko PJ, Knight B (2016) Knight's Forensic Pathology, 4th edn. CRC Press, Boca Raton, Florida, United States
8. Adelson L (1961) A microscopic study of dermal gunshot wounds. Am J Clin Pathol 35(5):393–402. https://doi.org/10.1093/ajcp/35.5.393
9. Dror IE, Pierce ML (2020) ISO standards addressing issues of bias and impartiality in forensic work. J Forensic Sci 65(3):800–808. https://doi.org/10.1111/1556-4029.14265
10. Vodanović M, Subašić M, Milošević D, Galić I, Brkić H (2023) Artificial intelligence in forensic medicine and forensic dentistry. J Forensic Odonto-Stomatol 41(2):30–41
11. Tournois L, Trousset V, Hatsch D, Delabarde T, Ludes B, Lefèvre T (2023) Artificial intelligence in the practice of forensic medicine: a scoping review. Int J Legal Med. https://doi.org/10.1007/s00414-023-03140-9
12. Lefèvre T, Tournois L (2023) Artificial intelligence and diagnostics in medicine and forensic science. Diagnostics 13(23):1–12. https://doi.org/10.3390/diagnostics13233554
13. Wankhade TD, Ingale SW, Mohite PM, Bankar NJ (2022) Artificial intelligence in forensic medicine and toxicology: the future of forensic medicine. Cureus 14(8). https://doi.org/10.7759/cureus.28376
14. Galante N, Cotroneo R, Furci D, Lodetti G, Casali MB (2022) Applications of artificial intelligence in forensic sciences: current potential benefits, limitations and perspectives. Int J Legal Med 445–458. https://doi.org/10.1007/s00414-022-02928-5
15. Margagliotti G, Bollé T (2019) Machine learning & forensic science. Forensic Sci Int 298:138–139. https://doi.org/10.1016/j.forsciint.2019.02.045
16. Cheng J, Schmidt C, Wilson A, Wang Z, Hao W, Pantanowitz J, Morris C, Tashjian R, Pantanowitz L (2024) Artificial intelligence for human gunshot wound classification. J Pathol Inform 15:100361. https://doi.org/10.1016/j.jpi.2023.100361
17. Oura P, Junno A, Junno J-A (2021) Deep learning in forensic gunshot wound interpretation—a proof-of-concept study. Int J Legal Med 135:2101–2106. https://doi.org/10.1007/s00414-021-02566-3
18. Molina DK, Dimaio V, Cave R (2013) Gunshot wounds: a review of firearm type, range, and location as pertaining to manner of death. Am J Forensic Med Pathol 34:366–371. https://doi.org/10.1097/PAF.0000000000000065
19. Marcel S, Rodriguez Y (2010) Torchvision the machine-vision package of torch. In: Bimbo AD, Chang S, Smeulders AWM (eds) Proceedings of the 18th international conference on multimedia

2010, Firenze, Italy, October 25-29, 2010, pp 1485–1488. Association for Computing Machinery, New York, New York, United States. https://doi.org/10.1145/1873951.1874254
20. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang EZ, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) Pytorch: an imperative style, high-performance deep learning library. In: Wallach HM, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox EB, Garnett R (eds) Advances in neural information processing systems 32: annual conference on neural information processing systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pp 8024–8035. https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html
21. Tietz M, Fan TJ, Nouri D, Bossan B (2017) skorch Developers: Skorch: a scikit-learn compatible neural network library that wraps PyTorch. https://skorch.readthedocs.io/en/stable/
22. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, Niculae V, Prettenhofer P, Gramfort A, Grobler J, Layton R, VanderPlas J, Joly A, Holt B, Varoquaux G (2013) Api design for machine learning software: experiences from the scikit-learn project, vol. abs/1309.0238. http://dblp.uni-trier.de/db/journals/corr/corr1309.html#BuitinckLBPMGNPGGLVJHV13
23. Mumuni A, Mumuni F (2022) Data augmentation: A comprehensive survey of modern approaches. Array 16:100258. https://doi.org/10.1016/J.ARRAY.2022.100258
24. Hataya R, Zdenek J, Yoshizoe K, Nakayama H (2020) Faster autoaugment: learning augmentation strategies using backpropagation. In: Vedaldi A, Bischof H, Brox T, Frahm J (eds) Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXV. Lecture Notes in Computer Science, vol 12370, pp 1–16. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-030-58595-2_1
25. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA (2020) Albumentations: fast and flexible image augmentations. Information 11(2):125. https://doi.org/10.3390/INFO11020125
26. Batista GEAPA, Prati RC, Monard MC (2004) A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor 6(1):20–29. https://doi.org/10.1145/1007730.1007735
27. Lemaître G, Nogueira F, Aridas CK (2017) Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res 18(17):1–5
28. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90. https://doi.org/10.1145/3065386
29. Huang G, Liu Z, Maaten L, Weinberger KQ (2017) Densely connected convolutional networks, pp 2261–2269. https://doi.org/10.1109/CVPR.2017.243
30. Tan M, Le QV (2019) Efficientnet: rethinking model scaling for convolutional neural networks, vol 97, pp 6105–6114
31. Tan M, Le QV (2021) Efficientnetv2: smaller models and faster training, vol 139, pp 10096–10106
32. Tu Z, Talebi H, Zhang H, Yang F, Milanfar P, Bovik AC, Li Y (2022) Maxvit: multi-axis vision transformer, vol 13684, pp 459–479. https://doi.org/10.1007/978-3-031-20053-3_27
33. Tan M, Chen B, Pang R, Vasudevan V, Sandler M, Howard A, Le QV (2019) Mnasnet: platform-aware neural architecture search for mobile, pp 2820–2828. https://doi.org/10.1109/CVPR.2019.00293
34. Sandler M, Howard AG, Zhu M, Zhmoginov A, Chen L (2018) Mobilenetv2: inverted residuals and linear bottlenecks, pp 4510–4520. https://doi.org/10.1109/CVPR.2018.00474
35. Howard A, Pang R, Adam H, Le QV, Sandler M, Chen B, Wang W, Chen L, Tan M, Chu G, Vasudevan V, Zhu Y (2019) Searching for mobilenetv3, pp 1314–1324. https://doi.org/10.1109/ICCV.2019.00140
36. Radosavovic I, Kosaraju RP, Girshick RB, He K, Dollár P (2020) Designing network design spaces, pp 10425–10433. https://doi.org/10.1109/CVPR42600.2020.01044
37. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition, pp 770–778. https://doi.org/10.1109/CVPR.2016.90
38. Ma N, Zhang X, Zheng H, Sun J (2018) Shufflenet V2: practical guidelines for efficient CNN architecture design, vol 11218, pp 122–138. https://doi.org/10.1007/978-3-030-01264-9_8
39. Iandola FN, Moskewicz MW, Ashraf K, Han S, Dally WJ, Keutzer K (2016) Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. CoRR arXiv:1602.07360
40. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: hierarchical vision transformer using shifted windows, pp 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986
41. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: Bengio Y, LeCun Y (eds) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. arXiv:1409.1556
42. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2021) An image is worth 16x16 words: transformers for image recognition at scale. In: International conference on learning representations. https://openreview.net/forum?id=YicbFdNTTy
43. Phelps C (1897) Traumatic injuries of the brain and its membranes. Appleton, New York, New York, United States
44. Pircher R, Preiß D, Pollak S, Thierauf-Emberger A, Perdekamp MG, Geisenberger D (2017) The influence of the bullet shape on the width of abrasion collars and the size of gunshot entrance holes. Int J Legal Med 131(2):441–445. https://doi.org/10.1007/s00414-016-1501-6
45. Rainio J, Lalu K, Ranta H, Penttilä A (2003) Morphology of experimental assault rifle skin wounds. Int J Legal Med 117(1):19–26. https://doi.org/10.1007/s00414-002-0308-9
46. Bartell TH, Mustoe TA (1989) Animal models of human tissue expansion. Plast Reconstr Surg 83(4):681–6. https://doi.org/10.1097/00006534-198904000-00014
47. Piraianu AI, Fulga A, Musat CL, Ciobotaru OR, Poalelungi DG, Stamate E, Ciobotaru O, Fulga I (2023) Enhancing the evidence with algorithms: how artificial intelligence is transforming forensic medicine. Diagnostics 13(18). https://doi.org/10.3390/diagnostics13182992
48. Smith JC (1964) Investigation of fatal gunshot wounds. Cleve-Marshall Law Rev 13(2):296–308
49. Svensson A, Wendel O, Nicol JD (1965) Techniques of crime scene investigation. American Elsevier Publishing Company, New York, New York, United States
50. Mahoney P, Carr D, Arm R, Gibb I, Hunt N, Delaney RJ (2017) Ballistic impacts on an anatomically correct synthetic skull with a surrogate skin/soft tissue layer. Int J Legal Med 1–12. https://doi.org/10.1007/s00414-017-1737-9
51. Humphrey C, Kumaratilake J (2016) Ballistics and anatomical modelling – a review. Legal Med 23:21–29. https://doi.org/10.1016/j.legalmed.2016.09.002
52. Ridge MD, Wright V (1966) The directional effects of skin. A bio-engineering study of skin with particular reference to Langer's lines. J Investig Dermatol 46(4):341–346. https://doi.org/10.1038/jid.1966.54
53. Maiden NR, Byard RW (2016) Unpredictable tensile strength biomechanics may limit thawed cadaver use for simulant

research. Aust J Forensic Sci 48(1):54–58. https://doi.org/10.1080/00450618.2015.1025842

54. Gu Q, Tian J, Li X, Jiang S (2022) A novel random forest integrated model for imbalanced data classification problem. Knowl-Based Syst 250:109050. https://doi.org/10.1016/j.knosys.2022.109050

Authors and Affiliations

Renato Queiroz Nogueira Lira[1] · Luana Geovana Motta de Sousa[1] · Maisa Luana Memoria Pinho[1] · Renan Cesar Pinto da Silva Andrade de Lima[1] · Pedro Garcia Freitas[2] · Bruno Scholles Soares Dias[3] · Andreia Cristina Breda de Souza[4] · André Ferreira Leite[1]

✉ Renato Queiroz Nogueira Lira
renato.lira@aluno.unb.br; renatolira@msn.com

Luana Geovana Motta de Sousa
luana.geovana@aluno.unb.br

Maisa Luana Memoria Pinho
maisa.pinho@aluno.unb.br

Renan Cesar Pinto da Silva Andrade de Lima
renan.lima@aluno.unb.br

Pedro Garcia Freitas
pedro.freitas@unb.br

Bruno Scholles Soares Dias
brunoscholles98@gmail.com

Andreia Cristina Breda de Souza
andreiacrisbreda@gmail.com

André Ferreira Leite
andreleite@unb.br

1 Department of Dentistry, University of Brasília, Campus Darcy Ribeiro, Asa Norte, Brasília 70910-900, Distrito Federal, Brazil

2 Department of Computer Science, University of Brasília, Campus Darcy Ribeiro, Asa Norte, Brasília 70910-900, Distrito Federal, Brazil

3 Department of Electrical Engineering, University of Brasília, Campus Darcy Ribeiro, Asa Norte, Brasília 70910-900, Distrito Federal, Brazil

4 Department of Forensic Dentistry and Public Health, Federal University of Rio de Janeiro, Cidade Universitária, Rio de Janeiro 21941-617, Rio de Janeiro, Brazil

ANEXO A - DOCUMENTO DE APROVAÇÃO PELO COMITÊ DE ÉTICA

DADOS DA VERSÃO DO PROJETO DE PESQUISA

Título da Pesquisa: ESTIMATIVA DE CALIBRES DE PROJÉTEIS DE ARMA DE FOGO E DE DISTÂNCIA DE DISPARO POR MEIO DE FOTOGRAFIAS E MACHINE LEARNING
Pesquisador Responsável: Renato Queiroz Nogueira Lira
Área Temática:
Versão: 4
CAAE: 54418221.9.0000.0030
Submetido em: 24/08/2022
Instituição Proponente: Departamento de Odontologia - Faculdade de Ciências da Saúde - UNB
Situação da Versão do Projeto: Aprovado
Localização atual da Versão do Projeto: Pesquisador Responsável
Patrocinador Principal: Financiamento Próprio

PLATAFORMA BRASIL COORDENADOR PLATAFORMA BRASIL

Comprovante de Recepção: PB_COMPROVANTE_RECEPCAO_1857834

ANEXO B – CARTA DE ACEITE

Outlook

IJLM: Your manuscript entitled Deep Learning-Based Human Gunshot Wounds Classification - [EMID:4a474adaf3743a84]

De Tony Fracasso <em@editorialmanager.com>
Data Sex, 18/10/2024 10:30
Para Renato Lira <renatolira@msn.com>

Ref.:
Ms. No. IJLM-D-24-00283R1
Deep Learning-Based Human Gunshot Wounds Classification
International Journal of Legal Medicine

Dear Mr Lira,

I am pleased to tell you that your work has now been accepted for publication in International Journal of Legal Medicine.

Thank you for submitting your work to this journal.

With kind regards

Tony Fracasso
Editor-in-Chief
International Journal of Legal Medicine